Queries for apls-62-09-05

This manuscript/text has been typeset from the submitted material. Please check this proof carefully to make sure there have been no font conversion errors or inadvertent formatting errors. Allen Press.

Removal of Surface Reflection from Above-Water Visible–Near Infrared Spectroscopic Measurements

NITIN K. SINGH, SREEKALA G. BAJWA,* and INDRAJEET CHAUBEY

World Wildlife Fund (US), 300 Buck Island Ranch Rd., Lake Placid, Florida 33852 (N.K.S.); Biological and Agricultural Engineering, University of Arkansas, Fayetteville, Arkansas 72701 (S.G.B.); and Agricultural and Biological Engineering, Purdue University, West Lafayette, Indiana 47907 (I.C.)

Water quality estimation in fresh and marine water systems with *in situ* above-water spectroscopy requires measurement of the volume reflectance (ρ_v) of water bodies. However, the above-water radiometric measurements include surface reflection (L_r) as a significant component along with volume reflection. The L_r carries no information on water quality, and hence it is considered as a major source of error in *in situ* above-water spectroscopy. Currently, there are no methods to directly measure L_r. The common method to estimate L_r assumes a constant water surface reflectance (ρ_s) of 2%, and then subtracts the L_r thus calculated from the above-water radiance measurements to obtain the volume reflection (L_v). The problem with this method is that the amount of ρ_s varies with environmental conditions. Therefore, a methodology was developed in this study for direct measurement of water volume reflectance above water at nadir view geometry. Other objectives of this study were to analyze the contribution of L_r to the total water reflectance under various environmental conditions in a controlled setup and to develop an artificial neural network (ANN) model to estimate ρ_s from environmental conditions. The results showed that L_r contributed 20–54% of total upwelling radiance from water at nadir. The ρ_s was highly variable with environmental conditions. Using sun altitude, wind speed, diffuse lighting, and wavelength as inputs, the ANN model was able to accurately simulate ρ_s, with a low root mean square error of 0.003. A sensitivity analysis with the ANN model indicated that sun altitude and diffuse light had the highest influence on ρ_s, contributing to over 82% of predictability of the ANN model. Therefore, the ANN modeling framework can be an accurate tool for estimating surface reflectance in applications that require volume reflectance of water.

Index Headings: **Above-water spectroscopy; Surface reflectance; Volume reflectance; Artificial neural networks; ANNs; Water quality; Visible–near-infrared spectroscopy.**

INTRODUCTION

Optical remote sensing is a fast and efficient method for large-scale monitoring and evaluation of the aquatic environment. Visible–near-infrared (VNIR) spectroscopy from satellite, air, or water-based platforms can be used for water quality estimation, particularly of optically active constituents (OAC) such as sediments and chlorophyll in both marine and fresh water systems.[1,2] The sensors operating from above-water platforms record the total upwelling radiance above the water surface (L_t), which includes three separate components, namely, volume reflection (L_v), bottom reflection (L_b), and surface reflection (L_r). The volume reflection is the only component of the total upwelling radiation from water that carries key information on water quality. The surface reflection refers to light reflected from the air–water interface, and hence, it carries no information on water quality, except for conditions that may affect the surface roughness of the water significantly. The L_b component is negligible for deep water bodies and is usually not an issue for deep parts of lakes and ocean. Since L_r is not useful in water quality estimation, the variation in L_r with environmental conditions can introduce significant errors in water quality estimations.[3–9] Therefore, it is important to remove the L_r from the total upwelling water radiance before using the above-water radiance data for water quality estimations.

There are no direct methods for measuring L_r or L_v from water bodies with sensors operating from above-water platforms. The commonly used indirect method for eliminating L_r from the above-water upwelling radiance measurements uses a correction factor (CF).[4,5,10] The CF is nothing but the water surface reflectance, ρ_s, and is defined as the fraction of sky radiance (L_s) reflected from the air–water interface at any specific zenith angle (Eq. 1).[4,5,11] The L_r can then be calculated as the product of downwelling sky radiation (L_s) and ρ_s (Eq. 2) and subtracted from the total upwelling water radiance to obtain the volume reflection (L_v) of the water body. Since L_v is the total upwelling radiance just under the water surface, and all radiometric systems used for water quality estimations operate from above-water platforms, direct under water measurement of L_v is difficult. But, if CF is known, the volume reflection of water can be estimated indirectly as the difference between total upwelling radiance measured above the water surface and the L_r estimated as the product of CF and downwelling sky radiance using Eq. 3:[5,7]

$$CF = \rho_s = \frac{L_r}{L_s} \tag{1}$$

$$L_r = \rho_s \times L_s \tag{2}$$

$$L_v = L_t - L_r \tag{3}$$

where the variables are as follows:

CF = Correction factor at a specific zenith angle, ratio
ρ_s = Surface reflectance at a specific zenith angle, ratio
L_r = Surface reflection from water at a specific zenith angle, W m^{-2} $sterdian^{-1}$
L_s = Downwelling sky radiation at a specific zenith angle, W m^{-2} $sterdian^{-1}$
L_v = Volume reflection from a water body at a specific zenith angle, W m^{-2} $sterdian^{-1}$
L_t = Total upwelling radiance above the water surface at a specific zenith angle, W m^{-2} $sterdian^{-1}$

Austin[11] proposed a CF of 0.02 for marine water systems at zenith angles of up to 30°, and 0.03 for a zenith angle of 40°. Arst et al.[12] validated that a CF of 0.02 is appropriate for calm

Received 25 February 2008; accepted 5 June 2008.
* Author to whom correspondence should be sent. E-mail: sgbajwa@uark.edu.

0003-7028/08/6209-000000$2.00/0

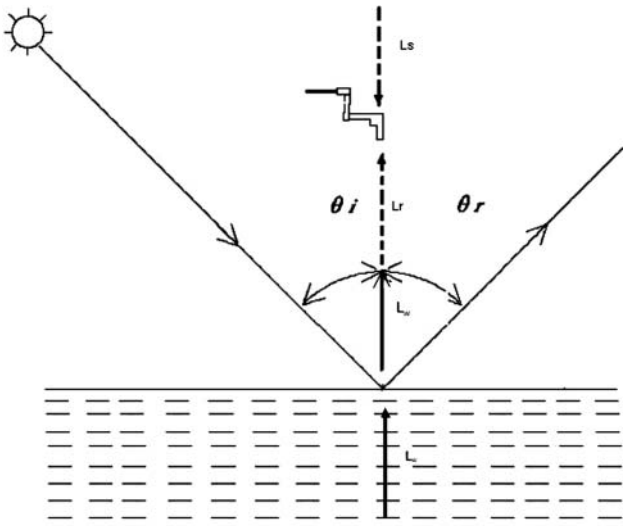

FIG. 1. Generation of surface reflection in above-water radiometric measurements. Light coming at an incidence angle of θ_i reflects from the air–water interface at angle θ_r. Light coming from nadir (L_d) with a zero incidence angle reflects back to nadir from the air–water interface (L_r) and merges with radiance coming from under the water surface (L_v) to generate the total upwelling radiance (L_t).

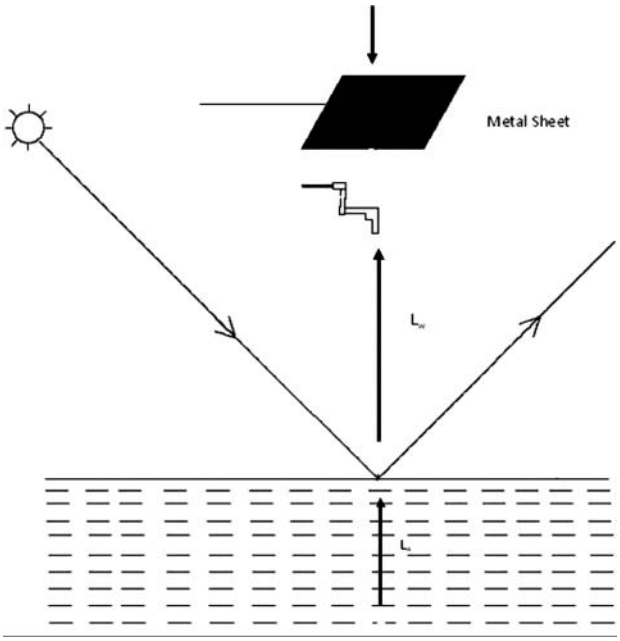

FIG. 2. An approach for removal of surface reflection from upwelling water radiance measured at nadir directly above the water surface by blocking the light coming from nadir.

water conditions and nadir view. Mobley[4] suggested a CF of 0.028 for the visible range of wavelengths at wind speeds less than 5 m/s. However, Mobley[4] also found that many environmental variables such as sun altitude and wind velocity, as well as sensor geometry, affected CF. The CF simulated with a radiative transfer model developed[4] for estimating upwelling water radiance corrected for L_r varied from 0.028 to 0.12 in response to sun altitude and wind speed for a sensor zenith angle of 40° and azimuth angle of 135°.

The constant CF of 0.02 is not very accurate at very high sun altitude or when there is a significant amount of sun glint.[11] In addition to sun glint, wind speed, diffuse light conditions, sun altitude, and viewing geometry of the sensor can influence the L_r as well as the CF.[3,4] Doxaran et al.[5] found that wavelength also affected the L_r, with L_r contributing approximately 50% of the total upwelling water radiance in the visible range and a higher percentage in the near-infrared (NIR) range in turbid water. This study also found that sensor viewing geometry (zenith angle) did not affect L_r, which was contradictory to the findings of Fougnie et al.[3] and Mobley.[4]

There are two basic problems in removing the L_r from upwelling water radiance. The first issue is the lack of an approach for directly measuring volume reflectance, ρ_v of water directly above the water surface. Second is the unavailability of an exact CF at specific wavelengths for different conditions of sun altitude, wind speed, and diffuse light. The CF of 0.02 suggested by Austin[11] and Arst et al.,[12] and 0.028 recommended by Mobley[4] for visible wavelengths, can lead to an over- or under-estimation of water reflectance.[5] Although fresh water systems are relatively less complex than marine systems, there is very little information on the CF and its dynamics with respect to environmental conditions for fresh water systems.

The goal of this study was to develop a methodology for direct measurement of volume reflection, L_v, from an above-water platform, and to understand the dynamics of L_r from fresh water under diurnal and annual variability in factors that affected CF. The specific objectives of this study were to:

(1) Develop an approach to directly measure upwelling water volume reflection from an above-water platform.
(2) Understand the contribution of L_r to total upwelling water radiance.
(3) Study the effect of sun altitude, diffuse light conditions, wind speed, and wavelength on CF.
(4) Develop an ANN model to simulate CF from known environmental conditions for specific Vis-NIR wavelengths.

MATERIALS AND METHODS

Above-Water Spectral Radiance. This study assumed that the total upwelling radiance just above the water surface at any given zenith angle is the sum of the volume reflection of the water body and the surface reflection from the air–water interface. The bottom reflection was ignored in this study since it was a constant and was less than 2% in the Vis-NIR wavelength range considered here. All other components of upwelling radiation such as those due to path scattering are considered negligible since the distance between the sensor and the water surface was considerably small.

The methodology used in the study also assumed that the water surface is calm and specular. According to the law of specular reflection, the angle of incident light (θ_i) is equal to the angle of surface reflected light (θ_r), as indicated in Fig. 1. Therefore, a sensor oriented at nadir will receive the surface reflected component (L_r) of downwelling radiation from nadir and the volume reflection coming from under the water surface (L_v). Only the L_v component of the upwelling light carries information on water quality. By placing a metal sheet directly above the sensor to block the incoming irradiance over a small area, the surface reflection, L_r, at nadir could be avoided without affecting L_v (Fig. 2). Since upwelling radiance from under the water surface at any given point or small area is

generated by the light intercepting water at all different angles in a larger surrounding area, blocking of light coming at nadir over a small area will not affect L_v. Therefore, the main assumption used in the experimental setup was that if the light coming at nadir is blocked, the nadir-oriented sensor will receive only volume reflection, L_v.

Since the downwelling and upwelling radiances have a hemispherical distribution, the radiance and reflectance measurements are usually made at specific zenith angles, at which the sensor is oriented. In this study, the sensor was oriented at nadir; therefore, the zenith angle was zero for all measurements. The surface reflected radiance, L_r, was calculated as the difference between total upwelling water radiance and volume reflection (Eq. 3).[5,7] The total upwelling water radiance was measured directly above the water surface without blocking the downwelling radiation at nadir. The CF or ρ_s was calculated as the ratio of L_r to L_s as shown in Eq. 1. The contribution of L_r to the total upwelling radiance, L_t is calculated as the ratio between the two (Eq. 4). The total water-body reflectance, which was the ratio of the total upwelling radiance to the total downwelling radiance above water surface, was calculated using Eq. 5, while volume reflectance of water was calculated with Eq. 6. The error introduced by L_r in the measured water reflectance was calculated as the ratio of L_r to L_v (Eq. 7).

$$c = \frac{L_r}{L_t} \times 100 \tag{4}$$

$$\rho_w = \frac{L_t}{\pi L_d} \times 100 \tag{5}$$

$$\rho_v = \frac{L_v}{\pi L_d} \times 100 \tag{6}$$

$$e\,\frac{\rho_w - \rho_v}{\rho_v} \times 100 = \frac{L_t - L_v}{L_v} \times 100 = \frac{L_r}{L_v} \times 100 \tag{7}$$

where the variables are as follows:

- c = Contribution of L_r to total upwelling radiance above water surface at a specific zenith angle, %
- L_d = Total downwelling radiance at a specific zenith angle, W m^{-2} $sterdian^{-1}$
- ρ_w = Total water-body reflectance at a specific zenith angle, ratio
- ρ_v = Volume reflectance of water at a specific zenith angle, ratio
- e = Error in measured water volume reflectance caused by surface reflection, %.

Tank Experiments and Data Collection. The study was conducted at the Arkansas Agricultural Experiment Station facility at Fayetteville (36.0999348 N, 94.175325 W). The experimental setup consisted of a rectangular tank (2.44 m length × 1.83 m width × 0.61 m height), with bottom and sides painted black, which resulted in a wall reflectance of less than 2%. The tank was filled with clear tap water up to 0.33 m depth. Three arms were fitted on the tank, two for holding two optical sensors, and the third one for holding a small metal sheet, painted black (Fig. 3). The 0.3 m × 0.3 m metal sheet was used to block the downwelling radiation coming at nadir to the water surface at the point of measurement. The metal plate was mounted 0.61 m above the water surface, causing the downwelling light at a solid angle of 23° to be blocked with respect to the view area.

The data collected included radiances L_t, L_v, L_d, L_s, wind velocity, and sun altitude. Radiances were measured with a portable ASD Field Spec Pro Dual VNIR spectro-radiometer (Analytical Spectral Devices, CO). The spectro-radiometer had two sensors simultaneously measuring radiance in 512 bands in the Vis-NIR range of 350-1050 nm. The two sensors were referred to as reference and target sensors, with the target sensor measuring the upwelling radiance from the water body and the reference sensor measuring the downwelling radiance simultaneously. The downwelling radiance was measured as the light reflected by a reference (Spectralon®) panel that had a 98% reflectance in the 400-1000 nm range. The measured reflectance was adjusted for the reflectance of the Spectralon® panel to obtain the downwelling radiance. When no metal sheet was blocking the downwelling radiance at nadir, the radiance measured by target sensor was L_t and the radiance measured by the reference sensor was L_d. When the black metal sheet was placed above the sensors to block the downwelling radiance at nadir, the target sensor recorded L_v while the reference sensor recorded L_s. The target sensor had a 8° field of view (FOV), and the reference had a 25° FOV. The target sensor was mounted at 0.3 m above the water surface, resulting in a view area of 14 cm^2. The reference sensor was mounted at 3 cm above the reference panel, resulting in a view area of 1.4 cm^2.

Wind speed was recorded during the data collection period with a hand-held anemometer (kestrel#1000, Kestrel Inc., Sylvan Lake, MI) mounted 1 m above the water surface. Sun altitude data for Fayetteville were obtained from the US Naval Lab (http://aa.usno.navy.mil/data/docs/AltAz.htmlusnaval.gov). Data were collected on five different dates (September 25, October 20, November 8, and December 8 and 9) in 2006, and one day (May 19) in 2007, under clear sky conditions. On each date, 14 to 16 complete sets of observations were collected between 10:30 a.m. and 2:30 p.m. at 15 min intervals.

Artificial Neural Network Model Development. An artificial neural network (ANN) is a parallel distributed processing unit that works similar to the human brain's neural network system.[13] Data driven models such as ANN have the capability to develop a numerical relationship between a set of inputs and outputs without any prior understanding of the system. This was a major advantage in using the ANN as a modeling tool for representing the highly complex and nonlinear relationship of CF with sun altitude, diffuse light, wind speed, and wavelength in this study.

In this study, a feed-forward multilayer perceptron (MLP) network was used for the development of the model.[13] In an MLP network the processing units or neurons are arranged in layers and the information flows forward from the input to output layer through connections between neurons in successive layers. Each connection has a weight associated with it. The inputs from the previous layer coming to a specific neuron are scaled by the respective connection weights and bias of the neuron, and then they are transformed into an output value of the neuron using the processing or activation function of the neuron. The four inputs into the model included sun altitude, diffuse light, wind speed, and wavelength, which were selected based on past research indicating their influence on CF. The model output was CF. The activation functions used were log-sigmoid at the first hidden layer, tan-sigmoid at the second hidden layer, and pure linear at the output layer.

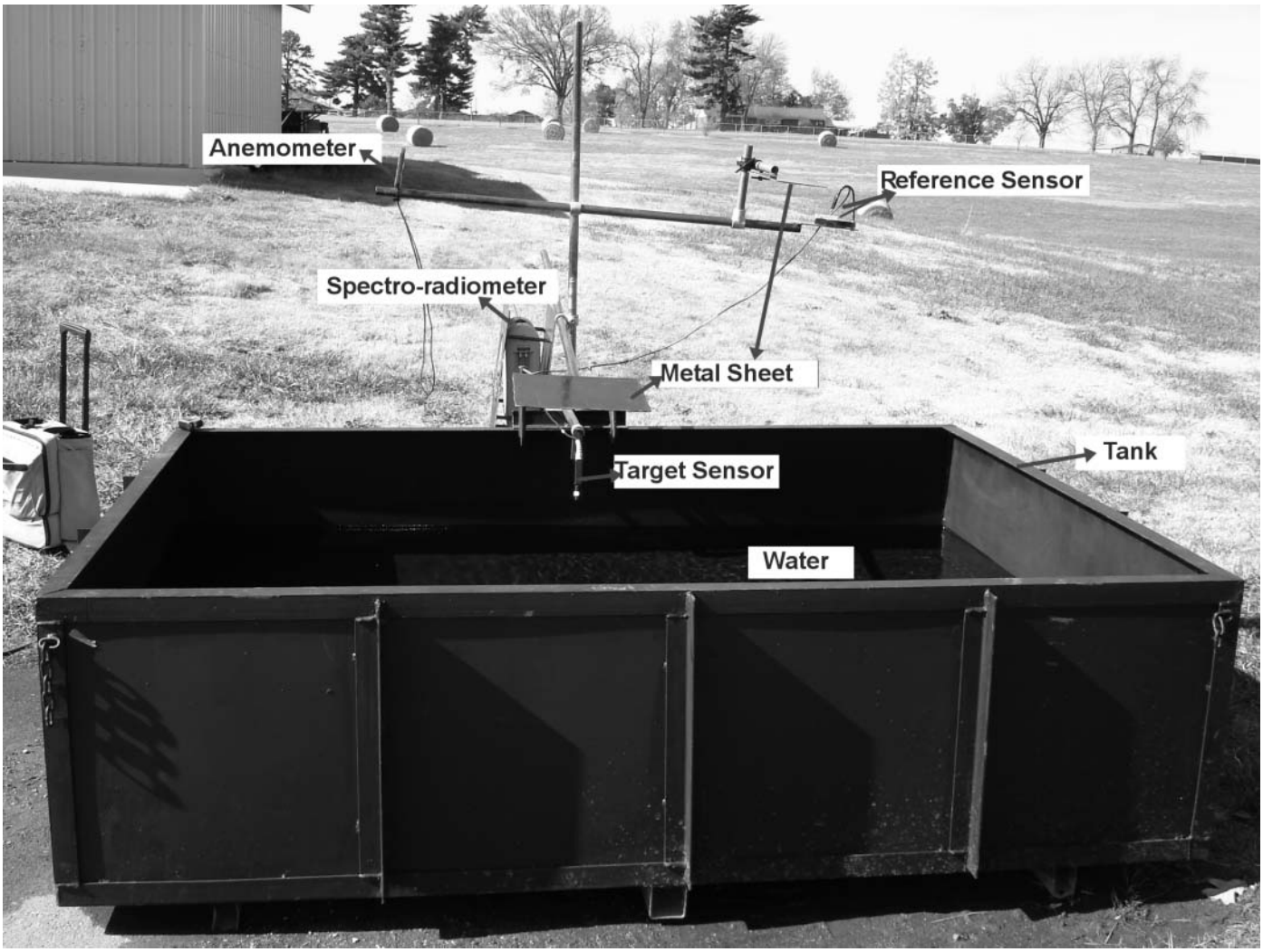

FIG. 3. Experimental setup developed at the University of Arkansas at Fayetteville for measuring upwelling water volume reflection without the surface reflected radiance.

For modeling, only wavelengths at 9 nm intervals within the range of 400-1000 nm were considered, resulting in 68 wavelengths. The complete set of data collected over six dates in 68 wavelengths included 6188 observations. These observations were divided into three groups of training, validation, and testing data. The data were organized by the time of data collection. A structured method that is commonly used in ANN model development was used to divide the data into the three groups. Every fourth observation was chosen for validation, every fifth observation for testing, and observations 1–3 for training the model. Thus, 3712 data points were used for training, 1238 for validation, and 1238 for independent testing of the model. The descriptive statistics of the three groups of data showed that they were comparable, with similar mean, standard deviation, and range for all variables (Table I). It should be noted that the data obtained in this study had a different range compared to the data used by Mobley.[4] The difference in experimental setup, water used, sensor orientation, and the location of the study all have contributed to this difference in the range of variables.

Error back propagation is a common method used for training feed-forward MLP networks. The automated regularization back propagation algorithm[14] available in Matlab (The

TABLE I. Descriptive statistics of the observed input variables (sun altitude, diffuse light conditions, wavelength, and wind speed) and output (correction factor) used for the ANN model development.

	Sun altitude (degrees)	Diffuse light	Wavelength (nm)	Wind speed (m/s)	Correction factor
Training					
Min	20.60	0.00196	400.0	0.8	0.0025
Max	73.70	0.4456	1000.0	6.8	0.139
Mean	41.17	0.07144	701.5	3.4	0.019
Stdev	13.50	0.05653	176.7	1.3	0.017
Validation					
Min	20.60	0.00196	400.0	0.8	0.003
Max	73.70	0.41030	1000.0	6.8	0.121
Mean	41.15	0.07138	701.5	3.4	0.019
Stdev	13.50	0.05667	176.7	1.3	0.017
Testing					
Min	20.60	0.00196	400.0	0.8	0.003
Max	73.70	0.37540	1000.0	6.8	0.127
Mean	41.19	0.07137	701.5	3.4	0.018
Stdev	13.51	0.05651	176.6	1.3	0.017

Mathworks, Inc., Natick, MA) was used for training the model. Automated regularization training works on the basis of a Bayesian framework.[14] This training algorithm tries to minimize the sum of square error (SSE) and updates weights and biases after each simulation of the network. Advantages of the backpropagation automated regularization training algorithm include high level of accuracy, reliability due to the probability based algorithm, less risk of over-fitting, simplicity in implementation, and capability to estimate the effective number of parameters affecting the model.[14] Training stopped when the stop criteria of either the number of epochs of 2100 (selected based on trial and error), or a SSE of 10^{-4} was reached.

The ANN model was optimized for the number of neurons in the hidden layer and the number of hidden layers in the network using a trial and error procedure. Performance of the model during optimization was tracked with the root mean square error (RMSE) and R^2 values on validation data. Model optimization started with one hidden layer with a minimum of one neuron. The numbers of neurons were systematically increased by one, in each successive run, till a satisfactory performance of the model was attained. This method identified 16 neurons in the hidden layer as the best, which resulted in an R^2 value of 0.98 and RMSE of 0.0001. However, the rule of thumb for optimization procedures is that the number of neurons in a hidden layer should not exceed twice the number of inputs.[15] Therefore, the number of neurons in a hidden layer should not exceed 8 since there were 4 inputs. Since model performance was relatively poor with 8 neurons in the hidden layer, a second hidden layer was added. The model was again optimized for the number of neurons in both the hidden layers. The model with two hidden layers containing 4 and 3 neurons, respectively, provided the best model performance, indicated by an R^2 value of 0.96 and RMSE of 0.0001.

The optimized model was trained with the training data while tracking the performance on the validation data in order to avoid over-fitting. Model training is a structured iterative process to identify the best set of values for the model parameters that minimizes the prediction error. In an ANN model, the model parameters include connection weight between the neurons, and the biases of each of the neurons. The fully trained model establishes a complex nonlinear relationship between the inputs and the output using the best set of weights and biases. The overall performance of the fully trained ANN model was evaluated on the independent test data set based on its predictive power measured as Nash–Sutcliff efficiency (R_{NS}^2) of the model,[16] in addition to the R^2 value and RMSE.

A sensitivity analysis was conducted to evaluate the effect of the four input variables on CF. The sensitivity of each input variable was assessed as a relative contribution of that variable to the prediction of the output. The connection weights between the neurons along the path from a specific input to the output layer can explain the relative importance of each variable in the model. Therefore, the sensitivity of an input was calculated as the weighted cumulative sum of connection weights between neurons connecting that input to the output, using Eq. 8.[17] This method of sensitivity analysis is only valid for ANN models. Biases were not considered in the sensitivity analysis procedure since the connection weights are multiplicative terms and biases are additive terms for an input. Also, the connection between the neurons facilitates the flow of information in a back-propagation network. Garson[17] also advised partitioning the connection weights rather than biases for sensitivity analysis.

$$S_r = \frac{\sum_{j=1}^{n_{H1}} \frac{I_{vi}}{\sum_{v=1}^{n_v} I_{vi}} H1_i O_2}{\sum_{v=1}^{n_v} I_{vi} \sum_{j=1}^{n_{H1}} \left(\frac{I_{vi}}{\sum_{v=1}^{n_v} I_{vi}} H1_i O_2 \right)} \quad (8)$$

where the variables are as follows:

S_r = Sensitivity of an input in predicting the output, %
n_{H1} = Number of neurons in the first hidden layer, number
I_{vi} = Weights connecting the i^{th} input to the first hidden layer, unitless
n_v = Number of input variables, number
i = Inputs in the input layer, number
j = Nodes in the hidden layers, number
H_{1i} = Weights connecting the first to the second hidden layer, unitless
O_2 = Weights connecting the second hidden layer to the output layer, unitless

RESULTS AND DISCUSSION

Surface Reflection from Water. The L_r, which is the radiance reflected from the air–water interface at nadir, ranged from 0.00001 to 0.0048 W m^{-2} sterdian^{-1} for the six days of data collection (Table II). The daily mean L_r was the highest on 19 May 2007 and lowest on 8 December 2006. The reason for the high L_r on 19 May 2007 may be the high sun altitude, which reached the highest value of 73.7° that day at the experiment location. Similar observations were made by Doxaran et al. (2004) and Austin (1974) that high sun altitudes dramatically increased the L_r. The L_r accounted for a daily mean of 5 to 65% of the total upwelling radiance in the wavelength range of 400 to 1000 nm (Fig. 4). The daily mean errors introduced by L_r in the measured water volume reflectance varied from 20–54%, which shows the importance of and need for removing L_r from above-water radiometric measurements. There was significant variation in the contribution of L_r to total upwelling radiance throughout the Vis-NIR region. In the visible range of 400–700 nm, the daily mean L_r contribution gradually decreased from the highest range of 24–35% at 400 nm to the lowest range of 5–17% at 700 nm. In the NIR region, the daily mean L_r contribution increased from the lowest range of 5–17% at 700 nm to a local peak of 15–30% at 750–760 nm, and then decreased to 10–24% at 815 nm, followed by a steep increase up to 920 nm.

Although the average L_r was the highest on May 19, 2007 (Table II), the relative contribution of L_r over the NIR wavelengths did not follow this trend (Fig. 4). The mean daily contribution of L_r to the total upwelling radiation in the 820-1000 nm range on 19 May 2007 was comparable to the data from 9 December 2006, with both values significantly lower than that for the four remaining days of data collection. On both of these dates, the highest mean daily contribution of L_r in the NIR range was 33%, compared to the 50–65% observed on

TABLE II. Summary of daily variation in measured variables such as surface reflected radiance (L_r), correction factor (ρ), diffuse light condition (ratio of downwelling diffuse radiance to total downwelling radiance), sun altitude (SA), and wind speed (WS) during the six days of tank experiment with tap water under clear sky conditions.

Variable	Date	Mean	Std.dev	Min	Max
Surface reflected radiance, L_r (W m^{-2} sterdian^{-1})	9/25/2006	0.00036	0.00035	0.00002	0.00143
	10/20/2006	0.00037	0.00036	0.00001	0.00143
	11/8/2006	0.00026	0.00029	0.00001	0.00113
	12/8/2006	0.00013	0.00014	0.00001	0.00057
	12/9/2006	0.00020	0.00022	0.00001	0.00089
	5/19/2007	0.00129	0.00103	0.00006	0.00477
Correction factor, CF or surface reflectance, ρ_s (ratio)	9/25/2006	0.0148	0.0043	0.0075	0.0278
	10/20/2006	0.0163	0.0060	0.0065	0.0512
	11/8/2006	0.0135	0.0042	0.0056	0.0269
	12/8/2006	0.0070	0.0022	0.0025	0.0127
	12/9/2006	0.0122	0.0087	0.0030	0.0521
	5/19/2007	0.0526	0.0258	0.0130	0.1391
Diffuse light condition (ratio)	9/25/2006	0.0603	0.0421	0.0196	0.2439
	10/20/2006	0.0810	0.0596	0.0257	0.3479
	11/8/2006	0.0818	0.0645	0.0242	0.3755
	12/8/2006	0.0934	0.0768	0.0244	0.4040
	12/9/2006	0.1028	0.0759	0.0290	0.4456
	5/19/2007	0.0692	0.0353	0.0314	0.2211
Sun altitude, SA (degrees)	9/25/2006	49.6	3.2	42.6	52.9
	10/20/2006	39.2	5.0	27.3	43.5
	11/8/2006	29.0	2.2	24.1	31.2
	12/8/2006	28.7	2.4	24.1	31.2
	12/9/2006	28.1	3.3	20.6	31.1
	5/19/2007	68.0	6.0	56.6	73.7
Wind speed, WS (m s^{-1})	9/25/2006	1.9	0.6	1.1	3.4
	10/20/2006	3.2	1.1	0.8	5.2
	11/8/2006	4.6	1.1	2.5	6.8
	12/8/2006	3.6	0.9	2.0	4.3
	12/9/2006	5.2	0.9	3.5	6.7
	5/19/2007	3.4	0.5	2.5	4.2

the remaining days. The reason for the relatively lower contribution of L_r in the NIR region on these two dates is not known with the limited data we have recorded. These two days were not similar in environmental conditions monitored (Table II). The diffuse radiance and wind speed were the highest, and sun altitude was the lowest on Dec. 9, whereas, the sun altitude was the highest, and the diffuse lighting and wind speed were below average, on May 19.

Correction Factor. The correction factor (CF) or ρ_s estimated from radiance data collected at 15 min intervals over six days varied from 0.0025 to 0.130 with a mean of 0.019 and standard deviation of 0.017 in the Vis-NIR wavelength range of 400 to 1000 nm (Table I). The coefficient of variation of the CF, calculated as the ratio of the standard deviation to the mean, was 0.92, which indicates that there was a high amount of variability in CF for the six days of data collection. The mean daily CF was slightly less than 0.02 recommended by Austin,[11] except for May 19 (Table II). One reason for the relatively smaller value of CF for the five dates may be the difference in how wind affected the water surface in a tank experiment. The surface roughness created on the water surface by similar wind speeds could be relatively less in a tank experiment compared to open and vast water bodies. Additionally, the wind velocity was measured approximately 1 m above the water surface. Therefore, the actual wind speed at the water surface could have been relatively smaller than the measured values.

The daily mean CF was compared to the value suggested by Austin[11] with Z-tests. The null hypothesis for the Z-test was that the daily mean of CF is equal to 0.02 as suggested by Austin.[11] The null hypothesis was rejected for each day of data collection, which implies that the daily mean CF found in this study was significantly different from 0.02. Although the six-day average of CF was close to 0.02, the daily mean CF was significantly lower than 0.02 on five dates. On May 19, the mean CF was 0.05, which was significantly higher than 0.02 ($P < 0.05$). Therefore, a constant CF of 0.02 will cause significant over or under estimation of L_r, as suggested by Doxaran et al.[5] It should be noted that a CF of 0.02 was suggested for calm marine water systems for a nadir sensor viewing geometry when the sun altitude and sun glint were low.[12] The conditions used in this study (fresh water in a tank) were very different

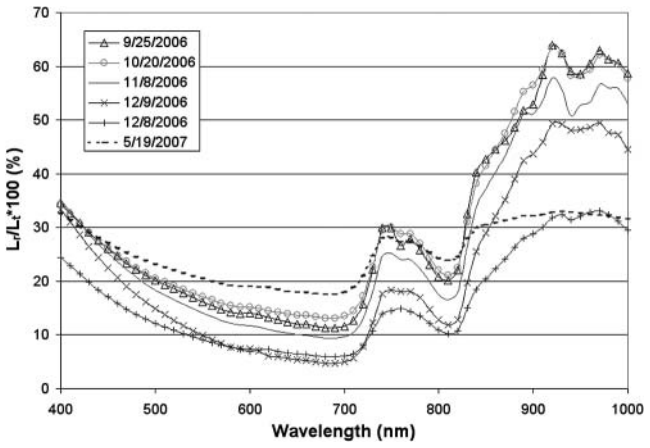

FIG. 4. Mean daily contribution of surface reflected radiance (L_r) to total upwelling water radiance (L_t) calculated as (L_r/L_t) × 100, for data collected from a tank filled with tap water on six different days.

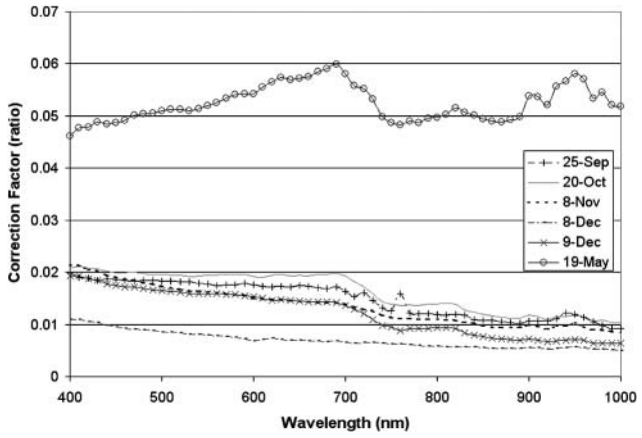

FIG. 5. Variation of mean daily correction factor with respect to wavelength on six different days of data collection between 10:30 a.m. and 2:30 p.m. for a tank of fresh water.

from those suggested by Arst et al.[12] Additionally, all environmental variables such as sun altitude, wind velocity, and diffuse light conditions in this study had considerable variability since the data were collected on six different days spread across nine months.

In general, the daily mean CF gradually decreased from 400 nm to 1000 nm, with a relatively sharp decrease in the 700 to 800 nm region (Fig. 5). The one exception to this trend was the data from May 19. On this date, the CF slowly increased from 0.046 at 400 nm to 0.06 at 690 nm. Additionally, CF showed higher variability within the visible wavelengths on May 19, compared to the other five days. The variations in CF over the visible wavelengths were comparatively small on all five dates except for May 19. Generally, the CF was much higher at visible wavelengths than at NIR wavelengths, with the exception of May 19. On May 19, the entire Vis-NIR region had comparable values of CF.

The daily mean CF was the highest on May 19 and the lowest on December 8 on all wavelengths in the 400-1000 nm range. The significantly high sun altitude on May 19 is the reason for the high CF on this date. While the sun altitude remained below 53° for the five other days of data collection, it varied from 56.6° to 73.7° on May 19. Other factors such as wind velocity and diffuse lighting conditions on May 19 were comparable to the other five days. A comparison of the CF from December 8, when the CF was the lowest, to that of December 9, an adjacent day, indicated large differences. The CF for December 9 was approximately twice as much as the CF for December 8 at all wavelengths. Both days were clear days with similar diffuse light conditions, with a mean of 0.09 and 0.10 (Table II). Being adjacent days, the sun altitudes were also very similar, with means of 28.7° and 28.1° respectively on December 8 and 9. Therefore, sun altitude and diffuse light were ruled out as possible causes of the observed differences in CF. The remaining factor, wind speed, was significantly different on these two days and was attributed as the reason for the large differences in CF. The average wind speed was 3.6 m s^{-1} on December 8, with a range of 2–4.3 m s^{-1}. On December 9, the average wind speed was 5.2 m s^{-1} with a range of 3.5–6.7 m s^{-1}. This result suggests that a 45% increase in wind speed can cause the increase of the CF by approximately 100%.

Figure 6 shows the variation of CF with respect to sun altitude and diffuse light conditions at two visible wavelengths

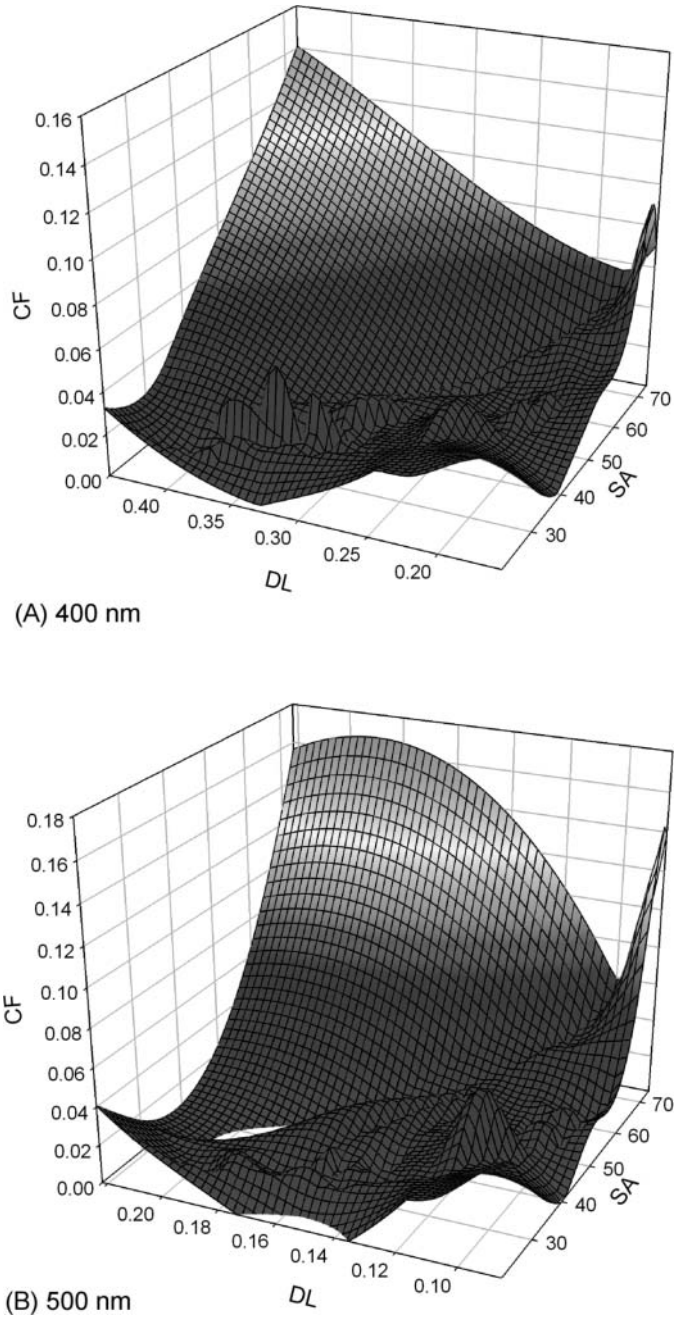

FIG. 6. Sample 3D mesh plots of diffuse light (DL), sun altitude (SA), and correction factor (CF) for data collected on six days at sample wavelengths of (**a**) 400 nm and (**b**) 500 nm.

of 400 nm and 500 nm. In general, the CF increased in a highly nonlinear fashion with respect to sun altitude and diffuse light at both sample wavelengths. A combination of the high sun altitude of above 40° and high diffuse light of above 0.25 resulted in sharp increases in the CF. Although the trend of CF with respect to sun altitude and diffuse light were somewhat similar, the specific patterns of the three-dimensional curve were very different at these two sample wavelengths. Similar nonlinear patterns were also observed between wind speed, sun altitude, and CF.

Artificial Neural Network Model Performance. The ANN model simulating CF from sun angle, wind velocity, diffuse light, and wavelength performed very well under both training

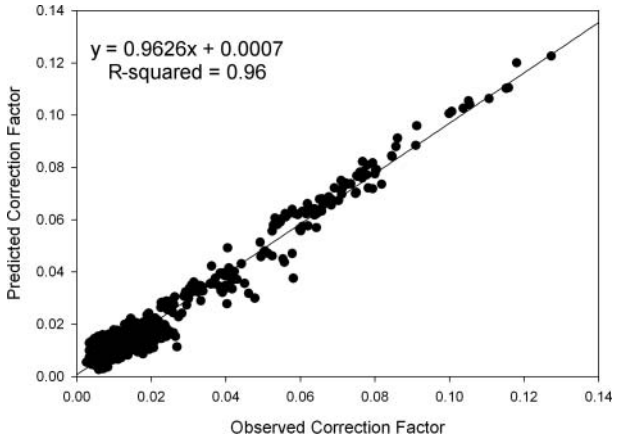

FIG. 7. Correction factor predicted by the trained ANN model plotted against observed values for an independent test data set.

and testing. The accuracy of the simulated CF was very high, which was indicated by a low RMSE of 0.0032 under training, validation, and testing. The ANN model was able to explain 96% of variability in CF in both the testing and validation data, and 97% of variability in CF in the training data. The tight distribution of the simulated CF with respect to the measured CF data around the 1:1 curve, and the high correlation between the simulated and measured CF ($R^2 = 0.96$), also indicated the superior performance of the ANN model (Fig. 7). The model also exhibited very high predictive power or efficiency, which was indicated by an R_{NS}^2 of 0.96 for test and validation data, and 0.97 for training data. A model with an R_{NS}^2 of unity is considered as 100% efficient in simulating the output variable.

With this ANN model, it would be relatively simple to estimate a highly accurate value of CF if the values of the four input variables are known. For any given location, it is simple to obtain the sun altitude values for any time of the day from the US Naval Lab. Wind speed data can be obtained from local, state, or federal agencies for established weather stations. Since wind speed typically shows high spatial variation, the best method for accurate measurement is to install multiple continuously recording anemometers distributed within the area of interest. If that is not practically possible, it is advisable to obtain the wind speed data from the closest possible weather station. The diffuse lighting condition would be the only other variable that needs to be measured. It can be easily measured in specific wavelengths with a hand-held spectro-radiometer. This study used fresh water stored in a tank to develop the basic understanding prior to conducting actual field tests. Since the ANN model worked extremely well for the data from the tank study, we expect this model to perform well with field data. This type of model would be easy to develop and implement for actual field data and hence offers an inexpensive and fast solution for estimating CF and L_r.

Input Sensitivity Analysis. A sensitivity analysis was performed with the calibrated and validated ANN model for estimating CF. The sensitivity analysis ranked the input variables based on their sensitivity or relative contribution to the output. Sun altitude and wind speed had the highest sensitivities of 46.4 and 36.1%, respectively, contributing to over 82.5% of the total predictive ability of the model. Diffuse light had a sensitivity of 14.5%, followed by wavelength with the lowest sensitivity of 2.65%. Although the contribution of wavelength to the predictive ability of the model was relatively low, it is important to include wavelength as a variable since the CF showed significant variability between visible and NIR wavelengths at low sun altitude, and high variability within the whole range of 400-1000 nm under high sun altitudes (Fig. 5). Additionally, the sensors used for remote sensing operate at specific wavelength ranges. Therefore, knowledge of CF at specific wavelengths or wavelength regions can improve the accuracy.

The ANN model developed for tank water was successful in simulating the CF as a function of sun altitude, diffuse light conditions, wind speed, and wavelength. ANN based models are data-driven models and might not work efficiently outside the training range of input data. This model was trained for a sun altitude range of 20.6° to 73.7°, which is typical for Fayetteville in Arkansas. Although sun altitude can be as high as 87° for the southern/southeastern United States, such as Texas and Florida, in the month of June, the maximum sun altitude in the rest of the US including Arkansas is approximately 77°. Therefore, the range considered in this study was representative of most of the regions in the US. The wind speeds ranged between 0.8 and 6.7 m/s during the 6 days of data collection, which is relatively low considering potential wind speeds over open water bodies at this location. It would be possible to retrain the model with additional data collected at higher wind speeds and run it under an extended range of wind speeds. The wind speed was measured at 1 m above the water surface, which may have been an over-estimation of the actual wind speed at water surface. This is especially true for a tank experiment since the walls of the tank covered and protected the water surface from wind somewhat. This model was developed for data measured with the sensor oriented at nadir and may not accommodate other sensor geometries.

A potential limitation of the model was the assumption of the smooth water surface, which was only possible under low surface roughness conditions. The ripples or waves created by the high wind speed, as well as excessive amount of optically active constituents at the surface, can create more diffuse reflectance conditions. The model was developed for clear unfiltered tap water, which had very low OAC compared to fresh water bodies or marine water systems. High levels of OAC can change the surface roughness and behavior of the air–water interface. Therefore, it is possible that the L_r and, hence, the CF obtained in this study are underestimations of actual CF for field conditions or fresh water bodies. This controlled study was conducted to understand the dynamics of L_r and CF on a relatively calm and smooth water surface.

CONCLUSION

A new approach was developed for removal of surface reflection from total water body reflection in above-water radiance measurements at nadir. With this new approach, the L_r can be easily removed by directly measuring the radiance coming from under the water surface at nadir. The L_r measured from clear tap water maintained in a tank contributed 5–65% of the total upwelling radiance, resulting in 20–54% of error in the measured above-water reflectance. The surface reflectance or CF showed large variability with a mean of 0.019 and coefficient of variation of 0.92 for the six days of data collection. The CF decreased with wavelength at low sun altitudes and exhibited a highly nonlinear relationship with sun altitude, diffuse light, and wind speed. Both wind velocity and high sun altitude dramatically increased the CF.

The CF was modeled on sun altitude, diffuse light, wind speed, and wavelength with ANN, a data-driven model framework. The model was very successful in simulating CF, which was indicated by the high accuracy (RMSE = 0.003), strong predictive ability ($R_{NS}^2 = 0.96$), and large amount of variability ($R^2 = 0.96$) in the CF explained by the model for a test data set. Sensitivity analysis of the ANN model showed that sun altitude and wind speed were the two most sensitive parameters affecting the predictability of CF. Together, they accounted for 82.5% of the model's predictive capability. Wavelength contributed less than 3% of the model's predictability. The study revealed how the experimental approach can be applied for obtaining CF with high accuracy for specific combinations of wavelength, sun altitude, wind velocity, and diffuse light for a nadir sensor view geometry by using an ANN-based modeling tool. If retrained with actual field data, the model is expected to perform well in predicting correction factors for actual field conditions with high accuracy, thus avoiding tedious field measurements.

1. A. Gittelson, F. Szilagyi, and K. H. Mittenzwey, Water Res. **27,** 1185 (1993).
2. K. E. Sawaya, L. G. Olmanson, N. J. Heinert, P. L. Brezonik, and M. E. Bauer, Remote Sens. Environ. **88,** 144 (2003).
3. B. R. Fougnie, R. Frouin, P. Lecomte, and P. Deschamps, Appl. Opt. **38,** 3844 (1999).
4. C. D. Mobley, Appl. Opt. **38,** 7442 (1999).
5. D. Doxaran, R. C. Nagur, K. Cherukuru, and S. J. Lavender, J. Pure Appl. Opt. **6,** 690 (2004).
6. D. A. Toole, D. A. Siegel, D. W. Menzis, M. J. Neumann, and R. C. Smith, Appl. Opt. **39,** 456 (2000).
7. K. Oki and Y. Yasuoka, Appl. Opt. **41,** 6463 (2002).
8. H. Arst, *Optical Properties and Remote Sensing of Multi-Componental Water Bodies* (Praxis Publishing, United Kingdom, 2003).
9. J. M. Froidefond and S. Ouillon, Opt. Exp. **13,** 2110 (2005).
10. T. Takashima, J. Meteorol. Geophys. **34,** 75 (1983).
11. R. W. Austin, "The remote sensing of spectral radiance from below the ocean surface", in *Optical Aspects of Oceanography,* N. G. Jerlov and E. S. Nielson, Eds. (Academic Press, London, London, p. 317.
12. H. Arst, D. Pozyadoknav, and A. Rosenstien, *Optical remote sensing in oceanology* (Tallinn, Valgus, Russia, 1990).
13. S. Haykin, *Neural Networks: A Comprehensive Foundation* (MacMillan, New York, 1994).
14. D. J. C. MacKay, Neural Comp. **4,** 448 (1992).
15. J. T. Heaton, *Introduction to Neural Networks with Java* (Heaton Research Inc., St. Louis, MO, 2005), p. 125.
16. D. G. Garson, AI Expert. **6,** 46 (1991).
17. J. E. Nash and J. V. Sutcliffe, J. Hydrol. **10,** 282 (1970).