Machine Learning-Based Ensemble Prediction of Water-Quality Variables Using Feature-Level and Decision-Level Fusion with Proximal Remote Sensing

Kyle T. Peterson, Vasit Sagan, Paheding Sidike, Elizabeth A. Hasenmueller, John J. Sloan, and Jason H. Knouft

Abstract

The objectives of this study were to accurately model relationships between spectral reflectance and water-quality parameters, including blue-green algae phycocyanin, chlorophyll a, total suspended solids, turbidity, and total dissolved solids; evaluate feature-level fusion to spectral data for water-quality modeling; and evaluate the effectiveness of machine learning regression techniques and decision-level fusion for water-quality variable prediction. We introduce the application of canonical correlation analysis fusion as a method for water-based spectral analysis to overcome the low signal-to-noise ratio of the data. Water-quality variables and spectral reflectance were used to create predictive models via machine learning regression models, including multiple linear regression, partial least-squares regression, Gaussian process regression, support vector machine regression, and extreme learning machine regression. The models were then combined using decision-level fusion. Results indicate that canonical correlation analysis feature-level fusion and machine learning techniques are superior to traditional methods.

Introduction

Degradation of freshwater resources has become increasingly problematic, particularly in areas with intensive agriculture such as the Midwestern United States. Inland freshwater systems, which provide drinking water, commodity transportation, fishing, recreation, agricultural irrigation, and hydroelectric power, among many other uses, are highly sensitive to changes in climate, land cover/land use, and anthropogenic activities. These changes can have many deleterious effects on water-quality, including problems with eutrophication, harmful algal blooms (HABs), and siltation, which have become increasing concerns in recent years.

HABs are a major water-quality problem because they can have significant socioeconomic and ecological costs (Carmichael and Boyer 2016). These HABs, often consisting of the liver-toxic *Mycrocystis* cyanobacteria, cause major disruptions to freshwater ecosystems, including fish kills, public-access closures, and contamination of drinking-water sources. Increasingly, eutrophic conditions in freshwater ecosystems, such as the record-setting algal blooms in Lake Erie (Kane *et al.* 2014), the Ohio River (Brooks *et al.* 2017), and the Gulf of Mexico (Rabalais and Turner 2017), are primarily attributed to an overabundance of nutrients (i.e., nitrogen and phosphorus species) in streams and rivers that drain agricultural lands. The production of row crops, such as corn and soybeans, in the Midwestern United States requires the copious use of fertilizers, resulting in runoff and tile drainage from agricultural fields that contains high concentrations of nutrients. These compounds can trigger algal blooms under certain conditions that are difficult to predict (Brooks *et al.* 2017). Fertilizers are transported along with sediment to streams and lakes where seasonal hypoxic conditions promote the solubilization of phosphorus, creating favorable conditions for algal blooms (Pearce *et al.* 2017).

Some effort has been made in using hyperspectral imagery to detect and measure HABs (Kudela *et al.* 2015); however, less has been focused on the use of hyperspectral data to measure other water-quality parameters that are associated with HABs. The use of remote sensing enables monitoring of key water-quality variables linked with HAB outbreaks such as blue-green algae phycocyanin (BGA-PC) and chlorophyll a (Chl-a). As with HABs, issues related to excess sediment transport can be monitored using variables such as total suspended solids (TSS) and turbidity, while other types of hydrologic processes can be determined by evaluating levels of total dissolved solids (TDS). Traditional monitoring and assessment of water-quality has relied on *in situ* measurements and lab analyses of indicators related to the physical, chemical, and biological properties of the water body. This most commonly entails field-collected water samples followed by laboratory analysis. Methods such as these, which are generally very accurate, are quite time consuming, expensive, and not feasible for regional-scale monitoring and research (Duan *et al.* 2013). Recent advances in remote sensing technologies and statistical modeling techniques provide an efficient solution for regional-scale water-quality monitoring. Using a remote sensing approach, limitations such as spatial and temporal coverage may be eliminated by the use of satellite or airborne sensors.

Background

Water-surface reflectance of incident solar radiation can be quantified utilizing wavelengths that lie within both the visible and near-infrared portions of the electromagnetic

Kyle T. Peterson, Vasit Sagan (vasit.sagan@slu.edu), Paheding Sidike, and Elizabeth A. Hasenmueller are with the Department of Earth and Atmospheric Science, Saint Louis University, Saint Louis, MO.

John J. Sloan is with the National Great Rivers Research and Education Center, East Alton, IL.

Jason H. Knouft is with the Department of Biology, Saint Louis University, Saint Louis, MO.

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 4, April 2019, pp. 269–280.
0099-1112/18/269–280

doi: 10.14358/PERS.85.4.269

spectrum, from approximately 350 to 1200 nm. Through analysis of the quantified absorption and scattering properties of the water surface, its inherent optical properties (IOPs) can be determined. IOPs are properties of a medium that depend not on an ambient light source but rather upon the constituent composition and physiochemical properties of the surface water (C. Huang *et al.* 2015). IOPs, when correlated with traditional *in situ* measurements and lab analyses, can help characterize numerous variables (e.g., Chl-a, TSS), providing an overall picture of water-quality.

Chl-a and BGA-PC have been used to detect HABs and other algae in freshwater (Gholizadeh *et al.* 2016; Kudela *et al.* 2015). Previous studies (Kutser 2009; Ogashawara *et al.* 2013) identifying HABs have relied on the optical properties of phycocyanin, a pigment present in freshwater algal blooms. The phycocyanin pigment is often detected from remote sensing by a broad absorption feature between 615 and 630 nm (Ogashawara *et al.* 2013). Chl-a, a photosynthetic pigment also found in algae, absorbs light in the blue and red wavelengths, with maximum fluorescence at 670 and 681 nm (Sathyendranath *et al.* 2001). Chl-a also reduces the light reflected in these regions, resulting in the green color of algal blooms (Stumpf *et al.* 2016). Remote sensing provides a useful means for estimating Chl-a concentrations and monitoring HAB growth by utilizing the unique nature of these pigments' interaction with light.

Additionally, measures of sediment content, like TSS and turbidity, are essential for determining the health of a freshwater ecosystem. Elevated TSS levels usually follow heavy rainfall events and are due to the erosion of sediments from agricultural lands, degraded stream banks, overgrazed rangelands, logged landscapes, mined areas, and zones of construction. Turbidity is directly related to TSS levels and is often used to infer sediment concentrations (Terrio *et al.* 2015). Excess concentrations of suspended sediment can transport inorganic nutrients (i.e., sorbed phosphorus) from agricultural landscapes and may exacerbate algal growth in freshwater systems.

The TDS level, which is the sum of all dissolved ions and organic matter present in a water sample, is an important indicator of overall water-quality. The main constituents are usually calcium, magnesium, sodium, and potassium cations, and bicarbonate, chloride, sulfate, and nitrate anions (DeZuane 1997). TDS poses a distinct challenge for remote sensing detection, as it is not optically active, and therefore traditional remote sensing approaches based on IOPs are not effective. Thus, nontraditional approaches such as indirect detection and estimation based on other water properties that may correlate with TDS (e.g., TSS and turbidity) may prove successful (Lewis and Saunders 1989).

Due to the spectral complexity of inland waters, their small spatial extent, and the degree of human impact, the monitoring of water bodies requires high spatial and spectral resolutions. With recent advances in sensor technology, hyperspectral data have become more available than ever before; both the sensors and data can contain hundreds of spectral bands with high spectral resolution, versus multispectral sensors that contain only several bands. Hyperspectral imaging technology potentially provides an improved approach for water-quality monitoring, compared to traditional multispectral data-processing methods that have been limited by the lack of spectral resolution. However, this new technological opportunity comes with inherent methodological challenges. Hyperspectral and imaging spectroscopy data include both highly correlated and noisy spectral bands, and frequently create statistical difficulties due to a small number of useful bands compared to the large number of available ones (Verrelst *et al.* 2016).

To resolve the issue of low signal-to-noise ratios and the high dimensionality of spectroscopy data, information-fusion techniques provide an opportunity to create more informative features and reduce the dimensionality of spectral data. Data fusion is a broad term for methods that use multiple sets of data obtained from various sources (e.g., different sensors or bands) to improve some performance metric (Castanedo 2013). The use of data fusion in remote sensing has been proven to enhance the predictive ability of regression models mainly using sensor and image fusion (Brezonik *et al.* 2005; Pohl and Van Genderen 1998). Spectroscopic data sets are often sampled in a spatial context, therefore resulting in highly intercorrelated spectral-band information (Feilhauer *et al.* 2015). Feature-level fusion techniques take unique characteristics of each individual feature and combine them to create a single, more informative feature. In fields such as computer vision, feature-level fusion techniques are relied upon as highly useful methods for eliminating dimensionality problems and reducing noise (Rattani *et al.* 2007). Other studies (e.g., Peterson *et al.* 2018; Zhong *et al.* 2016) have successfully demonstrated the application of feature-fusion techniques to hyperspectral remote sensing data, indicating that a number of different fusion approaches can significantly increase the predictive ability of models. The use of fusion in a regression problem also serves as a dimension-reduction tool, providing a more computationally effective approach.

Empirical correlations between spectral measurements and levels of water-quality variables can be determined using statistical regression models. Popular parametric regression methods, such as partial least-squares regression (PLSR), have demonstrated sufficient results in the estimation of optically active water-quality variables such as Chl-a and TSS (Song *et al.* 2014; Wang *et al.* 2017). Like PLSR, other parametric models such as least-squares (Zheng *et al.* 2015) and multilinear regression (MLR; Brezonik *et al.* 2005) have been investigated in the literature, yet these models are often too simplistic and unable to capture the complex relationships between remote sensing data and the constituents of the water body. This has led to the introduction of machine learning (ML) and nonlinear regression techniques in the field of remote sensing. Nonparametric methods, such as Gaussian process regression (GPR), have been used to predict variable concentrations where linear regression techniques fail to achieve satisfactory results (Pasolli *et al.* 2010; Verrelst *et al.* 2012). Other nonparametric ML regression methods such as support vector regression (SVR; Mountrakis *et al.*, 2011) and extreme learning machine regression (ELR; Huang *et al.* 2012; Peterson *et al.* 2018; Sidike *et al.* 2017) have been shown to capture complex relationships in both vegetation and water remote sensing studies by significantly improving model correlations (Camps-Valls *et al.*, 2006; Maimaitijiang *et al.* 2017; Mountrakis *et al.* 2011).

Along with ML regression approaches, ensemble regression or decision-level fusion has gained popularity in the remote sensing community. Decision-level fusion in a regression problem is simply the integration of multiple models and model outputs to produce a single model (Pohl and Van Genderen 1998). This method leverages the strengths and limits the potential biases of utilizing a single modeling technique. Feilhauer *et al.* (2015) have proposed that decision-level fusion can create more stable predictions in regression problems. The use of decision-level fusion may also increase the transferability of models and capture a wide range of correlations that are not present in models relying on a single regression technique (Ghulam *et al.* 2011).

The primary objectives of this study are to accurately model relationships between water-quality variables and spectral reflectance, implement and evaluate feature-level fusion of spectral data for water-quality modeling, and evaluate the effectiveness of various ML regression techniques and decision-level fusion for prediction of water-quality variables.

Study Area

The study area consisted of three independent Midwestern water bodies (Figure 1): two reservoirs in Illinois—Carlyle Lake (an impoundment of the Kaskaskia River) and Lake Decatur (an impoundment of the Sangamon River)—and sections of the unimpounded Meramec River in Missouri. The three water bodies were selected based on surrounding watershed land cover/land use and geographic location to create a regional-scale study. Sampling sites for each watershed were selected to maximize the spatial distribution within the water body for the most representative sample of water conditions across the study areas.

Carlyle Lake is one of two large reservoirs along the Kaskaskia River, which is located in south-central Illinois and drains into the Mississippi River. The Kaskaskia River watershed is 14,152 km^2 of predominantly agricultural (67%) land, largely consisting of corn and soybean fields (Chiang *et al.* 2014). The reservoir has a total surface area of 105 km^2, with a mean depth of 3.4 m (Romano *et al.* 2009). Carlyle Lake was created in 1967 and provides flood control, recreation, and water supply for much of the surrounding region. The reservoir is composed of two main sections: a shallow northern region (mean depth < 1 m) located near the Kaskaskia River entrance, and a deeper, main lake region to the south.

Lake Decatur, located in central Illinois, is the largest reservoir along the Sangamon River. The reservoir has a total surface area of 11 km^2, with a mean depth of 2.4 m (Kohler *et al.* 1993). Lake Decatur's watershed is 2396 km^2 and, similar to Carlyle Lake, is dominated by cultivated agricultural land-use practices. Constructed in 1922, the reservoir is the main source of water for the surrounding areas, including the city of Decatur, Illinois, and the local bioenergy refining industry. Over the last several decades, Lake Decatur has experienced extensive siltation issues due to soil erosion and runoff from agricultural fields within the Sangamon River watershed (Kohler *et al.* 1993). Lake Decatur and Carlyle Lake represent watersheds that are highly affected by intensive agriculture, thus creating ideal conditions for algal blooms due to an overabundance of nitrogen and phosphorus combined with shallow depths, calm surface waters, and warm temperatures (Pearce *et al.* 2017).

The Meramec River, located in east-central Missouri, was chosen for this study because it is one of the few large, unimpounded waterways in the United States and contains a much lower percentage of agricultural land use (22% agriculture, 68% forest) than the other two locations. It also represents a free-flowing fluvial environment (compared to the two lentic systems), providing a wider range in water-quality parameters for this study. The Meramec River watershed encompasses a total catchment area of 10,300 km^2

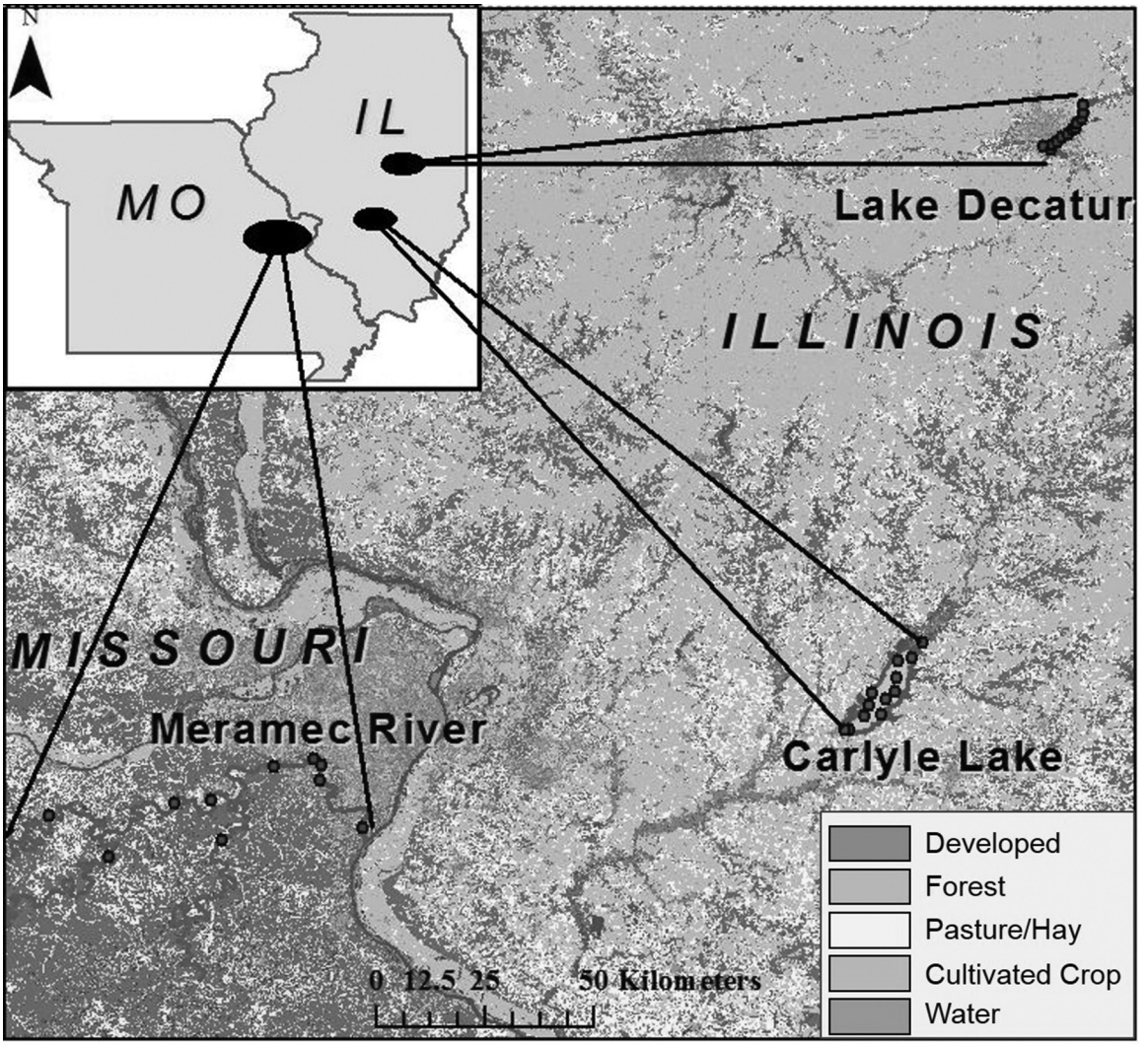

Figure 1. Map of the study area and sampling locations. The map also details the regional land cover/land use.

and includes two smaller rivers, the Bourbeuse River and the Big River (Hasenmueller and Criss 2013). Municipal water districts, including cities within the St. Louis region, utilize the river for drinking water and household use, making water-quality monitoring crucial.

Data

Data for this study consisted of *in situ* spectroscopy as well as water-quality measurements obtained using *in situ* sensors and lab analysis of discrete water samples (Table 1). Field sampling was conducted between June 2016 and November 2017 to gather water-quality data under a range of seasons and atmospheric conditions. A total of 105 samples were collected (73 during summer and 32 during autumn), comprising 74 samples from Carlyle Lake, 12 from Lake Decatur, and 19 from along the Meramec River.

Table 1. Descriptive statistics for all water-quality variable measurements.

Statistic	BGA-PC (μg/L)	Chl-a (μg/L)	TSS (mg/L)	Turbidity (FNU)	TDS (mg/L)
n	96	105	93	96	86
Minimum	0.1	0.5	1.0	0.4	8.0
Maximum	6.0	86.4	183.0	179.0	384.0
Mean	2.3	20.8	32.1	24.9	226.8
Variance	2.5	284.1	909.3	803.9	5834.2
Standard deviation	1.6	16.9	30.2	28.4	76.4

BGA-PC = blue-green algae phycocyanin; Chl-a = chlorophyll a; TSS = total suspended solids; FNU = formazin nephelometric units; TDS = total dissolved solids.

Field spectroscopy data were collected using the Spectral Evolution PSR-3500 (Spectral Evolution, Inc., Lawrence, Massachusetts) handheld spectroradiometer. This instrument has a spectral range of 350–2500 nm, with a spectral resolution of 3.5 nm in the 350–1000-nm range, 10 nm in the 1000–1900-nm range, and 7 nm in the 1900–2500-nm range. Spectral-reflectance measurements were captured between 10:00 A.M. and 2:00 P.M. with clear skies from a 1-m height above the water surface at nadir. Three separate water-surface reflectance samples were collected at each location and averaged to reduce noise in the data.

Initial processing of raw spectral data was conducted by interpolating spectral-reflectance values at 1-nm intervals to cover the full 350–2500-nm spectrum. The 1-nm intervals were used in this study to retain maximum spectral information at narrow wavelengths. The data were then normalized by removing excess noise and spectra within the 1700–2000-nm and 2350–2500-nm ranges.

Immediately following spectroscopy, surface-water samples were analyzed *in situ* using a YSI EXO2 sonde or YSI ProPlus Multiparameter handheld meter (Source: Yellow Springs Instruments, Yellow Springs, Ohio) with sensors for Chl-a, BGA-PC, TDS (calculated from specific conductance), and/or turbidity. In the Meramec River watershed, turbidity was measured with a Hach Turbidimeter (Hach Company, Loveland, Colorado). A Van Dorn sampler was used to collect discrete surface-water samples that were returned to the lab for determination of TSS and Chl-a. The water samples were collected in acid-washed polyethylene bottles and stored on ice until being returned to the lab; all samples were processed within 16 hr of collection. Sample TSS content was determined via EPA method 160.2 (Environmental Protection Agency 1971). Briefly, samples were vigorously mixed before analysis to resuspend sediments, then 100-mL aliquots were passed through glass-fiber filters using a vacuum-filtration device. The filters were then carefully removed and dried overnight at 105°C. Chl-a and BGA-PC were measured *in vivo* with a dual-channel YSI Total Algae PC Smart Sensor integrated into an EXO2 sonde (Yellow Springs Instruments, Yellow Springs, Ohio). Combining the *in situ* measurements and lab results, we developed a water-quality data set with corresponding reflectance measurements. The spectral and water-quality data set was divided such that 60% of the data were used for model training and 40% were used for testing before further analysis.

Methods

The methods used in this study (Figure 2) involve several stages, beginning with the compilation of a pair-wise data set containing water-quality measurements and spectral reflectance. Feature-level fusion is then performed on the spectral-reflectance data to eliminate noise, reduce dimensionality, and create a single, highly predictive reflectance variable. The fused spectral data and water-quality measurements are then inputted into four different regression models, and their outputs are combined using decision-level fusion.

Feature-Level Fusion

To combat issues related to the high dimensionality and redundancy of hyperspectral data, feature-level fusion is applied to the input raw spectral-reflectance values obtained from handheld spectroradiometry. In this study, we employed a form of CCA to fuse raw spectral features. This information-fusion method, based on work by Sun *et al.* (2005), extracts canonical correlation features between multiple feature vectors to form a single highly discriminant feature. The CCA technique has been found to work effectively when applied to complex water-quality-based spectral studies, as displayed by Peterson *et al.* (2018). The CCA method and algorithmic description are detailed by Sun *et al.* (2005) and Peterson *et al.* (2018).

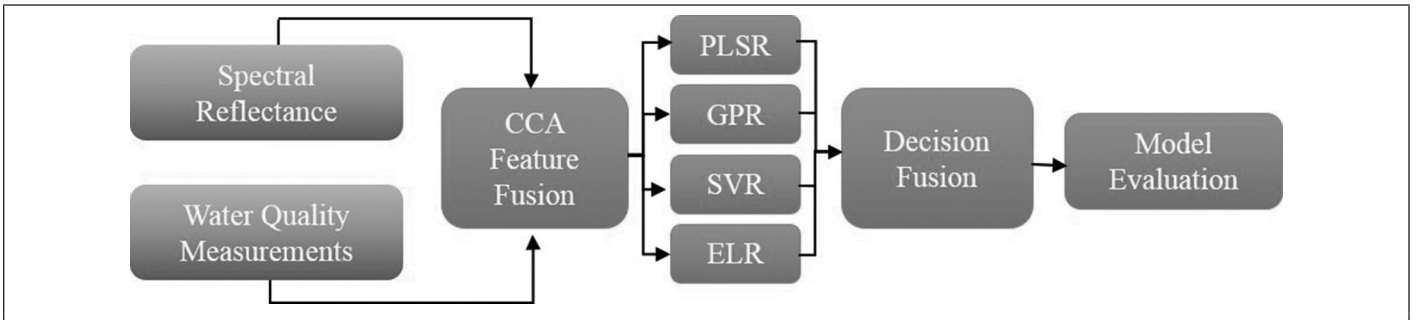

Figure 2. Visualized workflow and methodology.

In this study, CCA feature-level fusion was performed on the spectral-reflectance data to eliminate excess noise, reduce the dimensionality of the spectral data, and capture spectral trends not inherent in the raw data. CCA fusion has been found to outperform other feature-level fusion techniques, such as singular-value decomposition and principal component analysis-based algorithms (Sun *et al.* 2005). The fused spectral feature vectors (Z) were then used as inputs, along with corresponding water-quality variable concentrations, for the regression models.

Regression Modeling

For the regression modeling, five ML-based techniques were applied, including both parametric and nonparametric models; they are MLR, PLSR, GPR, SVR, and ELR. The use of MLR in this study was to serve as a baseline for comparison among other ML-based techniques. PLSR was included as a parametric regression technique due to its recent popularity in remote sensing studies. This regression approach, first utilized in chemometrics and widely used in vegetation remote sensing studies, has been proven to effectively link spectroscopy data to nutrient and biochemical concentrations in tree and plant canopies (Gökkaya *et al.* 2015). Recently, methods within the Bayesian nonparametric family have gained popularity. Using an implementation of GPR, Lázaro-Gredilla *et al.* (2014) predicted concentrations of both vegetative and oceanic Chl-a with high accuracy based on *in situ* and airborne hyperspectral data. Several studies, including Feilhauer *et al.* (2015), have successfully utilized SVR in spectroscopic remote sensing-based regressions to predict leaf biochemical traits. Neural network-based regressions such as ELR have also been applied to predict biophysical traits in plants, and ELR has shown promise by outperforming the parametric methods that have been traditionally used (Maimaitijiang *et al.* 2017). A brief review of these regression models follows.

The general PLSR model is formulated as

$$y = B_1 x_1 + B_2 x_2 + \ldots B_i x_i + \varepsilon \tag{1}$$

where the response variable y is a vector of the water-quality variable concentrations, $i - 1$ is the spectral wavelength, x is the water-surface reflectance, B is the weighted regression coefficient, and ε is the residual-error vector (Wold *et al.* 2001).

GPR is a Bayesian approach which assumes a Gaussian behavior of function values. The basic GPR model is expressed as

$$f(x) \sim N\left(0, K\left(\theta, x, x'\right)\right) \tag{2}$$

where x is the input vector, N is the number of data points, and $K(\theta, x, x')$ is the covariance matrix between all possible pairs (x, x') for a given set of hyperparameters θ. The GPR model used in this study employed a linear kernel function and quasi-Newton optimizer (Nocedal and Wright 2006) to tune the hyperparameters.

SVR was first identified by Vapnik (1992) and is a popular ML algorithm for regression-based remote sensing studies due to its robust ability to capture nonlinear trends. SVR is considered a nonparametric technique, as it relies on kernel functions. To train the SVR model, the following minimization problem must be solved:

$$\text{minimize}\,\frac{1}{2}\|w\|^2, \text{ subject to} \begin{cases} y_i - [w, x_i] - b \le \varepsilon \\ [w,\ x_i] + b - y_i \le \varepsilon \end{cases} \tag{3}$$

where x_i is a training sample with the target value y_i, b is a real constant, and w is a measure of error. The inner product plus intercept $[w, x_i] + b$ is the prediction for that sample, and ε is a threshold parameter whose predictions must all be within range (Vapnik, 1992). This implementation of SVR also utilized a Bayesian optimization function (Pelikan *et al.* 1999) for optimized hyperparameter tuning, where each parameter of the SVR function was tuned based on the results of previous iterations.

ELM is a single-hidden-layer feed-forward neural network that contains a single input layer, one hidden layer, and one output layer. ELM was included as a cutting-edge regression technique, as it has only recently been applied to remote sensing studies and has been shown to outperform other ML techniques when applied to water-quality parameter estimation (Peterson *et al.* 2018). The weights of the hidden layer within ELR can be randomly produced without iterative optimization (Huang *et al.* 2006), leading to significantly less computational time in training a model. To create the model, distinct pairs of samples $\{x_i y_i\}_{i=1}^N$ are selected from a given training set of N input vectors $x_i \in \mathbb{R}^d$ with the corresponding N output values $\{y_i\}_{i=1}^N$. Using ELR, the goal is to determine the relationship between x_i and the desired output y_i. To determine this relationship, a cost function for L hidden nodes is denoted as

$$\min \| y_i - \widehat{y_i} \|_{i=1}^N = \min \sum_{i=1}^N (y_i - \sum_{j=1}^L \beta_j h_j (w_j \times x_i + b_j)) \tag{4}$$

where $\hat{y}_i$ is the predicted output. The j output weight vector, denoted as β_j, is the output weight and links the j hidden node and the output node. $\mathcal{R}(x_i) = g(w_j, b_j, x_i)$ is the output of the j hidden node with respect to the input $x_i g(.)$, which is a nonlinear piecewise function (e.g., sigmoid function); $w_j \in \mathbb{R}^d$, the weight vector; and b_j, the bias of the j hidden node (Maimaitijiang *et al.* 2017). To train a traditional single-hidden-layer feed-forward network, hidden-layer parameters (j, b_j) are optimized through gradient-descent or global search methods, but in ELR these are randomly generated without iterative tuning, making the learning process much faster and capable of processing large amounts of data. A detailed description of the ELM method can be found in Huang *et al.* (2006).

The Proposed Decision-Level Fusion

Once all of the models were run individually, the outputs of each model in the testing stage were fused using the weights obtained from the training stage. In this study, we propose a weighted-prior (WP) method for decision-level fusion (DLF). The WP fusion method automatically assigns weights to the fusion model based on the results generated from the training data set's models. Once individual regression models have been run on the training data, the normalized absolute difference between the models' predicted values $\hat{y}_{t,i}$ and observed values y_t are calculated as

$$\boldsymbol{D}_i = \frac{|y_t - \hat{y}_{t,i}|}{|y_t + \hat{y}_{t,i}|}, \boldsymbol{D}_i \in \mathbb{R}^N \tag{5}$$

where $i = 1,2,\ldots, l$ indicates the ith model used for DLF, l represents the total number of models, t signifies the data from the training phase, and N is the total number of samples in the training set. Lower values of $\boldsymbol{D}_i$ indicate less divergence from the observed values, and thus better predicted results of a model. To find a weight that can be used for each model in the testing phase, we first computed the mean of $\boldsymbol{D}_i$, denoted by $m_{\boldsymbol{D}_i}$, as

$$m_{\boldsymbol{D}_i} = \frac{\sum \boldsymbol{D}_i}{N}. \tag{6}$$

Then the associated weight variable w_i was obtained by

$$w_i = \frac{1 - m_{D_i}}{\sum_{i=1}^{I}\left(1 - m_{D_i}\right)}. \quad (7)$$

As in the consensus-theory weight-selection scheme, w_i represents the goodness of fit for each separate training model (Benediktsson and Kanellopoulos 1999). We subtracted m_{D_i} from 1 to assign higher weight values to models which yielded better prediction in the training phase, and vice versa. The resulting weight variables were applied to the predicted values $\hat{y}_{s,i}$ of the testing data for each respective model and then summated to create a single fused prediction, expressed as

$$y_{\mathrm{WP}} = \sum_{i=1}^{I} w_i \hat{y}_{s,i}, \quad (8)$$

where y_{WP} is the final predicted result of the proposed WP method, and the subscript s represents the data from the testing phase. The aim of the DLF in this study is to reduce the model's bias introduced by utilizing a single technique and to increase the potential transferability of the resulting model.

Model Evaluation

The results of all regression models were evaluated using the R^2 correlation coefficient and the root-mean-square error (RMSE). The RMSE statistic is calculated as

$$\mathrm{RMSE} = \frac{\sqrt{\Sigma_{i=1}^{S}(\check{y}_i - y_i)^2}}{S} \quad (9)$$

where y_i and $\check{y}_i$ are the measured and predicted parameters and S is the total number of testing samples. Smaller RMSE

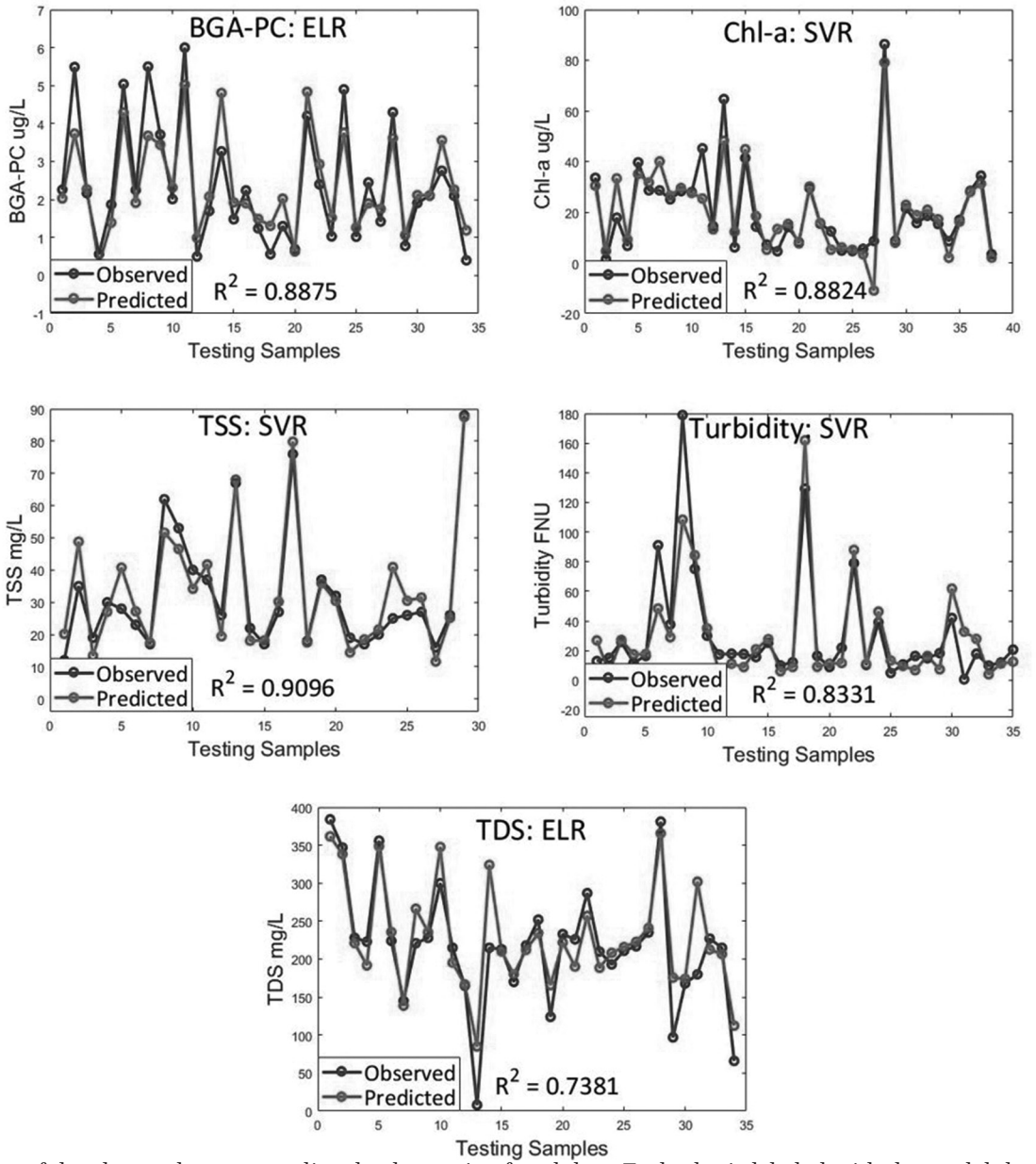

Figure 3. Plots of the observed versus predicted values using fused data. Each plot is labeled with the model that generated the best results.

values indicate a higher overall accuracy of the model. The coefficient R^2 represents the proportion of variation in the responses that is explained by the original model using predictor values from the test data, and it indicates overall accuracy of the model. Higher R^2 values indicate a higher overall correlation of the model. Each model used a randomized data partitioning of 60% training and 40% testing data. The regression functions and all data manipulation were conducted using MATLAB R2017a.

Spectral-Band Importance

To determine the impact of each wavelength or spectral band on the regression model, a variable-importance analysis was conducted by using a single-band reflectance value as the input feature to predict the corresponding water-quality variable concentration. This was done using ELR to capture both linear and nonlinear trends in the spectral data. Testing each band individually in ELR results in R^2 values that serve as a measure of the relative importance of the given spectral band.

Table 2. Results of the regression models using nonfused spectral data.

	BGA-PC	Chl-a	TSS	Turbidity	TDS
MLR					
Training					
R^2	.566	.3485	.7158	.3894	.3377
RMSE	0.992	32.2781	19.9445	31.0821	83.8559
Testing					
R^2	.428	.2158	.683	.3571	.3188
RMSE	1.846	34.9833	20.5647	33.8972	88.8973
PLSR					
Training					
R^2	.6612	.7071	.7957	.627	.4994
RMSE	0.9162	12.2005	14.1169	16.861	53.8638
Testing					
R^2	.4813	.4702	.7272	.2684	.0707
RMSE	1.1216	9.1673*	17.0546	26.8159*	71.6659
GPR					
Training					
R^2	.8987	.9478*	.9429*	.8459	.994*
RMSE	0.5302	3.6786	6.1561*	8.9587	0.0199*
Testing					
R^2	.4036	.4022	.6924	.2867	.1131
RMSE	1.0606*	13.5432	15.8975	28.899	66.818
SVR					
Training					
R^2	.9426	.5204	.7665	.6893	.6513
RMSE	0.9426	12.1642	14.8548	9.163	56.3449
Testing					
R^2	.4568	.4982*	.7415	.367	.3492*
RMSE	1.1269	13.6431	17.9005	32.187	65.3449*
ELR					
Training					
R^2	.9509*	.9333	.9323	.9516*	.7792
RMSE	0.2201*	2.7602*	8.6202	4.387*	29.938
Testing					
R^2	.5695*	.4857	.8126*	.4818*	.3138
RMSE	1.1353	12.6907	10.3026*	27.517	67.1659
Mean					
Testing					
R^2	.442425	.393475	.711025	.3198	.2041
RMSE	1.30745	18.71508	17.85433	30.62505	73.63678

BGA-PC = blue-green algae phycocyanin; Chl-a = chlorophyll a; TSS = total suspended solids; TDS = total dissolved solids; MLR = multilinear regression; RMSE = root-mean-square error; PLSR = partial least-squares regression; GPR = Gaussian process regression; SVR = support vector regression; ELR = extreme learning machine regression. *Best model results for each variable.

Results

Regression-Model Results

The regression models using the nonfused features generated relatively low correlation statistics and high RMSE values for all variables and all modeling techniques (Table 2). Model results for TSS exhibited the strongest overall correlations, explaining an average of 71% of the observed values from the testing data set. The BGA-PC models generated an average R^2 of .442 but showed the lowest RMSE values, with a mean of 1.307. Models generated for all other variables displayed relatively weak correlations, with $R^2 \leq .5$. With the low correlation values and high overall RMSE values, the data sets appeared to be plagued by excess noise, thus decreasing the accuracy of the models. Nevertheless, ELR performed the best on the testing data, followed by SVR. In general, the results produced by the ELR and GPR models for nonfused data have very high R^2 values for training data and much lower ones when applied to testing data. This is likely due to overfitting of the models and could potentially be reduced by fine-tuning of the respective function parameters.

Predictive accuracy for all variables showed improvements after application of CCA feature-level fusion (Table 3). TSS again showed the highest predictive power, explaining 87%

Table 3. Results of the regression models using fused spectral data.

	BGA-PC	Chl-a	TSS	Turbidity	TDS
MLR					
Training					
R^2	.7826	.6147	.8715	.6483	.6281
RMSE	1.0946	7.8496	13.6582	13.8477	25.489
Testing					
R^2	.6754	.5098	.7894	.5371	.4984
RMSE	1.5867	8.499	16.9489	17.9491	58.4197
PLSR					
Training					
R^2	.9648	.9886	.9983	.984	.9969
RMSE	0.3462	2.0597	0.089	0.249	0.6814
Testing					
R^2	.8545	.8494	.8842	.8094	.6998
RMSE	0.9168	4.1879	12.987	8.678	40.689
GPR					
Training					
R^2	.968*	.9906	.9984	.991	.9945
RMSE	0.3157	1.019	0.3579	0.3687	0.018*
Testing					
R^2	.8744	.8568	.8859	.8146	.712
RMSE	0.8861	4.593	7.067	8.2546	34.5258
SVR					
Training					
R^2	.9648	.987	.9989*	.999*	.9994*
RMSE	0.1674	0.1683*	0.0782*	0.2873	0.2188
Testing					
R^2	.8728	.8824*	.9096*	.8331*	.7358
RMSE	0.799	5.8741	6.0548*	7.9471*	32.5177
ELR					
Training					
R^2	.9678	.9907*	.9988	.989	.995
RMSE	0.051*	0.2688	0.1964	0.2166*	0.9875
Testing					
R^2	.8875*	.8572	.9019	.8295	.7381*
RMSE	0.8673*	1.3857*	10.5909	8.2909	30.5499*
Mean					
Testing					
R^2	.819275	.7683	.86535	.74765	.6615
RMSE	1.04715	5.7885	11.89845	10.79315	41.53805

BGA-PC = blue-green algae phycocyanin; Chl-a = chlorophyll a; TSS = total suspended solids; TDS = total dissolved solids; MLR = multilinear regression; RMSE = root-mean-square error; PLSR = partial least-squares regression; GPR = Gaussian process regression; SVR = support vector regression; ELR = extreme learning machine regression. *Best model results for each variable.

of the observed measurements in the testing data, followed by BGA-PC (82%), Chl-a (77%), turbidity (75%), and TDS (66%). Applying the feature-level fusion data, SVR and ELR displayed the best overall model results, while MLR and PLSR generated the worst results in terms of R^2 and RMSE values. This is likely due to the high complexity of the water–light interactions captured in the spectral data for each given variable. This is supported by the fact that the least predictive models were generated using the parametric MLR and PLSR techniques, while the most predictive models were generated by nonparametric techniques such as SVR and ELR that capture nonlinear trends or correlations in the data.

The application of CCA feature-level fusion, on average, increased model R^2 values by 112% and decreased RMSE values by 46% when compared to nonfused spectral data

Table 4. Mean model improvements (%) as a result of the canonical correlation analysis feature-level fusion. Mean improvement was calculated by taking the mean testing results of each variable and finding the difference between the mean nonfused and fused testing results.

	BGA-PC	Chl-a	TSS	Turbidity	TDS
R^2	85.18	95.26	21.7	133.79	224.11
RMSE	−19.91	−69.07	−33.63	−64.76	−43.59

BGA-PC = blue-green algae phycocyanin; Chl-a = chlorophyll a; TSS = total suspended solids; TDS = total dissolved solids; RMSE = root-mean-square error.

Table 5. Decision-level fusion (DLF) testing results and improvement over the mean results of the feature-level fusion models.

	BGA-PC	Chl-a	TSS	Turbidity	TDS
DLF-Mean					
R^2	.8672	.8545	.8907	.8178	.7159
RMSE	0.8673	4.885	10.215	8.4078	35.9108
DLF-WP					
R^2	.8838	.8646	.8972	.8249	.7224
RMSE	0.8516	4.5948	9.378	7.9667	32.4937
Mean improvement (%)					
R^2	1.91	1.18	0.73	0.87	0.91
RMSE	−1.18	−5.94	−8.19	−5.25	−9.52

BGA-PC = blue-green algae phycocyanin; Chl-a = chlorophyll a; TSS = total suspended solids; TDS = total dissolved solids; DLF-Mean = arithmetic mean of the fused testing model results; RMSE = root-mean-square error; DLF-WP = weighted prior of the fused testing model results.

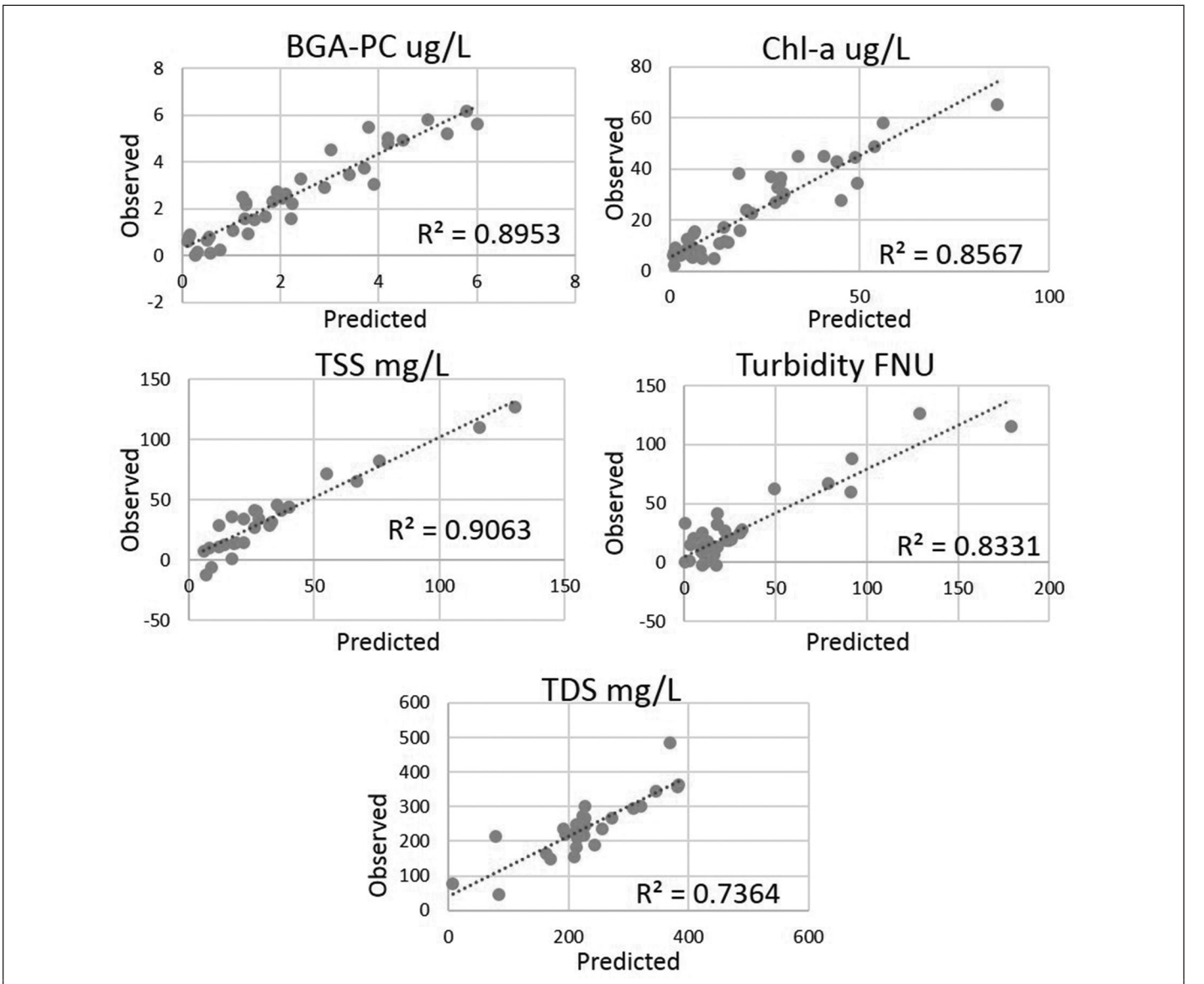

Figure 4. Plots of the observed values versus the values predicted by weighted-prior decision-level fusion for the various water-quality parameters.

(Table 4). Feature-level fusion improved the TDS results the most, increasing the average R^2 value by 224% and decreasing the average RMSE value by 44%. The CCA fusion technique had the greatest impact on lower-signal-to-noise data sets such as TDS and turbidity. Nevertheless, the method also increased the accuracy of results for higher-signal-to-noise ratio data sets (TSS, BGA-PC, and Chl-a).

Decision-Level Fusion-Model Results

The performance of the WP decision-level fusion models shows that the proposed method was better than the arithmetic mean. Indeed, R^2 values on average increased by 1% and RMSE values decreased by 7% (Table 5). Figure 4 shows the plotted results of the WP fusion method. Although the WP fusion method showed only minor improvement over simple mean fusion, the WP method gave considerably higher R^2 values than the least predictive single model and slightly lower than the highest model for each variable. This decision fusion displays the best results when applied to the variables with the highest variance of model outputs, such as TDS and turbidity. WP decision fusion effectively improved model correlations and eliminated intermodal variance for all variables.

Spectral-Band Importance

ML methods demonstrated strong competence in regression analysis and modeling. However, it is challenging to connect models from ML algorithms to remote sensing. For this reason, we investigated the contribution of each spectral region to the model prediction using a spectral-band importance analysis (Figure 5). As expected, the results for BGA-PC indicate that the most informative wavelengths or spectral bands were located between 400 and 450 nm and near 560, 615 (i.e., phycocyanin absorption), and 670 nm. Results for Chl-a showed that the bands at 450 (i.e., Chl-a absorption), 550, 555, 690 (i.e., Chl-a absorption), 715, and 825 nm were the most informative. The resulting R^2 values confirm that BGA-PC and Chl-a share many of the same important spectral bands, as both variables can be identified through the chlorophyll absorption in the red and near-infrared regions. Numerous studies have identified these key absorption regions for both BGA-PC and Chl-a; a summary is given by Stumpf *et al.* (2016). Although BGA-PC and Chl-a share many of the same key spectral bands and properties, BGA-PC can be separated from Chl-a using the phycocyanin absorption near 615 nm. The TSS results generated the highest

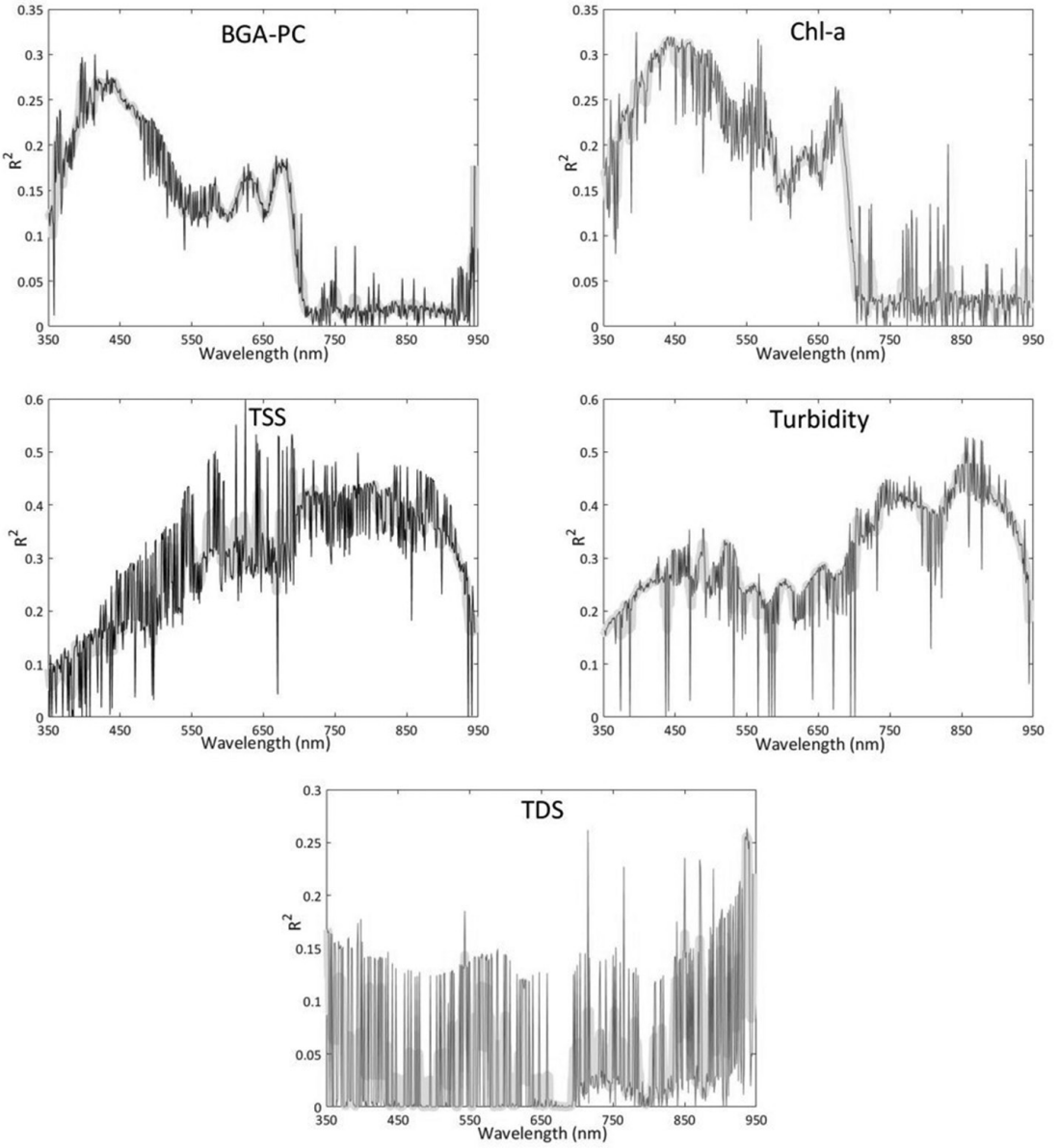

Figure 5. Results of the spectral-band importance analysis for each water-quality variable.

R^2 values near 600 nm, between 730 and 740 nm, and between 815 and 840 nm. Turbidity followed a similar pattern to TSS, displaying spectral importance around 450, 750, and 850 nm. These results coincide with previous studies that have utilized reflectance values in the near-infrared region to estimate sediment concentrations (Feng *et al.* 2014; Zheng *et al.* 2015). We hypothesize that TDS is indirectly estimated with our technique. This is supported by the band-importance analysis, which indicates a generally even distribution of R^2 scores but higher correlations in the red and near-infrared regions. Identifying the most important spectral bands for each variable helps us detect the underlying spectral properties that allow us to predict variable concentrations.

Discussion

The results of this study demonstrate that water-quality variable concentrations can be accurately predicted using spectral-reflectance models and data-fusion techniques. The models resulting from the fused spectral data can be used to predict water-quality variable levels for BGA-PC, Chl-a, TSS, turbidity, and TDS throughout the study area. This approach proved to be quite robust, creating accurate predictive models for a range of water-quality variables related to algae, sediment, water clarity, and dissolved matter. Utilizing data obtained from a range of temporal, spatial, and hydrological conditions, the developed models aimed at creating a more representative prediction of water-quality conditions. With the inclusion of TDS, a variable that was not optically active could also be predicted and modeled with the use of ML algorithms. We were able to observe a correlation in optical indicators between algae (i.e., BGA-PC and Chl-a concentrations) and TDS due to variable discharge. Indeed, both algae and TDS levels tend to decrease during floods and increase during low-water periods. This relationship in rivers could explain why we are able to predict TDS, despite the fact that this parameter is not optically active. Our results demonstrated that the use of nonparametric regression techniques have a significant benefit when applied to water-based spectral data.

The application of CCA feature-level fusion to spectral-reflectance data greatly increases the predictive power of all models for all variables. As seen in other studies (Jimenez *et al.* 1999), feature-level fusion techniques have demonstrated the ability to reduce noise in highly dimensional spectral data sets. Variables such as TDS and turbidity, with low signal-to-noise ratios, highlight this advantage by greatly improving model predictions. Based on these results and other similar studies, the CCA feature-level fusion method may serve as an ideal technique for working with noisy hyperspectral and spectroradiometer data. Moreover, the high dimensionality of hyperspectral data sets often creates both modeling and computational issues. With the use of feature-level fusion, we can reduce the spectral information from over 1000 features to just one highly informative feature. Other studies, such as by Kang *et al.* (2014), have displayed the ability of feature-level fusion to greatly reduce computational time and generate higher correlations by decreasing the number of variables in the equation.

Along with feature-level fusion, incorporating a decision-level fusion method into the modeling approach helps achieve more consistent and reliable results, as indicated in several previous remote sensing studies (Du *et al.* 2012; Jimenez *et al.* 1999). By enabling more consistent results, this methodology may enhance the potential transferability of the models to other water bodies, which has been a persistent issue related to water-quality remote sensing models. Decision-level fusion of multiple regression methods has also demonstrated promise for the predictive ability of remote sensing-based models and has even been referred to as a key methodology for future remote sensing studies (Benediktsson *et al.* 2007). Utilizing multiple modeling techniques and decision-level fusion provides an even greater benefit to water-based remote sensing studies because of the small sample sizes generally used in these studies due to the difficulty and expense of collecting and analyzing large quantities of water samples. Given the lower sample sizes ($n < 100$) for these studies, variation between model outputs is far greater than for studies with large data sets, thus creating the need for a unified and consistent model that can be provided via decision-level fusion. In this study, the proposed WP fusion method showed only minor improvements over a simple mean fusion technique. In theory, this would have a greater benefit in scenarios containing higher intermodal variance, as the weighting method is based on normalized absolute differences in predicted values.

In the future, this research and methodology will be applied at the airborne and satellite levels to test the applicability of the aforementioned methods at a larger scale. In theory, airborne hyperspectral data should produce similar results to those in this study. Scaling up to the satellite level will allow us to test model transferability. However, this may prove challenging, as the spatial, temporal, and radiometric resolution are far inferior to those of spectrometer-based data. Platforms such as the European Space Agency's (Lathrop and Lillesand 1989) *Sentinel 2* and *3* would likely be the best choice of satellite sensors, as they have band designations in key water-quality spectral locations and contain more bands at higher spatial resolutions than *Landsat 7* or *8*. Other ML regressions should also be tested within the ensemble methodology, such as neural networks and deep learning, and may prove superior to those used in this study. Further research should also focus on exploring other computer-vision and data-science fusion techniques that have not yet been explored in the context of remote sensing.

Conclusion

This study demonstrates that several key water-quality variables can be predicted with high relative accuracies using machine learning techniques including multiple linear regression, partial least-squares regression, Gaussian process regression, support vector machine regression, and extreme learning machine regression. A canonical correlation analysis feature-level fusion was developed for spectral analysis of water-quality variables. An ensemble forecasting method based on a decision-level fusion approach, combining the advantages of various ML techniques, was developed and proven to be most effective. Our main conclusions follow:

The results of the regression modeling highlighted the benefit of feature-level fusion applied to low-signal-to-noise spectral data. Combining feature-level fusion and ML regression with decision-level fusion increased the overall accuracy of the final model predictions by enhancing the quantitative evaluations and eliminating model biases. The application of CCA feature-level fusion, on average, increased model R^2 values by 85.18%, 95.26%, 21.7%, 133.79%, and 224.11%, respectively, for BGA-PC, Chl-a, TSS, turbidity, and TDS predictions.

Applying the CCA feature-level fusion technique in combination with the newly proposed weighted-prior decision-level fusion was found to improve the regression-modeling results. Thus, this method could be ideal for future water-based spectral modeling studies. The WP decision-level fusion models improved accuracy by 1.91%, 1.18%, 0.73%, 0.87%, and 0.91% for all water-quality variables while significantly eliminating intermodal variance.

The results of the spectral-band importance analysis identified the potential impact of each spectral band on the resulting models. As expected, results for BGA-PC indicated that bands located between 400 and 450 nm and near 560,

615 (i.e., phycocyanin absorption), and 670 nm had the highest correlations. Chl-a showed similar results to BGA-PC, with the highest importance at 450 (i.e., Chl-a absorption), 550, 555, 690 (i.e., Chl-a absorption), 715, and 825 nm. Results for TSS resulted in the highest R^2 values near 600 nm, between 730 and 740 nm, and between 815 and 840 nm. Turbidity followed a comparable pattern to TSS, displaying spectral importance around 450, 750, and 850 nm. TDS had a generally even distribution of R^2 scores, but with higher correlations in the red and near-infrared regions, where peaks seemed to mirror BGA-PC and Chl-a.

Acknowledgments

This study was made possible by funding from the Saint Louis University President's Research Fund, The Nature Conservancy (Award Number 061716-01), and the National Great Rivers Research and Education Center (NGRREC). We would like to thank all those who assisted in data collection and data analysis, including Miles Corcoran, Andrew Shaughnessy, David Pan, Camille Buckley, Mason Maimaitijiang, Matthew Maimaitiyiming, and Ethan Shavers. Data related to this study can be accessed by contacting the corresponding author.

References

Asselman, N. E. M. 1999. Suspended sediment dynamics in a large drainage basin: The River Rhine. *Hydrological Processes* 13 (10):1437–1450.

Benediktsson, J. A., J. Chanussot and M. Fauvel. 2007. Multiple classifier systems in remote sensing: From basics to recent developments. Pages 501–512 in *Mutiple Classifier Systems. MCS 2007*. Edited by M. Haindl, J. Kittler, and F. Roli. Lecture Notes in Computer Science vol. 4472. Berlin: Springer.

Benediktsson, J. A. and I. Kanellopoulos. 1999. Classification of multisource and hyperspectral data based on decision fusion. *IEEE Transactions on Geoscience and Remote Sensing* 37 (3):1367–1377.

Brezonik, P., K. D. Menken and M. Bauer. 2005. Landsat-based remote sensing of lake water-quality characteristics,including chlorophyll and colored dissolved organic matter (CDOM). *Lake and Reservoir Management* 21 (4):373–382.

Brooks, B. W., J. M. Lazorchak, M. D. A. Howard, M.-V. V. Johnson, S. L. Morton, D. A. K. Perkins, E. D. Reavie, G. I. Scott, S. A. Smith and J. A. Steevens. 2017. In some places, in some cases, and at some times, harmful algal blooms are the greatest threat to inland water-quality. *Environmental Toxicology and Chemistry* 36 (5):1125–1127.

Camps-Valls, G., L. Bruzzone, J. L. Rojo-Alvarez and F. Melgani. 2006. Robust support vector regression for biophysical variable estimation from remotely sensed images. *IEEE Geoscience and Remote Sensing Letters* 3 (3):339–343.

Carmichael, W. W. and G. L. Boyer. 2016. Health impacts from cyanobacteria harmful algae blooms: Implications for the North American Great Lakes. *Harmful Algae* 54:194–212.

Castanedo, F. 2013. A review of data fusion techniques. *The Scientific World Journal* 2013:704504.

Chiang, L.-C., Y. Yuan, M. Mehaffey, M. Jackson and I. Chaubey. 2014. Assessing SWAT's performance in the Kaskaskia River watershed as influenced by the number of calibration stations used. *Hydrological Processes* 28 (3):676–687.

DeZuane, J. 1997. *Handbook of Drinking Water-quality*. New York: John Wiley & Sons.

Du, P., J. Xia, W. Zhang, K. Tan, Y. Liu and S. Liu. 2012. Multiple classifier system for remote sensing image classification: A review. *Sensors* 12 (4):4764–4792.

Duan, W., K. Takara, B. He, P. Luo, D. Nover and Y. Yamashiki. 2013. Spatial and temporal trends in estimates of nutrient and suspended sediment loads in the Ishikari River, Japan, 1985 to 2010. *Science of the Total Environment* 461–462:499–508.

Feilhauer, H., G. P. Asner and R. E. Martin. 2015. Multi-method ensemble selection of spectral bands related to leaf biochemistry. *Remote Sensing of Environment* 164:57–65.

Feng, L., C. Hu, X. Chen and Q. Song. 2014. Influence of the Three Gorges Dam on total suspended matters in the Yangtze Estuary and its adjacent coastal waters Observations from MODIS. *Remote Sensing of Environment* 140:779–788.

Gholizadeh, M. H., A. M. Melesse and L. Reddi. 2016. A comprehensive review on water-quality parameters estimation using remote sensing techniques. *Sensors* 16(8):1298.

Ghulam, A., T. M. Kusky, T. Teyip and Q. Qin. 2011. Sub-canopy soil moisture modeling in *n*-dimensional spectral feature space. *Photogrammetric Engineering and Remote Sensing* 77 (2):149–156.

Gökkaya, K., V. Thomas, T. L. Noland, H. McCaughey, I. Morrison and P. Treitz. 2015. Prediction of macronutrients at the canopy level using spaceborne imaging spectroscopy and LiDAR data in a mixedwood boreal forest. *Remote Sensing* 7 (7):9045–9069.

Haghighat, M., M. Abdel-Mottaleb and W. Alhalabi. 2016. Fully automatic face normalization and single sample face recognition in unconstrained environments. *Expert Systems with Applications* 47:23–34.

Hasenmueller, E. A. and R. E. Criss. 2013. Water balance estimates of evapotranspiration rates in areas with varying land use. In *Evapotranspiration: An Overview,* edited by S. Alexandris, 1–22. London: IntechOpen.

Huang, C., X. Chen, Y. Li, H. Yang, D. Sun, J. Li, C. Le, L. Zhou, M. Zhang and L. Xu. 2015. Specific inherent optical properties of highly turbid productive water for retrieval of water-quality after optical classification. *Environmental Earth Sciences* 73 (5):1961–1973.

Huang, G.-B., H. Zhou, X. Ding and R. Zhang. 2012. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 42 (2):513–529.

Huang, G.-B., Q.-Y. Zhu and C.-K. Siew. 2006. Extreme learning machine: Theory and applications. *Neurocomputing* 70 (1–3):489–501.

Jimenez, L. O., A. Morales-Morell and A. Creus. 1999. Classification of hyperdimensional data based on feature and decision fusion approaches using projection pursuit, majority voting, and neural networks. *IEEE Transactions on Geoscience and Remote Sensing* 37 (3):1360–1366.

Kane, D. D., J. D. Conroy, R. P. Richards, D. B. Baker and D. A. Culver. 2014. Re-eutrophication of Lake Erie: Correlations between tributary nutrient loads and phytoplankton biomass. *Journal of Great Lakes Research* 40 (3):496–501.

Kang, X., S. Li and J. A. Benediktsson. 2014. Feature extraction of hyperspectral images with image fusion and recursive filtering. *IEEE Transactions on Geoscience and Remote Sensing* 52 (6):3742–3752.

Kohler, C. C., R. J. Sheehan and J. J. Sweatman. 1993. Largemouth bass hatching success and first-winter survival in two Illinois reservoirs. *North American Journal of Fisheries Management* 13 (1):125–133.

Kudela, R. M., S. L. Palacios, D. C. Austerberry, E. K. Accorsi, L. S. Guild and J. Torres-Perez. 2015. Application of hyperspectral remote sensing to cyanobacterial blooms in inland waters. *Remote Sensing of Environment* 167:196–205.

Kutser, T. 2009. Passive optical remote sensing of cyanobacteria and other intense phytoplankton blooms in coastal and inland waters. *International Journal of Remote Sensing* 30 (17):4401–4425.

Lathrop, R. G. Jr. and T. M. Lillesand. 1989. Monitoring water-quality and river plume transport in Green Bay, Lake-Michigan with Spot-1 imagery. *Photogrammetric Engineering and Remote Sensing* 55:349–354.

Lázaro-Gredilla, M., M. K. Titsias, J. Verrelst and G. Camps-Valls. 2014. Retrieval of biophysical parameters with heteroscedastic Gaussian processes. *IEEE Geoscience and Remote Sensing Letters* 11(4):838–842.

Lewis, W. M. Jr. and J. F. Saunders III. 1989. Concentration and transport of dissolved and suspended substances in the Orinoco River. *Biogeochemistry* 7 (3):203–240.

Maimaitijiang, M., A. Ghulam, P. Sidike, S. Hartling, M. Maimaitiyiming, K. Peterson, E. Shavers, J. Fishman, J., Peterson, S. Kadam, J. Burken and F. Fritschi. 2017. Unmanned Aerial System (UAS)-based phenotyping of soybean using multi-sensor data fusion and extreme learning machine. *ISPRS Journal of Photogrammetry and Remote Sensing* 134:43–58.

Mountrakis, G., J. Im and C. Ogole. 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (3):247–259.

Nocedal, J. and S. J. Wright. 2006. *Numerical Optimization,* 2nd ed. New York: Springer.

Ogashawara, I., D. R. Mishra, S. Mishra, M. P. Curtarelli and J. L. Stech. 2013. A performance review of reflectance based algorithms for predicting phycocyanin concentrations in inland waters. *Remote Sensing* 5 (10):4774–4798.

Pasolli, L., F. Melgani and E. Blanzieri. 2010. Gaussian process regression for estimating chlorophyll concentration in subsurface waters from remote sensing data. *IEEE Geoscience and Remote Sensing Letters* 7 (3):464–468.

Pearce, A. R., L. G. Chambers and E. A. Hasenmueller. 2017. Characterizing nutrient distributions and fluxes in a eutrophic reservoir, Midwestern United States. *Science of the Total Environment* 581–582:589–600.

Pelikan, M., D. E. Goldberg and E. Cantú-Paz. 1999. BOA: The Bayesian optimization algorithm. Pages 525–532 in *GECCO'99 Proceedings of the 1st Annual Conference on Genetic and Evolutionary Computation, Vol. 1,* held in Orlando, Fla., 13–17 July 1999. Edited by W. Banzhaf, J. M. Daida, M. H. Garzon and V. Honavar. San Francisco: Morgan Kaufmann.

Peterson, K. T., V. Sagan, P. Sidike, A. L. Cox and M. Martinez. 2018. Suspended sediment concentration estimation from Landsat imagery along the lower Missouri and middle Mississippi Rivers using an extreme learning machine. *Remote Sensing* 10 (10):1503.

Pohl, C. and J. L. Van Genderen. 1998. Multisensor image fusion in remote sensing: Concepts, methods and applications. *International Journal of Remote Sensing* 19 (5):823–854.

Rabalais, N. N. and R. E. Turner. 2017. Louisiana Universities Marine Consortion, Official Press Release.

Rattani, A., D. R. Kisku, M. Bicego and M. Tistarelli. 2007. Feature level fusion of face and fingerprint biometrics. In *2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems,* held in Crystal City, Va., 27–29 September 2007. Edited by J. Editor. City, St.: Publisher.

Romano, S. P., S. G. Baer, J. J. Zaczek and K. W. J. Williard. 2009. Site modelling methods for detecting hydrologic alteration of flood frequency and flood duration in the floodplain below the Carlyle Dam, lower Kaskaskia River, Illinois, USA. *River Research and Applications* 25 (8):975–984.

Sathyendranath, S., G. Cota, V. Stuart, H. Maass and T. Platt. 2001. Remote sensing of phytoplankton pigments: A comparison of empirical and theoretical approaches. *International Journal of Remote Sensing* 22 (2–3):249–273.

Sidike, P., E. Krieger, M. Z. Alom, V. K. Asari and T. Taha. 2017. A fast single-image super-resolution via directional edge-guided regularized extreme learning regression. *Signal, Image and Video Processing* 11 (5):961–968.

Song, K., L. Li, S. Li, L. Tedesco, H. Duan, Z. Li, K. Shi, J. Du, Y. Zhao and T. Shao. 2014. Using partial least squares-artificial neural network for inversion of inland water chlorophyll-a. *IEEE Transactions on Geoscience and Remote Sensing* 52 (2):1502–1517.

Stumpf, R. P., T. W. Davis, T. T. Wynne, J. L. Graham, K. A. Loftin, T. H. Johengen, D. Gossiaux, D. Palladino and A. Burtner. 2016. Challenges for mapping cyanotoxin patterns from remote sensing of cyanobacteria. *Harmful Algae* 54:160–173.

Sun, Q.-S., S.-G. Zeng, Y. Liu, P.-A. Heng and D.-S. Xia. 2005. A new method of feature fusion and its application in image recognition. *Pattern Recognition* 38 (12):2437–2448.

Terrio, P. J., T. D. Straub, M. M. Domanski and N. A. Siudyla. 2015. *Continuous monitoring of sediment and nutrients in the Illinois River at Florence, Illinois, 2012–13.* U.S. Geological Survey Scientific Investigations Report 2015-5040. Reston, Va.: U.S. Geological Survey.

Vapnik, V. 1992. *The Nature of Statistical Learning Theory*. New York: Springer.

Verrelst, J., L. Alonso, G. Camps-Valls, J. Delegido and J. Moreno. 2012. Retrieval of vegetation biophysical parameters using Gaussian process techniques. *IEEE Transactions on Geoscience and Remote Sensing* 50 (5):1832–1843.

Verrelst, J., J. P. Rivera, A. Gitelson, J. Delegido, J. Moreno and G. Camps-Valls. 2016. Spectral band selection for vegetation properties retrieval using Gaussian processes regression. *International Journal of Applied Earth Observation and Geoinformation* 52:554–567.

Wang, Z., K. Kawamura, Y. Sakuno, X. Fan, Z. Gong and J. Lim. 2017. Retrieval of chlorophyll-*a* and total suspended solids using iterative stepwise elimination partial least squares (ISE-PLS) regression based on field hyperspectral measurements in irrigation ponds in Higashihiroshima, Japan. *Remote Sensing* 9 (3):264.

Wold, S., M. Sjöström and L. Eriksson. 2001. PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58 (2):109–130.

Yellow Springs Instruments. 2017. *EXO User Manual*, revision G, April 2017 ed. City, St.: Publisher.

Zheng, Z., Y. Li, Y. Guo, Y. Xu, G. Liu and C. Du. 2015. Landsat-based long-term monitoring of total suspended matter concentration pattern change in the wet season for Dongting Lake, China. *Remote Sensing* 7 (10):13975–13999.

Zhong, Z., B. Fan, K. Ding, H. Li, S. Xiang and C. Pan. 2016. Efficient multiple feature fusion with hashing for hyperspectral imagery classification: A comparative study. *IEEE Transactions on Geoscience and Remote Sensing* 54 (8):4461–4478.