Inverse Modeling of Beaver Reservoir's Water Spectral Reflectance

V. Garg, I. Chaubey, C. Maringanti, S. G. Bajwa

ABSTRACT. *Estimation of inherent optical properties (IOP) needed for water quality evaluation by remote sensing models is very complex, primarily due to the large number of model simulations needed to find optimal parameter values. This study presents an approach for optimally parameterizing the IOP values of a physical hyperspectral optical - Monte Carlo (PHO-MC) model. An artificial neural network (ANN) based pseudo simulator combined with the Nondominated Sorted Genetic Algorithm II (NSGA II) was used to efficiently perform a large number of model simulations and to search the optimal parameter values for IOP determination. Concentrations of suspended matter (sm), chlorophyll-a (chl), and total dissolved organic matter (DOM) along with the reflectance data at 16 different wavelengths were measured at 48 sampling stations in the Beaver Reservoir, Arkansas, between 2003 and 2005 and were used to evaluate the IOP values. Measured concentrations and reflectance data from 24 sampling stations were used to optimize IOP parameter values for sm, chl, and DOM. The data collected from the remaining 24 sampling stations were used for the validation of PHO-MC model-predicted reflectance by using optimized IOP values. PHO-MC predicted reflectance values were significantly correlated (r = 0.90, p < 0.01) with the corresponding measured reflectance values, indicating that the pseudo simulator combined with the NSGA II accurately estimated the IOP values. An estimated 10^{10} years of calculation time was reduced to less than 3 min by using the pseudo simulator and NSGA II to supplement the PHO-MC model for estimating the IOP values.*

Keywords. ANN, Beaver Reservoir, GA, Inherent optical properties, Inverse modeling, Remote sensing, Water quality.

Major optically active constituents (OAC) of water that are predictable using remote sensing are chlorophyll-a (chl), total suspended matter (sm), and dissolved organic matter (DOM) (Dekker and Bukata, 2002). In physics-based models of remote sensing, light interactions with water and its OAC are solved in the framework of radiative transfer theory to calculate the reflectance either leaving the water surface or at some depth from the water surface. Light interactions with water and its OAC are determined based on the inherent optical properties (IOP) of the OAC (Mobley, 1994). Factors affecting prediction accuracy include model structure, IOP of water quality constituents of interest, and boundary conditions. Among these factors, IOP values of OAC are considered insufficiently known because a large regional and temporal variability may exist in their values (Yacobi et al., 1995; Morel and Maritorena, 2001). For example, IOP values of suspended matter may vary by more than an order of magnitude due to their variation in shape, size, and composition (Sathyendranath et al., 1987; Hoepffner and Sathyendranath, 1993; Ciotti et al., 2002; Cota et al., 2003; Lee and Carder, 2004).

The IOP values specific to OAC of a water body are determined either by using direct in-situ measurements or by inverse modeling. Direct in-situ measurements of IOPs require special instruments, such as a hyperspectral absorption and attenuation meter (AC-S, WET-Labs, Philomath, Ore.), multispectral measurement instrument for both backscattering and fluorescence (HydroScat-6, HobiLabs, Inc., Bellevue, Wash.), and beam scattering and attenuation sensor (c-beta, HobiLabs, Inc., Bellevue, Wash.) (Stramska et al., 2000; Babin and Stramski, 2002). Full-scale experiments involving this equipment are not only expensive and time-consuming but also present technical difficulties. The accuracy of scattering measurements by instruments is affected by a number of factors such as radiometric calibration; sensor-response function and optical geometry (involving the scattering volume, illumination beam, detection of scattered light, and path length in the water); proper angular resolution; temperature and pressure effects; as well as optical and mechanical imperfections of the instruments (Stramski et al., 2004a). In inverse modeling, parameter values of IOP are determined based on known reflectance of a water body and known OAC concentrations in that water body. Inverse modeling has been attempted in an effort to overcome the difficulties faced during the in-situ measurements and has been reported to improve the accuracy of remote sensing of water quality (Sathyendranath et al.,

The authors are **Vijay Garg,** Former Graduate Student, Department of Biological and Agricultural Engineering, University of Arkansas, Fayetteville, Arkansas; **Indrajeet Chaubey, ASABE Member,** Associate Professor, Department of Agricultural and Biological Engineering, Department of Earth and Atmospheric Sciences, and Division of Environmental and Ecological Engineering, Purdue University, West Lafayette, Indiana; **Chetan Maringanti, ASABE Member Engineer,** Graduate Student, Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, Indiana; and **Sreekala G. Bajwa, ASABE Member Engineer,** Associate Professor, Department of Biological and Agricultural Engineering, University of Arkansas, Fayetteville, Arkansas. **Corresponding author:** Indrajeet Chaubey, Department of Agricultural and Biological Engineering, 225 S. University Street, Purdue University, West Lafayette, IN 47907; phone: 765-494-5013; fax: 479-496-1115; e-mail: ichaubey@purdue.edu.

2001). In recent years, inverse modeling of IOP for water quality prediction has received much attention, as evident by a number of studies (e.g., Gordon and Boynton, 1997; Barnard et al., 1999; Loisel and Stramski, 2000; Stramska et al., 2000; Loisel et al., 2001; Stramski et al., 2001; Babin and Stramski, 2002; Risovic, 2002; Babin et al., 2003; Cota et al., 2003; McKee et al., 2003; Babin and Stramski, 2004; Stramski et al., 2004b).

A major limitation of inverse modeling to estimate IOP parameter values is the large number of model runs needed to search the entire parameter space of the possible solutions. Therefore, previous efforts to determine IOP values using inverse modeling have used semi-analytic remote sensing models (Barnard et al., 1999; Stramska et al., 2000; Loisel et al., 2001; Lee et al., 2002; Hamre et al., 2003; McKee et al., 2003), which are basically simple approximation of the radiative transfer equations in one spatial dimension (Mobley et al., 1993; Mobley, 1994). Even though the semi-analytic models are computationally very efficient, they may not account for all experimental conditions depending upon assumptions made in the development of these models. Alternatively, stochastic Monte Carlo models are flexible for any geometrical situation and can be used even when boundary conditions and water constituent vary in all three spatial dimensions (Mobley, 1994; Doyle and Rief, 1998). However, Monte Carlo models are computationally inefficient and running a model can be time consuming, especially when large numbers of runs are required. Thus, inverse modeling of IOPs using a Monte Carlo model may need years of computation time, primarily because the model has to perform a very large number of simulations to search parameter values that meet a given objective function.

The computation time of running a model can be reduced by replacing the original model with a suitable pseudo simulator, which can mimic the behavior of the model. Artificial neural networks (ANN) can be a viable choice for such a pseudo simulator. ANN algorithms have the ability to generalize patterns of input data and to synthesize a complex model without prior knowledge of exact input-output relationships. Application of a proper search mechanism determines the number of searches required to arrive at the optimum solution. Therefore, a proper search mechanism is one of the most decisive factors in successful implementation of inverse modeling of IOP values. When the objective function is not continuously differentiable throughout the solution domain, techniques such as gradient descent may fail to converge. A global search algorithm, such as a genetic algorithm (GA), may be preferable for this purpose.

The overall goal of this study was to reduce the computational burden in estimating IOP parameter values using inverse modeling of an analytical remote sensing model (Physical Hyperspectral Optical - Monte Carlo, or PHO-MC model). The large computation requirements were reduced by utilizing an artificial neural network (ANN) based pseudo simulator and Nondominated Sorted Genetic Algorithm II (NSGA II), which supplemented the PHO-MC model. The optimized IOP parameter values for chl, sm, and DOM were validated using measured spectral reflectance and concentration data from an independent set of 24 sampling locations that were not used in the IOP parameter optimization.

MATERIALS AND METHODS
DESCRIPTION OF THE PHO-MC MODEL

The Physical Hyperspectral Optical - Monte Carlo (PHO-MC) remote sensing model used in this study solves the radiative transfer equation using a Monte Carlo approach (Mobley et al., 1993; Mobley, 1994):

$$\mu \frac{dL(z;\xi;\lambda)}{dz} = -c(z;\lambda)L(z;\xi;\lambda)$$

$$+ \iint\limits_{\xi \in \Xi} L(z;\xi';\lambda)\beta(z;\xi' \to \xi;\lambda)d\Omega(\xi') + S(z;\xi;\lambda) \quad (1)$$

where $L(z;\xi';\lambda)$ is the unpolarized spectral radiance at wavelength λ, depth z, and in direction $\xi = (\theta, \varphi)$; c is the total attenuation coefficient; β is the volume scattering function; and S is the internal source of radiance.

In the PHO-MC model, a photon packet path is traced in a three-dimensional space by accounting for directional and state changes suffered by a photon at various points in the water defined by suitable probability distributions based on IOP properties of water and its OAC. The tracing of a photon packet path starts just before the point when it enters into the water body. The path is followed until the termination of the photon packet, due either to its complete absorption or its escape through the air-water interface. Reflectance $R(\lambda)$ at wavelength λ is calculated by initiating sufficiently large numbers of photons. To generate a hyperspectral graph of reflectance of a water medium, the model is run for photons of different wavelengths. The IOP of water and its OAC, which vary with the wavelength, are used as an input in the model. Other inputs to the model are percent diffused light, reflectance characteristics of the bottom, sun angle, refractive index of air, refractive index of water, and depth of water. The model has been shown to simulate reflectance of both clear and sediment-laden water bodies accurately (fig. 1). Detailed descriptions of the model can be obtained from Garg (2006) and Garg et al. (2009) and are not provided here for brevity.
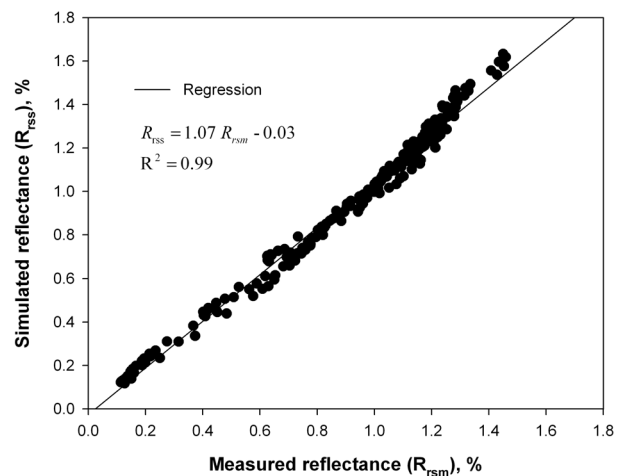

Figure 1. PHO-MC model simulated reflectance (R_{rss}) versus measured reflectance (R_{rsm}) of the clear water surface in a tank study for 246 combinations of six water depths (0.75, 0.60, 0.50, 0.40, 0.20, and 0.10 m) and 41 wavelengths (10 nm apart, in the range of 400 to 800 nm).

STUDY SITE AND DATA COLLECTION
METHODOLOGY

This study was conducted in the Beaver Reservoir, located in the Ozark Plateau of northwest Arkansas. Eight to ten water quality and reflectance data from the Beaver Reservoir were collected on five clear sky days of 15 September 2003, 19 December 2003, 19 October 2004, 13 December 2004, and 12 March 2005. A total of 48 water samples were collected concurrent with spectral measurements from different locations of the reservoir (fig. 2). Equal amounts of water samples were collected from three depths (surface, 1 m below surface, and 2 m below surface), which were composited to make a 4 L water sample. Water samples were stored in the dark on ice until returned to the laboratory. An amount of 1,000 to 1,500 mL was filtered onto Gelman GF/C 47 mm filters for both chl and sm determination. Filters for chl extraction were macerated in 5 mL of 90% acetone solution. Extracts were then cleared by centrifugation and analyzed spectrometrically (APHA, 1998). Filters for determination of sm were pre-weighed. After sample filtration, filters were oven-dried at 105°C for 24 h and re-weighed to determine sm. Dissolved organic matter was estimated by chemical analysis of total organic carbon of a filtered 250 mL sample by EPA Method 415.1.

The upwelling radiance of water and solar radiance from a Lambertian Spectralon reference panel were measured by a portable spectroradiometer (FieldSpec Pro Dual VNIR, Analytical Spectral Devices, Boulder, Colo.). The
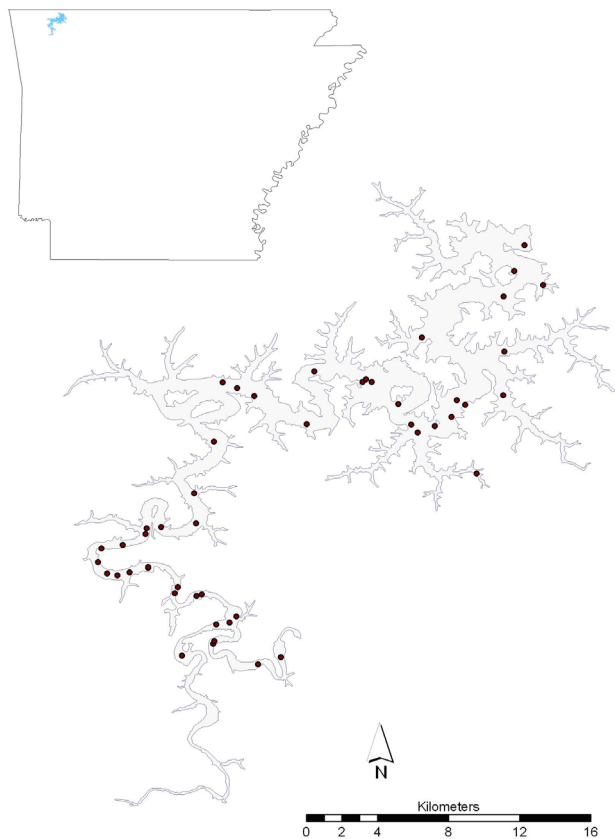

Figure 2. Location of the Beaver Reservoir in Arkansas and sampling sites (shown as filled circles) where water quality and remote sensing data were collected in this study.

instrument collected data in 512 discrete, contiguous bands with a nominal bandwidth of 1.438 nm and spectral range of 336 to 1,071 nm for two sensors. In this analysis, reflectance data between 400 to 800 nm were used. Upwelling radiance from the water surface $L_w(\lambda)$ was collected at nadir view angle holding the target sensor at about 30 cm above the water surface. Downwelling radiance arriving at the water surface $L_s(\lambda)$ was measured by positioning the reference sensor above a Lambertian white reference panel (0.07 × 0.07 m) made of barium sulfate. All measurements were made five times, and mean values were used. The percent reflectance was calculated as:

$$R(\lambda) = 100 \cdot L_w(\lambda)/L_s(\lambda) \qquad (2)$$

DESCRIPTION OF IOP REPRESENTATION
IN THE PHO-MC MODEL

The IOP values of chl, sm, and DOM are a function of their concentrations in water and wavelength of the electromagnetic radiation. Therefore, IOP values for these OAC vary not only from one sampling location to another, but also for the different wavelengths used to measure reflectance at the same sampling location. Hence, IOP values of OAC were developed as a function of their concentrations and wavelength. IOP values are defined by two coefficients: (1) a part of the electromagnetic radiation of wavelength λ is absorbed by the OAC and is represented by the absorption coefficient $a(\lambda)$; (2) another part of the electromagnetic radiation of wavelength λ is scattered by the OAC and is represented by the scattering coefficient $b(\lambda)$. The absorption and scattering coefficients for various OAC are additive in nature and can be written as (Sathyendranath et al., 1989; Aleshin, 2001; Arst, 2003):

$$a(\lambda) = a_w(\lambda) + a_{chl}(\lambda) + a_{sm}(\lambda) + a_{DOM}(\lambda)$$

$$= a_w(\lambda) + a^*_{chl}(\lambda)C_{chl} + a^*_{sm}(\lambda)C_{sm}$$

$$+ a^*_{DOM}(\lambda)C_{DOM} \qquad (3)$$

$$b(\lambda) = b_w(\lambda) + b_{chl}(\lambda) + b_{sm}(\lambda) = b_w(\lambda)$$

$$+ b^*_{chl}(\lambda)\, C_{chl} + b^*_{sm}(\lambda)\, C_{sm} \qquad (4)$$

where a and b are the absorption and scattering coefficients at wavelength λ and concentration C, respectively. Subscripts represent the contribution of pure water (w), chl, DOM, and sm. An asterisk (*) denotes the specific value (value per unit concentration) of the absorption or scattering coefficients.

The values of $a_w(\lambda)$ and $b_w(\lambda)$ for water are considered known and were taken from Pope and Fry (1997) for the wavelength range 400-700 nm. Values from Smith and Baker (1981) were used for other wavelengths. The absorption coefficient of chl, $a_{chl}(\lambda)$, was calculated by the equations given by Devred et al. (2006):

$$a_{chl}(\lambda) = a^*_1(\lambda)C_1 + a^*_2(\lambda)(C_{chl1} - C_1) \qquad (5)$$

where $a^*_1(\lambda)$ and $a^*_2(\lambda)$ are the specific absorption coefficients of the two populations of chl adopted from table 3 of Devred et al. (2006). In equation 3, C_1 was

calculated by adopting average values reported in tables 2 and 3 of Devred et al. (2006) as:

$$C_1 = 0.62[1 - \exp(-1.61 \cdot C_{chl}]$$ (6)

The absorption coefficient of sm, $a_{sm}(\lambda)$, was calculated as an exponentially decreasing function with increasing wavelength as:

$$a_{sm}(\lambda) = X[1] (C_{sm})^{X[2]} \exp(-X[3]\lambda)$$ (7)

where $X[1]$, $X[2]$, and $X[3]$ are the parameters to be optimized during the inverse modeling.

The absorption coefficient of DOM, $a_{DOM}(\lambda)$, was calculated as a generally accepted exponentially decreasing function with wavelength (Mobley, 1994):

$$a_{DOM}(\lambda) = X[4] C_{DOM} \exp(-X[5](\lambda - 400))$$ (8)

where $X[4]$ and $X[5]$ are the parameters to be optimized during the inverse modeling.

The scattering coefficient of chl, $b_{chl}(\lambda)$, was calculated by the equation proposed by Morel and Maritorena (2001) with multiplication of the parameter $X[6]$ to the equation to adjust it for the local Beaver Reservoir phytoplankton optical characteristics as:

$$b_{chl}(\lambda) = X[6](C_{chl})^{0.766}\{0.002 + 0.01$$
$$\cdot [0.50 - 0.25 \log_{10}(C_{chl})][(\lambda / 550)^v]\}$$ (9)

The varying exponent v was defined as:

$$v = (1/2)(\log_{10}(C_{chl}) - 0.3), \quad 0.02 < C_{chl} < 2 \text{ mg m}^{-3}$$
$$v = 0, \quad\quad\quad\quad\quad\quad C_{chl} > 2 \text{ mg m}^{-3}$$ (10)

The scattering coefficient of suspended matter, $b_{sm}(\lambda)$, was calculated assuming a weaker dependence of the wavelength on the scattering coefficient as (Aleshin 2001):

$$b_{sm}(\lambda) = X[7] C_{sm} [(550 / \lambda)]^{0.3}$$ (11)

where $X[7]$ was the parameter to be optimized during the inverse modeling.

DESCRIPTION OF PARAMETER ESTIMATION USING INVERSE MODELING

Inverse modeling was used to optimize the value of the seven parameters ($X[1]$ to $X[7]$) of the IOP equations (eqs. 3 to 11). Inverse modeling requires known reflectance values for a set of different concentrations of the OAC and different wavelength of electromagnetic radiation. In this study, 384 measured data points collected at 24 different locations in the Beaver Reservoir representing known concentrations of chl, sm, and DOM and their reflectance characteristics at 16 different wavelengths (450, 480, 550, 570, 580, 590, 600, 610, 630, 650, 670, 690, 710, 730, 750, and 800 nm) were used.

In the inverse modeling, a set of possible values of $X[1]$ to $X[7]$ was randomly selected. This enabled the IOP values determination of OAC for all the 384 cases of different wavelengths and sampling locations calculated using equations 3 to 11. Based on these calculated IOP values, 384 reflectances were predicted using the PHO-MC model.

Predicted reflectances were compared with the corresponding observed reflectances to quantify the mean sum of squares error (MSE) as:

$$\text{MSE} = \frac{1}{N \cdot P} \sum_{n=1}^{N} \sum_{p=1}^{P} [R_M(\lambda_n) - R_S(\lambda_n)]^2$$ (12)

where $R_M(\lambda_n)$, and $R_S(\lambda_n)$ are, respectively, the measured and the PHO-MC model-predicted remote sensing reflectances at wavelength λ_n and sampling location p. The value of N was 16 and P was 24 in this study. The objective of the optimization was to obtain a set of $X[1]$ to $X[7]$ values that minimized the MSE.

Predicting 384 reflectances from the PHO-MC model to calculate MSE values, which evaluated one set of IOP parameters $X[1]$ to $X[7]$, required approximately 2.4 h on a desktop computer (Intel Pentium 4, 2.8 GHz). Searching the entire parameter space of all seven parameters ($X[1]$ to $X[7]$) may require 10^{14} calculations of MSE (assuming only 100 different realizations of each parameter) and may need approximately 10^{10} years of computational time on a single desktop computer. Two steps were taken to reduce this computational burden (fig. 3). One was to reduce the running time of the PHO-MC model by supplementing it with a pseudo simulator model. The pseudo simulator model was developed based on data set generated by running the actual PHO-MC model to calculate reflectances for a fixed set of absorption and scattering coefficients. This set was then used to develop a pseudo simulator consisting of a relationship between reflectance and values of absorption and scattering coefficients using ANN. The pseudo simulator model was used as a supplement to the PHO-MC model to interpolate PHO-MC model values for a given set of a and b values. This considerably reduced the computation time required by the pseudo simulator to calculate reflectance for any combination of absorption and scattering coefficient values.

The computation time needed to optimize IOP values was further reduced by using an optimization technique. An optimization technique requires a lesser number of searches for getting optimum values of the set of $X[1]$ to $X[7]$ compared to searching the entire possible parameter space of $X[1]$ to $X[7]$. A genetic algorithm (GA) was used as an optimization algorithm. The optimized parameter values of $X[1]$ to $X[7]$ of the IOP equations were verified using an independent data set consisting of 384 measured reflectances from the Beaver Reservoir (16 wavelengths at 24 sampling locations).

DESCRIPTION OF THE ANN-BASED PSEUDO SIMULATOR

A pseudo simulator, developed using an ANN, was used due to its ability in interpolating nonlinear functions. Through learning procedures, ANNs have the ability to approximate any nonlinear relationship that exists between a set of independent variables as input and their corresponding set of dependent variables as outputs (Lacroix et al., 1997). Research evidence regarding the superiority of the ANN technique for nonlinear data modeling has been provided by several researchers (Zhuang and Engel, 1990; Ranaweera et al., 1995; Panda and Panigrahi, 2000). An ANN attempts to mimic human mental and neural structures and functions (Hsieh, 1993) to develop a relationship
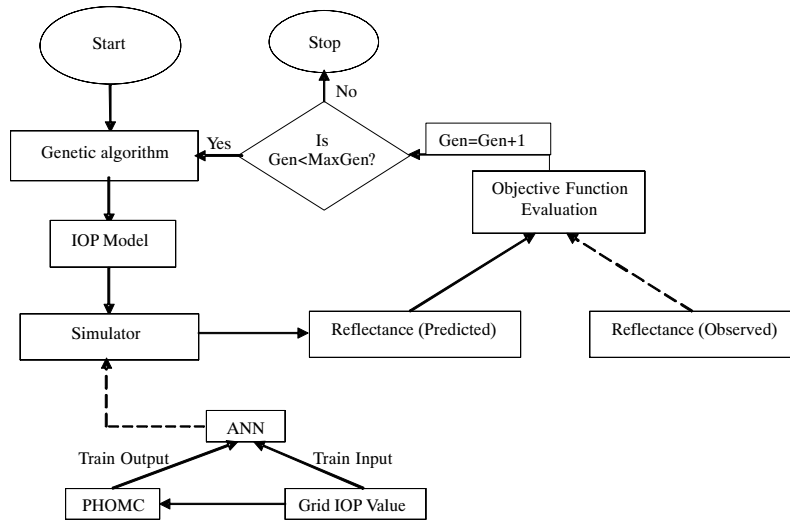
Figure 3. Flowchart of inverse modeling to determine the IOPs using ANN as a pseudo-simulator and GA as an optimizer.

between dependent and independent variables. This approach is contrary to the statistical regression modeling approach in which estimates of the coefficients of independent variables and constants of a mathematical equation are used to predict the dependent variable. In an ANN, the network topology consists of a set of nodes (neurons) as the input layer which are equal to the number of independent variables, one or more intermediate layers with hidden neurons, and an output layer consisting of one or more neurons depending upon the number of dependent variables. Each node in a layer receives and processes weighted input from the previous layer and transmits its output to the nodes in the following layer through links. Each link is assigned a weight, which is a numerical estimate of the connection strength. The weighted summation of inputs to a node is converted to an output according to a transfer function.

The ANN-based simulator model was developed based on two parameters, $a(\lambda)$ and $b(\lambda)$, as the independent variables and the PHO-MC model-predicted reflectance as the dependent variable. A total of 1,000 combinations of $a(\lambda)$ selected randomly from the range 4.5×10^{-6} to 3.5×10^{-2} m^{-1} and $b(\lambda)$ values selected randomly from 3.0×10^{-6} to 6.0×10^{-3} m^{-1} provided grid points for which the PHO-MC model was run to calculate the reflectance. A three-layer feedforward network (fig. 4) was trained using the ANN toolbox of MATLAB (ver. 2000, The MathWorks, Natick, Mass.) to obtain the weights and biases of each network. A sigmoid transfer function for five hidden neurons (optimized by trial and error) and one output neuron was used (eq. 13). A sigmoid function uses values between zero and one; therefore, input and output data sets were standardized by scaling the values between zero and one:

$$f(x) = \frac{1}{1 + e^{-x}} \qquad (13)$$

The supervised training was accomplished with the help of a back-propagation algorithm (Levenberg-Marquardt) as implemented in MATLAB. Twenty-one parameters of a trained neural network were used to calculate the values of output reflectance based on given input values of IOP as:

$$\overline{\mathbf{O}} = f\left(\left(\left(\sum_{i=1}^{5}\left[f\left(\left[a \cdot WA(i) + b \cdot WB(i)\right] + bH(i)\right)\right]\right.\right.\right.$$
$$\left.\left.\left. \cdot WH(i)\right) + bO(1)\right)\right) \qquad (14)$$

where O is the output reflectance for the IOP values a and b, WA and WH represent the weights connecting inputs to the hidden layer and hidden layer to output layer of neurons, and bH and bO represent the bias of the hidden layer and output layer neurons. Refer to figure 4 for the nomenclature of weights and bias values. Effectiveness of the simulator model was tested by comparing the ANN-predicted reflectances at 384 different combinations with the PHO-MC model-predicted reflectances.

DESCRIPTION OF THE OPTIMIZATION ALGORITHM

The Nondominated Sorted Genetic Algorithm II (NSGA II) (Deb, 2001; Deb et al., 2002) was used for optimization in this study. A genetic algorithm (GA) is a heuristic search method based on the theory of evolution that uses the concept of evaluation of the objective functions where the variables undergo mutation and crossover of the population (entities) in each generation until the global optima are reached. A GA consists of a population of chromosomes (solutions) with variables coded in the form of genes. The initial population of chromosomes is randomly generated for a given population size. During the selection process at each successive generation (iteration), the existing solutions are picked and/or duplicated, using probabilistic methods, based on fitness of the individuals: the greater the fitness of an individual, the larger the chance of it being selected into the mating pool. The individuals in the mating pool then undergo genetic operations (crossover and mutation). In this study, we used a roulette wheel based selection technique. Crossover, also called recombination or reproduction, produces child solutions from the parent solutions present in the mating pool. Crossover is necessary to generate a population for the next generation that shares many of the positive characteristics of the parent. During mutation, a bit in the chromosome sequence of the population
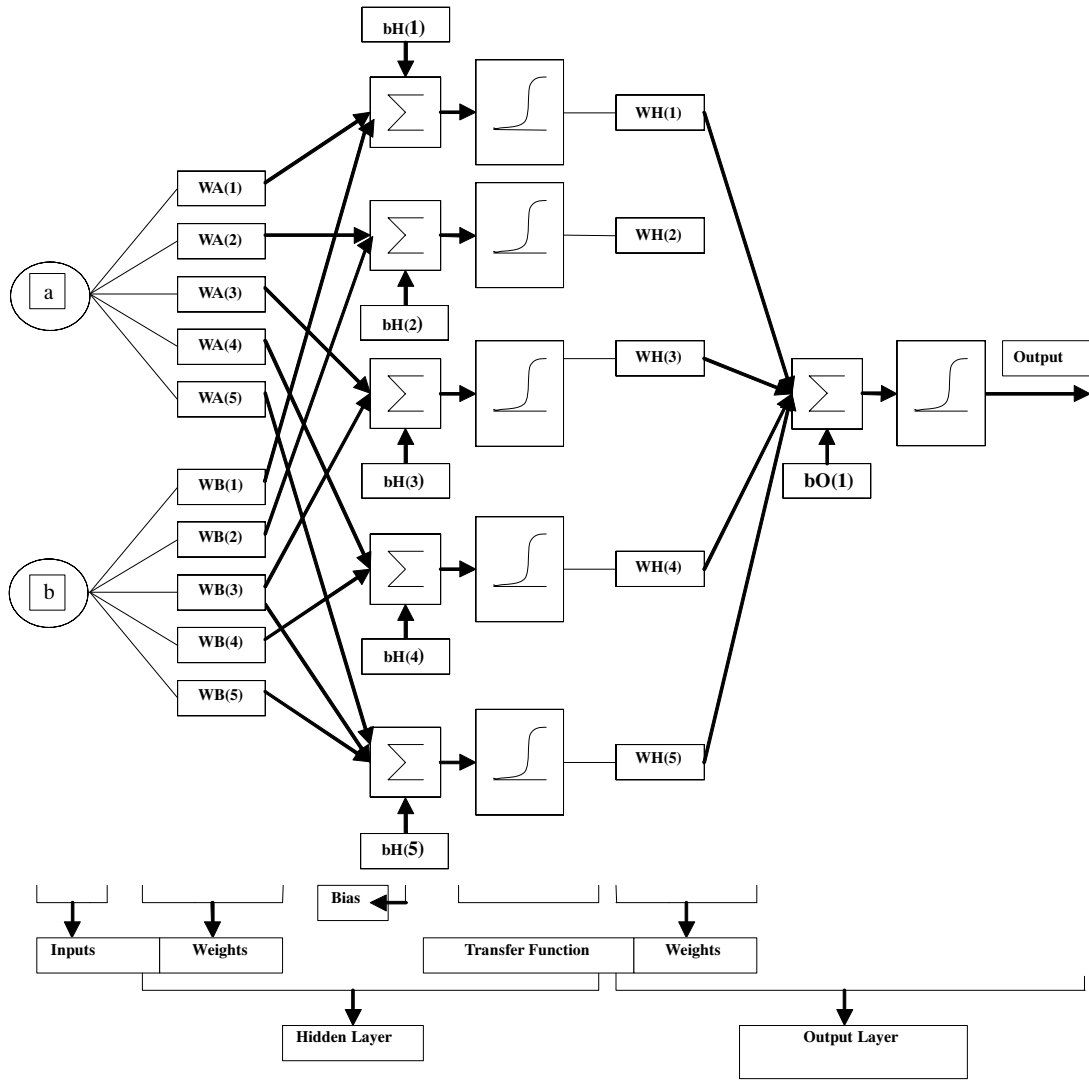
Figure 4. Artificial neural network (feedforward) structure used in this study as a pseudo-simulator: *a* and *b* are the inputs, *WA* and *WH* are the weights connecting the inputs to hidden layer and hidden layer to the output layer, respectively, and *bH* and *bO* represent the biases of the hidden layer and output layer, respectively.

is selected randomly and is altered from its original state. Mutation is used to maintain genetic diversity from one generation of solutions to the next. Goldberg (1989) introduced the "mutation clock" operator to identify the next bit to be mutated by skipping $\eta = -p_m \ln(1 - r)$ bits from the present bit for any random number (r) and mutation probability (p_m), thereby reducing the number of random numbers to be generated by $O(1/p_m)$.

The objective of multi-objective optimization is to search for solutions in the global Pareto-optimal region (i.e., optimal for all the objective functions) and to achieve solutions that are separate from one another to the maximum possible extent in the nondominant front. In a multi-objective optimization problem, if g_i, $\{i = 1, ..., M\}$ are the objective functions that need to be minimized, a solution $x^{(1)}$ is said to dominate $x^{(2)}$ if both the following conditions are true (Zitzler et al., 2000):

$$\forall i \in \{1,....M\} : g_i\left(x^{(1)}\right) \le g_i\left(x^{(2)}\right) \wedge$$

$$\exists j \in \{1,....M\} : g_j\left(x^{(1)}\right) < g_j\left(x^{(2)}\right) \qquad (15)$$

That is, $x^{(2)}$ is dominated by $x^{(1)}$, or in other words $x^{(1)}$ is nondominated by $x^{(2)}$. If each individual in a population of size N has solutions that are nondominated, then the representative of the solutions in the objective space determines the Pareto-optimal front. This also helps in checking the premature convergence of the optimization process (Deb et al., 2002).

There always exists a set of best solutions at each generation, which can be comparable to the population size N that can go along to the next generation. Such solutions that are nondominated among all the individual generations are called elite solutions and are stored in an external set called the elite set. After every generation, a percentage of the population is replaced by individuals from the elite set. The NSGA-II uses crowding distance to ensure that the solutions generated at each generation are well spread along the Pareto-optimal front and are far apart in the solution space. The crowding distance is defined as the sum of the side lengths of the cuboid that touches the neighboring solutions in case of the non-extreme solutions and is infinite for the extreme solutions.

GAs performs well in large and complex search spaces, and when properly implemented, a GA is capable of both exploration (broad search) and exploitation (local search) of the search space (Goldberg, 1989). In addition, probabilistic transition rules, instead of deterministic rules, decide the optimal solutions in GA. The most important feature of GAs is their robust nature and the balance between efficiency and efficacy necessary for survival in many different envi-ronments. Therefore, a GA-based optimization model can help in intelligently selecting the next set of parameters of the IOP equations based on the results of the previous set of parameter performances, instead of randomly selecting their values for successive performance evaluations.

RESULTS AND DISCUSSION

Water quality in the Beaver Reservoir varied temporally from 2003 to 2005 and spatially among sampling locations. Of the 48 sampling locations, measured data from randomly selected 24 sampling locations were used in inverse modeling (calibration) and data from the remaining 24 sampling locations were used to validate the inverse modeling results. Table 1 summarizes the water quality variation of the calibration and validation groups. Means of the OAC concentrations were not statistically different between the calibration and validation data sets ($p < 0.05$). Reflectance variability of all 48 sampling locations ranged from 0.12% to 10.01% in the wavelength range from 400 to 800 nm (fig. 5).

PSEUDO SIMULATOR MODEL PERFORMANCE

The reflectance values predicted by the PHO-MC model followed the general trend of being directly proportional to $b(\lambda)$ and inversely proportional to $a(\lambda)$ (fig. 6). For the same $a(\lambda)$ and $b(\lambda)$ values, the PHO-MC and pseudo simulator models gave similar reflectance values. The coefficient of determination (R^2) was calculated between the pseudo simulator and the PHO-MC model-predicted reflectance values. For the calibration data set, R^2 was greater than 0.99 for 1,000 different combinations of $a(\lambda)$ and $b(\lambda)$. Similarly for the validation data set, R^2 was greater than 0.99 for 384 different combinations of $a(\lambda)$ and $b(\lambda)$, indicating that the
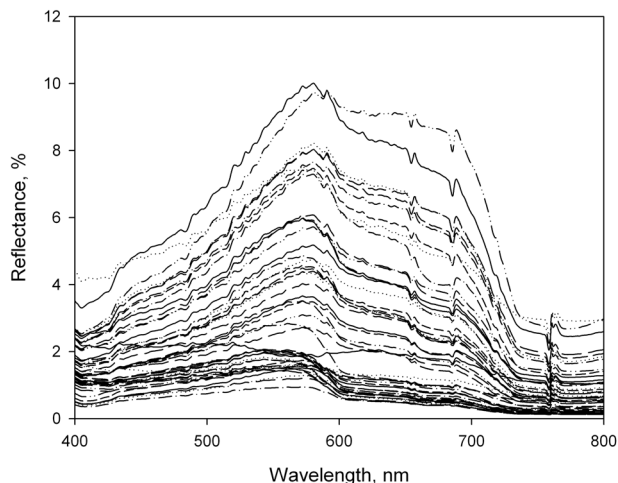

Figure 5. Reflectance (R) trend observed for 48 sampling locations in the Beaver Reservoir, Arkansas.

pseudo simulator accurately mimicked the behavior of the PHO-MC model.

The PHO-MC model run time for a set of $a(\lambda)$ and $b(\lambda)$ combination varied from 37 to 145 s with an average value of about 50 s. Based on the average model run time, it would take approximately 5.3 h if the actual PHO-MC model was run to simulate 384 reflectances to check error for one set of parameters for IOP determination. In the case of an enumerative scheme, searching the entire parameter space of all seven parameters ($X[1]$ to $X[7]$) may require 10^{14} calculations of MSE (assuming only 100 different values for each parameter), which may approximately 10^{10} years of computational time on a single desktop computer. Since the optimization process by GA involved 20,000 sets (100 populations × 200 generation) during its search for the global optima of the parameters, the required computer time would be 106,000 h (>12 years). However, due to use of an ANN as
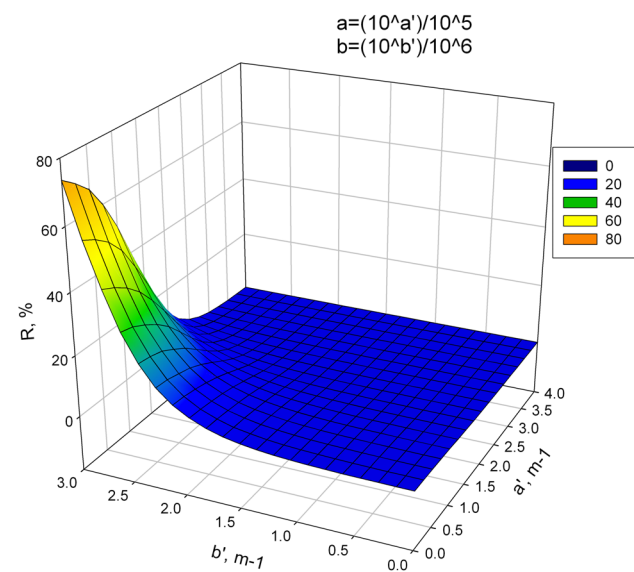

Figure 6. Reflectance (R) trend predicted by PHO-MC model depending upon total absorption coefficient, $a(\lambda)$, and scattering coefficient, $b(\lambda)$, of the OAC of water and water itself. To plot the graph, $a(\lambda)$ was transformed to a' (= 0. 5 + $\log_{10}a(\lambda)$) and $b(\lambda)$ was transformed to b' (= 0. 5 + $\log_{10}b(\lambda)$).

<center>

Table 1. Statistical data of water quality parameters of the Beaver Reservoir.[a]

</center>

Statistical Parameter	Concentration		
	sm (g m^{-3})	chl (mg m^{-3})	DOM (g m^{-3})
For all 48 sampling locations			
Range	0.15-16.30	0.45-12.60	1.2-6.2
Mean	4.05	3.87	2.3
CV	88.2	68.3	28.2
For 24 locations used for calibration			
Range	0.15-16.30	0.45-11.28	1.40-6.20
Mean	4.18	4.13	2.43
CV	89.2	60.2	38.2
For 24 locations used for validation			
Range	0.22-11.70	0.56-12.60	1.20-2.70
Mean	3.92	3.61	2.21
CV	86.9	76.8	16.7

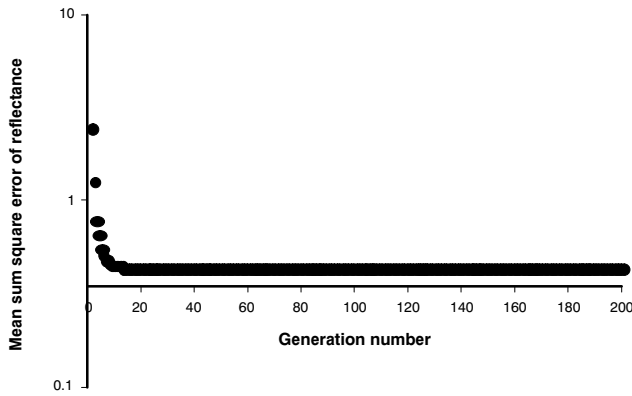[a] sm = suspended matter, chl = chlorophyll-a, DOM = dissolved organic matter, and CV = coefficient of variation.

Figure 7. Mean square error (MSE) of predicted vs. measured reflectances (on a log scale) as the number of generation proceeds in genetic algorithm optimization during inverse modeling of the PHO-MC to optimize parameter values of IOP equations.

Figure 8. Mass-specific backscattering coefficient of suspended matter (b_{sm}/2) of Beaver Reservoir water (estimated using inverse modeling approach).

pseudo simulator model and the GA to supplement the PHO-MC model, the total run time was reduced to less than 3 min.

GENETIC ALGORITHM PERFORMANCE

Error propagation of the objective function with the generation is depicted in figure 7. Errors decreased rapidly in the first 50 generations and then decreased slowly until becoming constant after about 100 generations. Hence, it can be said that the parameters of the IOP equations identified in successive generations were, on an average, better (smaller mean square error) than the previous generations for the first 100 generations, beyond which no significant improvements were observed. This showed that the GA was capable of learning from the previous parameter values and selected a better set of parameters in the successive generation until it reached global optima.

IOP PARAMETER VALUES

The optimized values of the seven parameters of IOP equations (X[1] to X[7]) were 21.48962, 0.75000, 0.01067, 0.08000, 0.00300, 0.12659, and 0.02500, respectively. In a laboratory study, Babin et al. (2003) reported average parameter values of X[1], X[2], and X[3] for non-algal particle absorption as 5.40, 1.0, and 0.0123, respectively. In the same study, the parameter X[4] value was approximately 0.05 (fig. 3 of Babin et al., 2003). Similarly, Arst (2003) reported that parameter X[5] ranged from 0.004 to 0.53 for oceanic water and from 0.013 to 0.020 for coastal and lake waters based on review of several studies. The inverse modeling approach estimated value for X[5], i.e., 0.003, was found to be lower than the lower range value for oceanic waters reported by Arst (2003). The scattering coefficient model of chl by Morel and Maritorena (2001), which was adopted in this study, had a constant value of 0.416 for X[6]. The optimized value of X[6] from the inverse modeling approach was about 70% lower for Beaver Reservoir water. It showed that Beaver Reservoir water phytoplankton had lower scattering coefficients than the generally reported phytoplankton scattering coefficients. Suspended mineral particle backscattering information before the Wozniak and Stramski (2004) studies is scanty and was not available in the literature to compare the estimated values. Optimized values of the suspended matter backscattering coefficient ($b_{sm}(\lambda)/2$) from the inverse modeling approach of this study (fig. 8) were
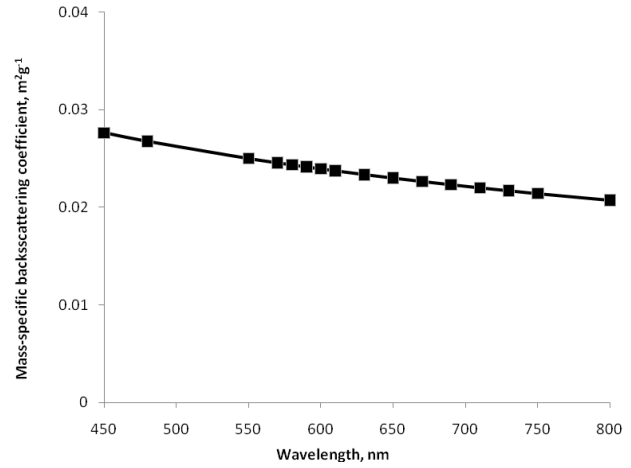
compared with the theoretically calculated suspended mineral backscattering coefficient values of Wozniak and Stramski (2004) and were found to be similar. These similarities of IOP parameter values with the literature values indicate merit of the proposed inverse modeling approach.

Based on estimated parameters of IOP, the mass-specific absorption coefficients for chl, sm, and DOM were calculated (fig. 9). In the 450 to 550 nm range, $a^*_{sm}(\lambda)$ was the dominant parameter affecting reflectance. For wavelengths greater than 550 nm, a_w was dominant. Figure 10 shows the mass-specific scattering coefficients for chl, sm, and DOM. The scattering coefficients of water, chl, sm, and DOM decreased with increasing wavelength. The mass-specific scattering coefficient was greatest for sm at all wavelengths.

Relative sensitivity analysis (S_r) of each parameter was calculated as follows:

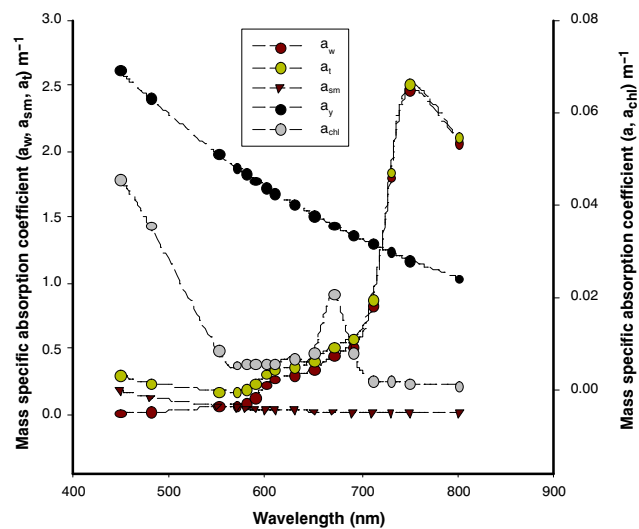$$S_r = \frac{P}{O} \frac{O_2 - O_1}{P_2 - P_1} \tag{16}$$

Figure 9. Estimated mass-specific absorption coefficient from inverse modeling technique of OAC (suspended matter, a_{sm}, and dissolved organic carbon, a_{DOM}, of Beaver Reservoir water) as well as adopted values of water (a_w) and chl (a_{chl}) absorption coefficients.
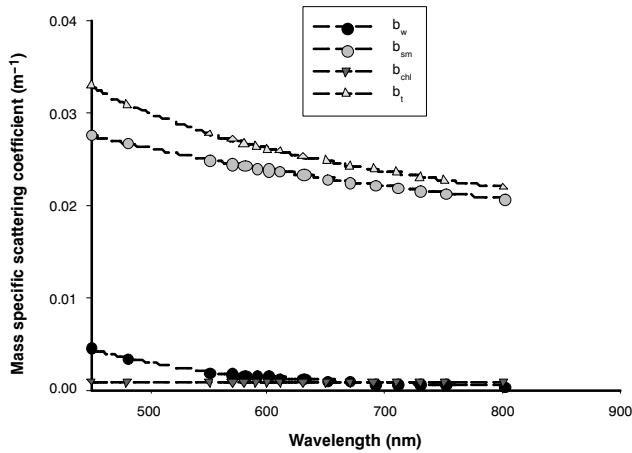
Figure 10. Mass-specific scattering coefficients of Beaver Reservoir water and its OAC estimated using inverse modeling approach.

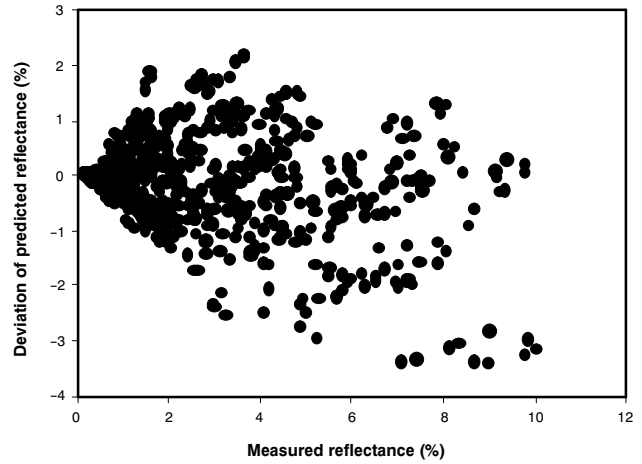**Figure 11. Deviation of PHO-MC model-predicted reflectance from the measured reflectance for the calibration and validation data sets.**

where P and O are the base values of input and output. The base value of each parameter was changed by $\pm 5\%$ from its optimum value while keeping the other six parameters constant at their optimum value to get P_1 and P_2 and the corresponding O_1 and O_2. The outputs considered were MSE and the average value of predicted reflectance (table 2). The relative sensitivity of reflectance was measured at 580 nm wavelength at one of the sampling data sets of the inverse modeling (C_{sm} = 16.3 g m^{-2}, C_{chl} = 1.05 mg m^{-2}, and C_{DOM} = 2.3 g m^{-2}). Reflectance error was measured at 580 nm wavelength, where reflectance was generally the greatest.

The most sensitive parameter was $X[3]$ followed by $X[2]$ (table 2). Two highly sensitive parameters belong to equation 7, which was used for determining the value of $a_{sm}(\lambda)$, indicating that the determination of absorption coefficient values of suspended matter was critical. Parameter $X[6]$ was not a very sensitive parameter, indicating that the scattering coefficient of chl was not a significant component of $b(\lambda)$.

The optimized values of the IOP parameters ($X[1]$ to $X[7]$) were validated by using the measured data from 24 sampling locations of the Beaver Reservoir. The IOP values were calculated for 16 different wavelengths at these locations. The IOP values thus determined were subsequently used in the pseudo simulator model to predict reflectances at these locations. A total of 384 reflectances were predicted (24 sampling locations × 16 wavelengths). Predicted reflectance values were compared with the corresponding measured reflectances. The MSE between predicted and measured reflectances for the validation data set was 1.03,

which was slightly greater than the MSE of 0.43 obtained for the inverse modeling data set. Figure 11 depicts the deviation of the predicted reflectances from the measured reflectances for the training and validation sets. Deviation was found to be less for the relatively smaller values of the measured reflectance. Figure 12 depicts the regression between the observed and predicted reflectances for the calibration and validation data sets. A highly significant Pearson product-moment correlation coefficient (r = 0.90, p < 0.01) was found

Table 2. Relative sensitivity of the parameters used to estimate IOP values at a wavelength and concentrations of suspended matter, chlorophyll-a, and dissolved organic matter.

Parameter	Relative Sensitivity for the Output	
	Reflectance	Sum of Error
$X[1]$	-0.557	0.0030
$X[2]$	-1.165	0.2530
$X[3]$	3.406	-0.5322
$X[4]$	-0.167	-0.0508
$X[5]$	0.090	-0.0048
$X[6]$	0.001	-0.0003
$X[7]$	0.608	-0.7257

(a)

$y = 0.9384x + 0.0486$
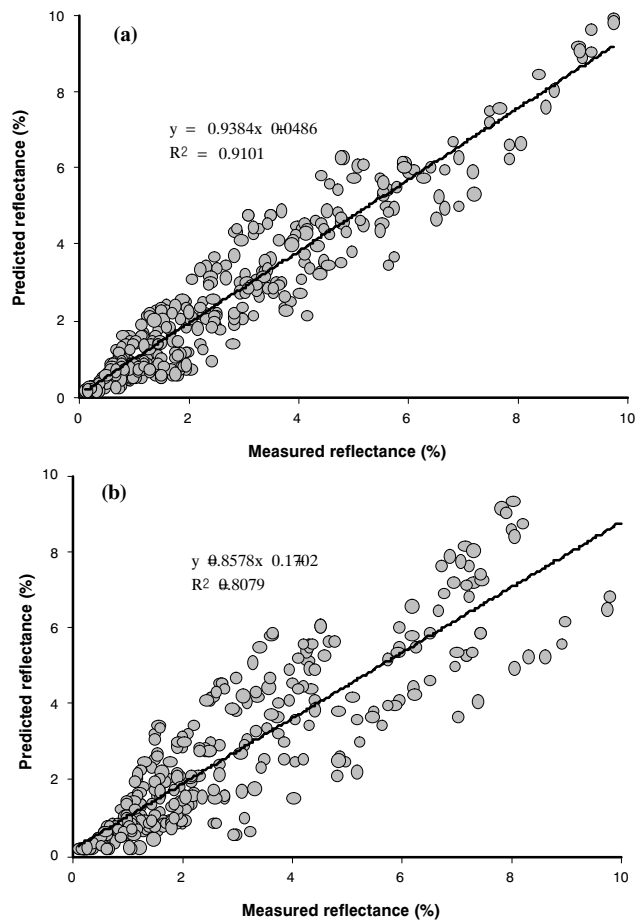$R^2 = 0.9101$

(b)

$y = 0.8578x + 0.1702$
$R^2 = 0.8079$

Figure 12. Scatter plot of measured and PHO-MC model-predicted reflectances for (a) the inverse modeling data set, and (b) the validation data set.

between predicted reflectance values and corresponding measured reflectance values for the validation data set. These results show that the proposed inverse modeling technique can be used to assess the IOP values of multiple optically active constituents of a water body. Successful inverse and forward application of the PHO-MC model showed its accuracy for remote sensing applications.

SUMMARY AND CONCLUSIONS

An inverse modeling approach was presented to calculate the inherent optical properties of the OAC of Beaver Reservoir water using a physical hyperspectral optical - Monte Carlo (PHO-MC) model. This approach is different from previous modeling approaches, which used simple semi-analytical models only. The PHO-MC model can account for different water depths, bottom reflectances, sun angles, and lighting conditions. The long computation time required to run the PHO-MC model was reduced substantially, from years to less than 3 min, by supplementing the PHO-MC model with an artificial neural network (ANN) based pseudo simulator and Nondominated Sorted Genetic Algorithm II (NSGA II) as an optimization algorithm. Pseudo simulators showed a high coefficient of determination ($R^2 >$ 0.99) with the PHO-MC model-predicted reflectance.

The IOP variation of chl, sm, and DOM were estimated with seven parameters. Mean sum of squares error between the PHO-MC model-predicted reflectance and measured reflectance was relatively small (0.43) for 384 reflectances. The IOP values were further validated using an independent data set from 24 sampling locations. A significant correlation ($r = 0.90$, $p < 0.01$) was found between predicted reflectance values and corresponding measured reflectance values for the validation dataset. In addition, no statistically significant difference was found between correlation coefficients for the calibration and validation data sets. The results indicated that the ANN-based pseudo simulators and NSGA II can be used to efficiently determine site-specific IOP values for the PHO-MC model. The approach developed should also be applicable to other remote sensing models.

REFERENCES

Aleshin, I. V. 2001. Optical methods and facilities for expeditious monitoring of the ecological status of sea water. *J. Opt. Tech.* 68(4): 261-268.

APHA. 1998. Standard method for the examination of water and wastewater. Washington, D.C.: American Public Health Association.

Arst, K. I. 2003. *Optical Properties and Remote Sensing of Multicomponental Water Bodies.* Berlin, Germany: Springer-Verlag.

Babin, M., and D. Stramski. 2002. Light absorption by aquatic particles in the near-infrared spectral region. *Limnol. and Oceanog.* 47(3): 911-915.

Babin, M., and D. Stramski. 2004. Variations in the mass-specific absorption coefficient of mineral particles suspended in water. *Limnol. and Oceanog.* 49(3): 756-767.

Babin, M., D. Stramski, G. M. Ferrari, H. Claustre, A. Bricaud, G. Obolensky, and N. Hoepffner. 2003. Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe. *J. Geophys. Res. Oceans* 108(C7): 4.1-4.20.

Barnard, A. H., J. R. V. Zaneveld, and W. S. Pegau. 1999. In situ determination of the remotely sensed reflectance and the absorption coefficient: Closure and inversion. *Applied Optics* 38(24): 5108-5117.

Ciotti, A. M., M. R. Lewis, and J. J. Cullen. 2002. Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient. *Limnol. and Oceanog.* 47(2): 404-417.

Cota, G. F., W. G. Harrison, T. Platt, S. Sathyendranath, and V. Stuart. 2003. Bio-optical properties of the Labrador Sea. *J. Geophys. Res. Oceans* 108(C7): 21.1-21.14.

Deb, K. 2001. *Multi-Objective Optimization Using Evolutionary Algorithms.* Hoboken, N. J.: John Wiley.

Deb, K., A. Pratap, S. Agarwal, and T. Meyarivan. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comp.* 6(2): 182-197.

Dekker, A. G., and R. P. Bukata. 2002. Chapter 23: Remote sensing of inland and coastal waters. In *Review of Radio Science: 1999-2002,* 977-519-534. W. R. Stone, ed. New York, N.Y.: John Wiley and Sons.

Devred, E., S. Sathyendranath, V. Stuart, H. Maass, O. Ulloa, and T. Platt. 2006. A two-component model of phytoplankton absorption in the open ocean: Theory and applications. *J. Geophys. Res. Oceans* 111: C03011, doi: 10.1029/ 2005JC002880.

Doyle, J. P., and H. Rief. 1998. Photon transport in three-dimensional structures treated by random walk techniques: Monte Carlo benchmark of ocean colour simulations. *Math. and Comp. in Simulation* 47(2-5): 215-241.

Garg, V. 2006. Development and evaluation of a physical hyperspectral optical-Monte Carlo model for an aquatic medium reflectance simulation. PhD diss. University of Arkansas.

Garg, V., I. Chaubey, and S. Singh. 2009. Evaluation of a hyperspectral optical-Monte Carlo remote sensing model in a tank study. *Trans. ASABE* 52(3): 759-769.

Goldberg, D. E. 1989. *Genetic Algorithm in Searching, Optimization, and Machine Learning.* Reading, Mass.: Addison-Wesley-Longman.

Gordon, H. R., and G. C. Boynton. 1997. Radiance-irradiance inversion algorithm for estimating the absorption and backscattering coefficients of natural waters: Homogeneous waters. *Applied Optics* 36(12): 2636-2641.

Hamre, B., O. Frette, S. R. Erga, J. J. Stamnes, and K. Stamnes. 2003. Parameterization and analysis of the optical absorption and scattering coefficients in a western Norwegian fjord: A case II water study. *Applied Optics* 42(6): 883-892.

Hoepffner, N., and S. Sathyendranath. 1993. Determination of the major groups of phytoplankton pigments from the absorption spectra of total particulate matter. *J. Geophys. Res. Oceans* 98(C12): 22789-22803.

Hsieh, C. 1993. Some potential applications of artificial neural networks in financial management. *J. Systems Mgmt.* 44(4): 12-15.

Lacroix, R., F. Salehi, X. Z. Yang, and K. M. Wade. 1997. Effects of data preprocessing on the performance of artificial neural networks for dairy yield prediction and cow culling classifications. *Trans. ASAE* 40(3): 839-846.

Lee, Z. P., and K. L. Carder. 2004. Absorption spectrum of phytoplankton pigments derived from hyperspectral remote-sensing reflectance. *Remote Sensing of Environ.* 89(3): 361-368.

Lee, Z. P., K. L. Carder, and R. A. Arnone. 2002. Deriving inherent optical properties from water color: A multiband quasi-analytical algorithm for optically deep waters. *Applied Optics* 41(27): 5755-5772.

Loisel, H., and D. Stramski. 2000. Estimation of the inherent optical properties of natural waters from the irradiance attenuation coefficient and reflectance in the presence of Raman scattering. *Applied Optics* 39(18): 3001-3011.

Loisel, H., D. Stramski, B. G. Mitchell, F. Fell, V. Fournier-Sicre, B. Lemasle, and M. Babin. 2001. Comparison of the ocean inherent optical properties obtained from measurements and inverse modeling. *Applied Optics* 40(15): 2384-2397.

McKee, D., A. Cunningham, and S. Craig. 2003. Estimation of absorption and backscattering coefficients from in situ radiometric measurements: Theory and validation in Case II waters. *Applied Optics* 42(15): 2804-2810.

Mobley, C. D. 1994. *Light and Water: Radiative Transfer in Natural Waters*. San Diego, Cal.: Academic Press.

Mobley, C. D., B. Gentili, H. R. Gordon, Z. H. Jin, G. W. Kattawar, A. Morel, P. Reinersman, K. Stamnes, and R. H. Stavn. 1993. Comparison of numerical models for computing underwater light fields. *Applied Optics* 32(36): 7484-7504.

Morel, A., and S. Maritorena. 2001. Bio-optical properties of oceanic waters: A reappraisal. *J. Geophys. Res. Oceans* 106(C4): 7163-7180.

Panda, S. S., and S. Panigrahi. 2000. Analysis of remotely sensed serial images for precision farming. ASAE Paper No. 003055. St. Joseph, Mich.: ASAE.

Pope, R. M., and E. S. Fry. 1997. Absorption spectrum (380-700 nm) of pure water: II. Integrating cavity measurements. *Applied Optics* 36(33): 8710-8723.

Ranaweera, D. K., N. F. Hubele, and A. D. Papalexopoulos. 1995. Application of radial basis function neural network model for short-term load forecasting. *IEEE Proc. Generation, Transmission and Distribution* 142(1): 45-50.

Risovic, D. 2002. Effect of suspended particulate-size distribution on the backscattering ratio in the remote sensing of seawater. *Applied Optics* 41(33): 7092-7101.

Sathyendranath, S., L. Lazzara, and L. Prieur. 1987. Variations in the spectral values of specific absorption of phytoplankton. *Limnol. and Oceanog*. 32(2): 403-415.

Sathyendranath, S., L. Prieur, and A. Morel. 1989. A 3-component model of ocean color and its application to remote-sensing of phytoplankton pigments in coastal waters. *Intl. J. Remote Sensing* 10(8): 1373-1394.

Sathyendranath, S., G. Cota, V. Struat, H. Maass, and T. Platt. 2001. Remote sensing of phytoplankton pigments: A comparison of empirical and theoretical approaches. *Intl. J. Remote Sensing* 22(2): 249-273.

Smith, R. C., and K. S. Baker. 1981. Optical properties of the clearest natural waters (200-800 nm). *Applied Optics* 20(2): 177-184.

Stramska, M., D. Stramski, B. G. Mitchell, and C. D. Mobley. 2000. Estimation of the absorption and backscattering coefficients from in-water radiometric measurements. *Limnol. and Oceanog*. 45(3): 628-641.

Stramski, D., A. Bricaud, and A. Morel. 2001. Modeling the inherent optical properties of the ocean based on the detailed composition of the planktonic community. *Applied Optics* 40(18): 2929-2945.

Stramski, D., E. Boss, D. Bogucki, and K. J. Voss. 2004a. The role of seawater constituents in light backscattering in the ocean. *Progress in Oceanog*. 61(1): 27-56.

Stramski, D., S. B. Wozniak, and P. J. Flatau. 2004b. Optical properties of Asian mineral dust suspended in seawater. *Limnol. and Oceanog*. 49(3): 749-755.

Wozniak, S. B., and D. Stramski. 2004. Modeling the optical properties of mineral particles suspended in seawater and their influence on ocean reflectance and chlorophyll estimation from remote sensing algorithms. *Applied Optics* 43(17): 3489-3503.

Yacobi, Y. Z., A. Gitelson, and M. Mayo. 1995. Remote-sensing of chlorophyll in Lake Kinneret using high-spectral-resolution radiometer and Landsat TM: Spectral features of reflectance and algorithm development. *J. Plankton Res.* 17(11): 2155-2173.

Zhuang, X., and B. Engel. 1990. Classification of multi-spectral remote sensing data using neural network vs. statistical methods. ASAE Paper No. 907552. St. Joseph, Mich.: ASAE.

Ziztler, E., K. Deb, and L. Thiele. 2000. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutional Computation* 8(2): 173-195.