Remote Sensing of Environment 123 (2012) 522–531

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Remotely sensed estimates of surface salinity in the Chesapeake Bay: A statistical approach

Erin A. Urquhart [a,*], Benjamin F. Zaitchik [a], Matthew J. Hoffman [b], Seth D. Guikema [c], Erick F. Geiger [d]

[a] Department of Earth and Planetary Sciences, The Johns Hopkins University, 3400 N. Charles Street, Olin Hall, Baltimore, MD 21218, USA
[b] School of Mathematical Sciences, Rochester Institute of Technology, 85 Lomb Memorial Drive Rochester, NY 14623, USA
[c] Department of Geography and Environmental Engineering, The Johns Hopkins University, 3400 N. Charles Street, Ames Hall, Baltimore, MD 21218, USA
[d] College of Earth, Ocean and Environment, University of Delaware, 11 Robinson Hall, Newark, DE 19716, USA

ARTICLE INFO

Article history:
Received 7 November 2011
Received in revised form 6 March 2012
Accepted 16 April 2012
Available online 17 May 2012

Keywords:
Chesapeake Bay
Salinity
Ocean color
Satellite remote sensing
Empirical algorithms
Statistical analysis

ABSTRACT

In coastal and estuarine environments, near-surface salinity varies significantly in space and time. As absolute salinity and salinity gradients are central to many physical and ecological processes in these environments, reliable and consistent salinity estimates are a priority for marine research and application communities. Satellite remote sensing has a great potential to meet this need, yet sensors and algorithms designed to monitor open ocean salinity are typically ill-suited for high resolution applications to coastlines and estuaries. Here we present results of multiple statistical models that predict daily, gridded surface salinity at 1 km resolution across Chesapeake Bay as a function of level 2 surface reflectance estimates from the NASA Moderate Resolution Imaging Spectroradiometer (MODIS), onboard the Aqua platform. Eight statistical methods were tested and it was found that sea surface salinity can be accurately predicted via remote sensed products with an accuracy that is more than sufficient for many physical and ecological applications. For the best-performing statistical model, mean absolute error was 1.82 relative to mean Chesapeake Bay salinity of 16.5.

1. Introduction

Sea surface salinity plays a vital role in circulation patterns, influences the spatial distribution of many marine organisms, and affects seawater density in both coastal systems and open oceans. In coastal and estuarine environments, even small changes in salinity can greatly alter the transportation course and lifecycle of organisms and the status of the ecosystems they comprise (Baird and Ulanowicz, 1989). For this reason, salinity is a core input to ecological analyses and to operational models designed to monitor physical and biological processes in coastal environments. Advances in coastal remote sensing and computer modeling technology have led to several successful operational products that employ sea surface salinity. The National Atmospheric and Ocean (NOAA) CoastWatch Program provides a near real-time product for forecasting harmful algal blooms and predicting the likelihood of where sea nettles exist in the Chesapeake Bay. NOAA's forecasting models are accomplished by applying surface salinity and temperature estimated from a numerical hydrographic model (ChesROMS) to species-specific habitat models for the Bay (National Oceanic and Atmospheric Administration (NOAA), 2010).

These applications point to the critical need for reliable, continuous, and spatially distributed estimates of salinity in coastal environments. In situ salinity measurements are a critical component of such monitoring efforts, but cost and logistics limit the temporal and spatial coverage of such measurements. The Chesapeake Bay Monitoring Program conducts routine bi-monthly water quality monitoring along the mainstem sections of Maryland and Virginia Bay waters. The monitoring program measures key components of the Bay ecosystem: habitat, living resources, pollutant inputs, and water quality. These monitoring efforts are used in both research and modeling of the Chesapeake Bay ecosystem (Maryland Department of Natural Resources, 2011). Both physical and biological processes in coastal systems can occur on spatial and temporal scales that are not observed through monthly environmental sampling at designated sites and transects.

Satellite remote sensing offers the potential to estimate salinity across entire water bodies at the frequency of satellite overpass, dramatically enhancing our monitoring capabilities relative to in situ observation networks. To date, however, satellite missions targeting salinity have focused on open ocean rather than coastal applications. NASA's Aquarius mission, launched in June 2011, and the European Space Agency's Soil Moisture and Ocean Salinity (SMOS) mission launched in November 2009, are capable of measuring sea surface salinity from space across the world's oceans, but the 150 km spatial and 7-day temporal resolution of Aquarius and the 250 km spatial and 10–30 day average temporal resolution of SMOS are too coarse for coastal and estuarine environments. The Chesapeake Bay, for

* Corresponding author at: Johns Hopkins University, 3400 N. Charles Street, 301 Olin Hall, Baltimore, MD 21218, USA. Tel.: +1 410 516 7135.
E-mail address: erinu@jhu.edu (E.A. Urquhart).

0034-4257/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2012.04.008

example has a maximum width of only 48 km (National Aeronautical Space Administration, NASA, 2011). The coarse resolution of these salinity missions stands in contrast to the 1 km spatial resolution estimates of sea surface temperature (SST) that are produced with near global coverage on a daily basis by MODIS and other sensors. Estimating high-resolution coastal and estuarine surface salinity from satellite is known to be a valuable tool, yet no proven or operational salinity algorithm exists for the Chesapeake Bay.

Attempts to successfully map sea surface salinity via remote sensing have ranged from Skylab photography (Lerner and Hollinger, 1977) to microwave radiometer measurements (Blume and Fedors, 1978), decametric wave ranges (Kachan and Pimenov, 1997), ESTAR measurements, and Landsat TM data (McKeon and Rogers, 1976). The use of satellite imagery to map sea surface salinity in an estuary was first performed in the San Francisco Bay by Khorram (1982). This pioneer study found correlations between Landsat TM color bands and sea surface salinity in an estuarine environment. Other studies (Del Vecchio and Blough, 2004; Bowers and Brett, 2008; Maisonet et al., 2009) explore the empirical relationships between colored dissolved organic matter (CDOM) and salinity using remotely sensed ocean color in a coastal setting. These studies showed that a straight-line relationship between CDOM and salinity is expected dependent on the ratio of the flushing time of an estuary and the timescale of the source variation.

The empirical relationship between colored dissolved organic matter (CDOM) and salinity is important in that CDOM serves as an intermediary function between remote sensing reflectance bands and sea surface salinity. This relationship assumes that fresh-high CDOM river waters mix conservatively with salty-low CDOM seawater, and therefore an inversely correlated relationship between CDOM and salinity (D'Sa and Miller, 2003). Since we can measure CDOM from space, we can also derive salinity values from remotely sensed observations. It is important to note that this method only works in systems in which there is conservative mixing between coastal waters and rivers. Flocculation and photodegradation could invalidate the assumptions of conservative mixing in this method, however previous work (Blough et al., 1993; Del Castillo et al., 1999) has shown that these effects have negligible impacts on CDOM concentration. Therefore, because sea surface salinity is a function of colored dissolved organic matter, it is also a function of remote sensing reflectance. Thus we are confident in our assumption that sea surface salinity can be expressed directly as a function of remotely sensed ocean color bands. To minimize the number of empirical models applied when deriving salinity from satellite registered radiance, and to capture any additional information on salinity contained in MODIS reflectance bands, we used the standard remote sensing reflectance bands in a multivariate regression model rather than a univariate model using solely CDOM.

The purpose of this study is to predict sea surface salinity in the Chesapeake Bay at 1 km resolution using MODIS-Aqua ocean color bands (Table 2). This effort is built on work by Geiger et al. (in press), in which Chesapeake Bay salinity fields were estimated at 1 km resolution using an artificial neural network (ANN) algorithm applied to MODIS-Aqua data. Here, we test the hypothesis that salinity predictions with smaller or similar errors can be achieved using simpler, more transparent statistical models. To explore a range of statistical modeling options, this study uses eight empirical models typically used when representing continuous response variable data. The eight statistical models are: a Categorical and Regression Tree model (CART), a Generalized Linear Model (GLM), a Generalized Additive Model (GAM), a Random Forest Model, a Mean model, an Artificial Neural Network (ANN), a Multivariate Adaptive Regression Spline (MARS), and a Bayesian Additive Regression Tree (BART). Each of these models includes the dependent response variable[1] sea surface salinity and a set of remotely sensed independent predictor variables[2] described in the data description section below. To test the generalizability of model-predicted sea surface salinity across the diverse salinity conditions of the Chesapeake Bay, we run six seasonal and regional cross validation tests using the top three performing salinity models. The spatial and temporal cross evaluation lends to a more generalizable salinity product than earlier Chesapeake Bay salinity products.

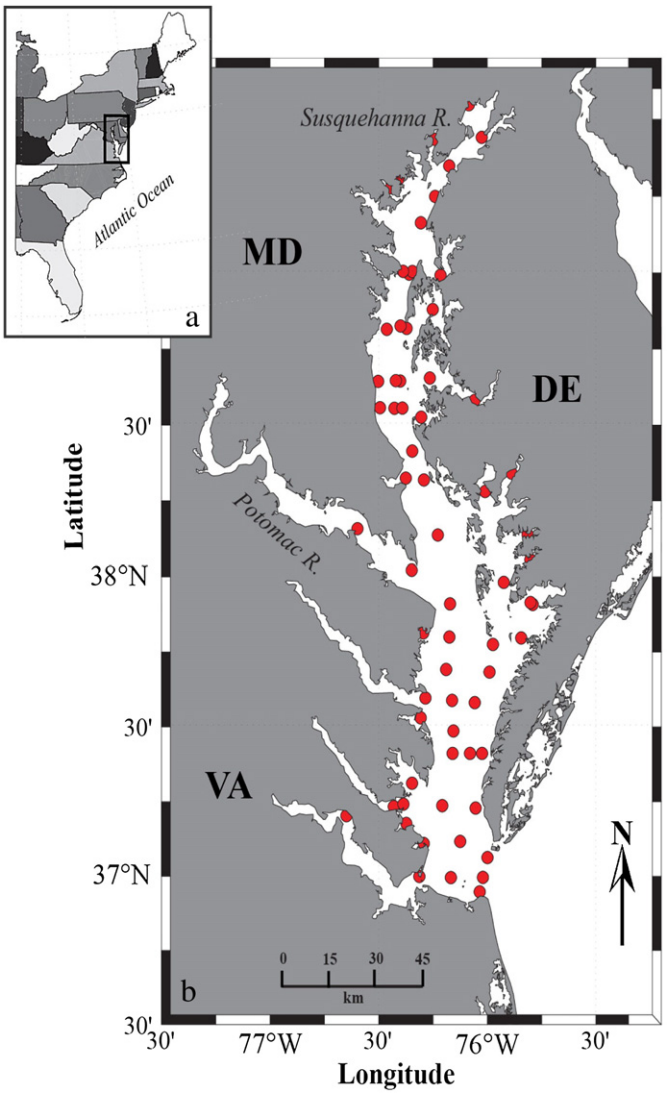

Fig. 1. The a) Mid-Atlantic coast and the b) inset of the Chesapeake Bay Estuary with 67 Chesapeake Bay Monitoring Program stations.

2. Data description

2.1. Study area

The Chesapeake Bay is the largest estuary in the United States, extending 332 km (from Havre de Grace, MD to Cape Charles, VA) along the Atlantic Coast (Fig. 1). The Chesapeake Bay estuary has a strong north-to-south salinity gradient that includes oligohaline (0–6), mesohaline (6–18), and polyhaline (18–30) zones (Baird and Ulanowicz, 1989)[3]. Sea surface temperatures in the Bay range from

[1] In a statistical experiment, a "dependent response variable" is the observed variable whose changes are determined by the presence of one of more independent variables (Brownlee, 1960).

[2] An "independent predictor variable" is a manipulated variable whose presence determines the change in the dependent variable (Brownlee, 1960).

[3] In situ and estimated salinity values reported in this study use a standard unitless measure.

Table 1
Data types, spatial resolution, temporal resolution, and sources of data.

Data type, parameters (period of record)	Spatial resolution	Temporal resolution	Source
In situ station data, salinity, surface temperature (2003–2010)	67 main-stem monitoring stations on Bay axis, 1 m vertical resolution	~20 surveys/yr, bi-monthly to monthly cruises	MDDNR[a]; VA DEQ[b] (Chesapeake Bay Monitoring Program)
L3-mapped ocean color and thermal SST satellite products[c] (2003–2010)	1 km spatial resolution	Daily satellite overpasses	Modis AQUA, National Aeronautical Space Administration

[a] Maryland Department of Natural Resources.
[b] Virginia Department of Environmental Quality.
[c] L2 Modis AQUA standard suite of products (see Table 2).

local wintertime lows of −0.5 °C to summertime highs of 31 °C. The oligohaline upper Bay has a mean depth of 4.5 m, the mesohaline middle Bay 10 m, and the polyhaline lower Bay 9 m, giving the overall Bay an average depth of 6.5 m (22 ft) (Baird and Ulanowicz, 1989).

The physical transport regime of the Chesapeake Bay estuary follows the classical estuarine circulation model of partially mixed estuaries, in that it is characterized by a 2-layer gravitational circulation scheme. As salt water enters the mouth of the Bay along the eastern shore, there is a net up-estuary flow of water, which occurs below the pycnocline, and a complementary net down-estuary flow as the fresh surface water makes its way from the head to the mouth of the Chesapeake Bay (Pritchard, 1952).

The drainage area of the Chesapeake Bay watershed encompasses 166,000 km^2. Freshwater flows into the Chesapeake Bay estuary from 25 main rivers and tributaries. The Susquehanna River is the largest tributary in the Chesapeake Bay and accounts for approximately 45% of freshwater flow into the Bay (Baird and Ulanowicz, 1989).

2.2. In situ measurements

The analysis performed in this paper made use of in situ environmental data collected by the Chesapeake Bay Monitoring Program (Table 1). Bi-monthly data was collected during various research cruises organized by the Maryland Department of Natural Resources (MDDNR) and the Virginia Department of Environmental Quality (VADEQ). The dataset included in situ salinity measurements from 67 monitoring stations (Fig. 1) along the Bay's axis collected from 2003 through 2010. Using the satellite diffuse attenuation coefficient for down-welling irradiance at 490 nm, we calculated the optical depth at each sampling location and found that the mean optical depth of our samples was 0.89 m. Therefore, sampling measurements more than 1 m in depth were excluded from this study for reasons of remotely sensed surface optical depth.

2.3. MODIS satellite measurements

The satellite remotely sensed ocean color products used in this study were from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua (Tables 1 and 2). Standard ocean color data products were downloaded from NASA's ocean color website (http://oceancolor.gsfc.nasa.gov/), and then batch processed in the SeaWIFS Data Analysis System (SeaDAS). Level-2 daytime standard suite ocean color products at 1 km spatial resolution were mapped directly to a cylindrical coordinate system and then standard quality control flags were applied. Daily satellite images were acquired for the same time period as in situ measurements.

For the purposes of in situ-satellite calibration, we matched in situ station data to the daily satellite measurements within a 1 km radius of the sampling station. Any remotely sensed measurements that were within 1 km of the monitoring station were averaged and thus representative of the unique value of that salinity “pixel”. This sampling procedure yielded 620 satellite and in situ matched measurements for use in statistical analysis.

3. Methods

3.1. Statistical models

This study presented eight different statistical models developed to predict sea surface salinity via remotely sensed ocean color measurements in the Chesapeake Bay. We chose the eight major types of empirical models that are typically used to regress continuous response variable data. A holdout cross validation was used with the eight statistical models in which 80% of the matchup data points was used to train the models and the remaining 20% was used for validation. Table 2 summarizes the twelve predictor variables that were used to train the eight empirical models presented below. Multivariate models were also compared to a univariate model that used the standard MODIS-Aqua CDOM product (Morel and Gentili, 2009) to predict salinity. The univariate model was found to underperform multivariate models, and will not be discussed further. All statistical computations were carried out in R Statistical Package 2.14 (R Development Core Team, 2011), on an Intel Xeon W3580 Processor, 3.33 GHz machine with 12 GB RAM. Computational time for all statistical models within the holdout validation test was less than one

Table 2
Variables used in model development.

Variable name	Variable description	Mean (μ)	Standard deviation	Maximum	Minimum
Salinity (predictor variable)	In situ salinity measurement at surface	16.49	4.69	31.65	0.00
Lat	Latitudinal data coordinate of in situ-satellite matchup	37.68	0.51	39.44	37.00
Lon	Longitudinal data coordinate of in situ-satellite matchup	−76.14	0.15	−75.79	−76.46
Rrs_412	Remote sensing reflectance at 412-nm	0.0014	0.0012	0.0058	−0.001
Rrs_443	Remote sensing reflectance at 443-nm	0.0022	0.0011	0.0067	0.0003
Rrs_469	Remote sensing reflectance at 469-nm	0.0029	0.0013	0.0083	0.0006
Rrs_488	Remote sensing reflectance at 488-nm	0.0035	0.0015	0.0094	0.0008
Rrs_531	Remote sensing reflectance at 531-nm	0.0055	0.0020	0.0126	0.0018
Rrs_547	Remote sensing reflectance at 547-nm	0.0060	0.0021	0.0140	0.0018
Rrs_555	Remote sensing reflectance at 555-nm	0.0059	0.0020	0.0139	0.0019
Rrs_645	Remote sensing reflectance at 645-nm	0.0030	0.0015	0.0145	0.0006
Rrs_667	Remote sensing reflectance at 667-nm	0.0022	0.003	0.0137	0.0002
Rrs_678	Remote sensing reflectance at 678-nm	0.0022	0.0012	0.0135	0.0003

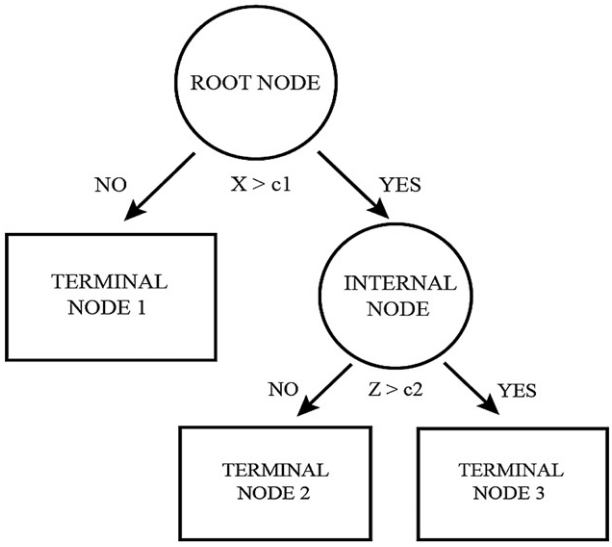

Fig. 2. Artificial neural network architecture.
Adapted from Lee and Park (1992).

hour, with the exception of the BART model which required up to seven hours of computational time.

3.1.1. Generalized linear model (GLM)

Generalized linear models are an extension of the standard Ordinary Least Squares (OLS) linear model that allows for regression analysis of both continuous and count data (Nelder and Wedderburn, 1972). An OLS standard model works to minimize the sum of vertical distances between the observed and the predicted response, commonly called the sum of squared residuals (Hastie et al., 1998). An OLS model is composed of two key elements: 1) the random component, which is the probability distribution of the response variable, y, given the predictor variables x_i and 2), the linear predictor, which is an equation that incorporates the data from the predictor variables. A generalized linear model generalizes the standard OLS model by adding a link function, which relates the linear predictor to a function of the predictor variables specifying the conditional mean (Cameron and Trivedi, 1998). The link function transforms the expectation of the linear predictor. The salinity measurements in this dataset exhibited a normal Gaussian distribution and therefore we used a normal identity link function $\mu = X\beta$ in the construction of the GLM.

3.1.2. Generalized additive model (GAM)

A GAM is a flexible statistical model that extends the traditional linear model by allowing for nonlinear relationship between the dependent response and independent predictor variables (Hastie and Tibshirani, 1986). This model replaces the $X\beta$ link function of the generalized linear model with a non-parametric smoothing function $f(X)$. The smoothing function can provide information about the relationship between the predictor variables and response variable that is not revealed using a traditional linear model. Nonlinear effects of the covariates on the response variable y can be expressed using GAM. For this study the standard smoothing approach, a cubic regression spline, was used. A cubic regression spline imposes a smoothness on the function $f(X)$, with a potential knot point at each of the unique values of x. Again, an identity link function was used to establish a relationship between the mean value of the response variable y and the smoothed function of the x together with a Gaussian conditional distribution (Hastie and Tibshirani, 1986).

3.1.3. Artificial neural network (ANN)

An artificial neural network (ANN) is commonly defined as a massive interconnected network composed of processors, which operate in parallel and learn from experience and training (Lee and Park, 1992). The idea of a neural network comes from the biological neural system; the processing elements of an ANN serve as the neurons, while the connections are like synapses from a biological system. The neurons in the ANN are interconnected by means of various information channels. A neural network has at least three basic layers: the inputs, the hidden layer, and the outputs. Input neurons send data via synapses or connections to the hidden layer then via more connections send data to the output neurons (Fig. 2). Each synapse has an unknown parameter called the "weight"; the weighted inputs

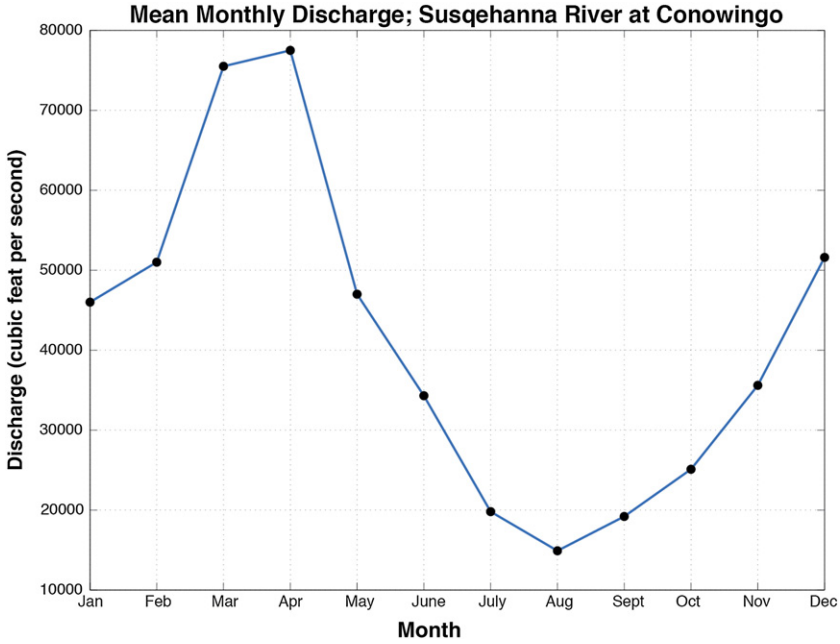

Fig. 3. Mean monthly (1970–2000) discharge at USGS 01578310 Susquehanna River at Conowingo, MD station.
Adapted from United States Geological Survey (2012).

Table 3
Comparison of holdout mean absolute errors (MAEs) based on 120 random holdout samples.
p-Values in **bold** represent statistically significant differences between models.

Model	MAE	p-Value GAM	p-Value CART	p-Value BCART	p-Value RF	p-Value MEAN	p-Value ANN	p-Value BART	p-Value MARS
GLM	1.93	3.4e−06	**2.2e−16**	**2.2e−16**	**1.5e−05**	**2.2e−16**	0.0006	**0.0001**	0.1407
GAM	1.82		**2.2e−16**	**2.2e−16**	**4.8e−15**	**2.2e−16**	0.2575	**5.9e−14**	**2.3e−09**
CART	2.39			0.7254	**2.2e−16**	**2.2e−16**	**2.2e−16**	**2.2e−16**	**2.2e−16**
BCART	2.38				**2.2e−16**	**2.2e−16**	**2.2e−16**	**2.2e−16**	**2.2e−16**
RF	2.06					**2.2e−16**	**7.7e−12**	0.5489	0.0015
MEAN	3.72						**2.2e−16**	**2.2e−16**	**2.2e−16**
ANN	1.85							**9.5e−11**	**2.1e−06**
BART	2.04								0.0093
MARS	1.98								

are added together and if the sum exceeds the pre-specified threshold then the neuron fires, giving an output (Lee and Park, 1992). To maximize prediction accuracy, we first tested two different neural networks, one with 40 hidden nodes and one with 45 hidden nodes, with these sizes selected based on Geiger et al. (in press). The neural network that exhibited the optimum node size for salinity prediction was then used in the holdout cross validation. In training our ANN models we did note a dependence on the randomly selected initiation points for the weights (i.e., the final trained network varied slightly for different initiation sets).[4] As a result, we trained 5 different ANNs of each size, with each network starting from a different, fixed seed numbers for the initial sampling of the weights. We report the error from the average of these five models below.

3.1.4. Multivariate Adaptive Regression Spline (MARS)

Multivariate adaptive regression spline (MARS) is a non-parametric regression method that can be seen as an extension of a linear model allowing for interactions and non-linearities in a dataset (Friedman, 1991). MARS behaves like a generalized linear model, but based on automatically selected basis functions. MARS builds models in the same fashion as recursive partitioning trees, but allows for a forward and backward pass (Hastie et al., 2008).

3.1.5. Tree-based data mining techniques

To further a different class of models for empirically predicting sea surface salinity in the Chesapeake Bay, the study used four tree-based data mining methods: classification and regression tree (CART) (Breiman et al., 1998), Bayesian additive regression trees (BART) (Chipman et al., 2010), bagged categorical and regression trees (BCART) (Sutton, 2005), and Random Forest model (RF) (Breiman et al., 1998). Each of the four tree-based data mining methods explores the relationship between the predictor variables and the dependent response variable, sea surface salinity. The dataset undergoes recursive binary partitioning at the nodes. Tree-based methods give a flexible description of relationships within the dataset while also providing a convenient visual for result interpretation.

3.1.6. Mean model

Each of the statistical models outlined in this section was compared to a mean statistical null model. Our mean model was simply the average value of the response variable salinity. For validation purposes, all nine models including the mean model were input into the holdout run.

3.1.7. Geographic model

Surface salinity, optical depth, and CDOM/salinity relationships are highly variable and dependent on location in the Chesapeake Bay. To test the added value of using ocean color bands, as well as the correlation between salinity and geographic location, a holdout validation test using only latitude and longitude was run employing the nine statistical methods outlined above.

3.2. Cross-validation of top statistical models

In order to develop the most effective remotely sensed salinity prediction model, we needed to test the generalizability of the empirical

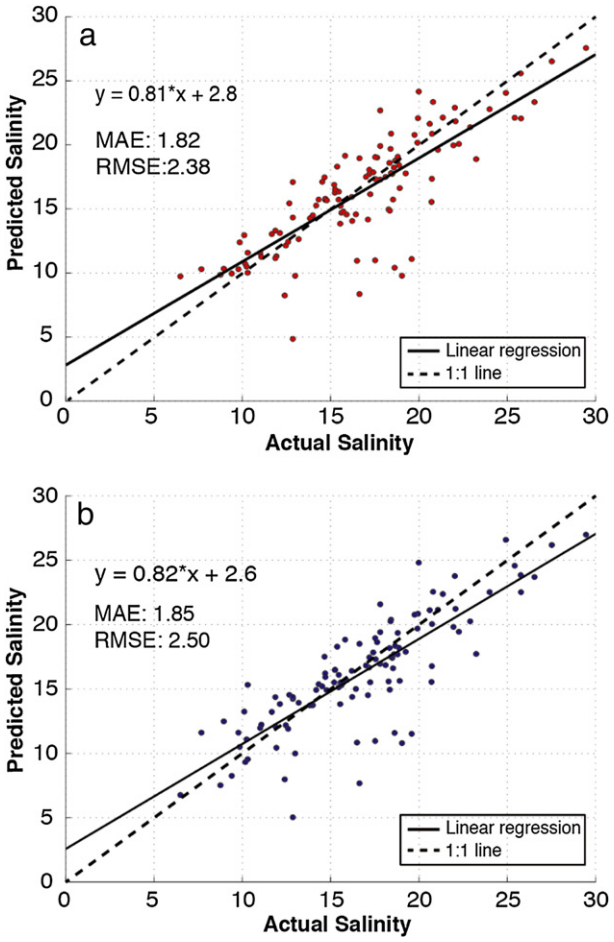

Fig. 4. One-to-one model regression between in situ salinity and predicted salinity for a) the GAM and b) the ANN. The mean absolute errors for each statistical model are: a) 1.83 and b) 1.85.

[4] This variability persisted well beyond the number of replications of the ANN training algorithms at which convergence was reported.

Table 4
Comparison of holdout mean squared errors (MSEs) based on 120 random holdout samples.
p-Values in **bold** represent statistically significant differences between models.

Model	MSE	p-Value GAM	p-Value CART	p-Value BCART	p-Value RF	p-Value MEAN	p-Value ANN	p-Value BART	p-Value MARS
GLM	6.40	0.0002	**2.2e−16**	**2.2e−16**	0.0003	**2.2e−16**	0.8135	0.0415	0.7253
GAM	5.67		**2.2e−16**	**2.2e−16**	**1.6e−10**	**2.2e−16**	0.1956	**5.3e−08**	0.0004
CART	9.17			0.6968	**9.7e−15**	**2.2e−16**	**1.1e−07**	**2.2e−16**	**2.2e−16**
BCART	9.08				**9.7e−14**	**2.2e−16**	**2.6e−07**	**2.2e−16**	**2.2e−16**
RF	7.14					**2.2e−16**	0.0800	0.0606	**6.9e−05**
MEAN	22.07						**2.2e−16**	**2.2e−16**	**2.2e−16**
ANN	6.28							0.3101	0.9162
BART	6.77								0.0137
MARS	6.33								

algorithms in the Chesapeake Bay. To validate the reliability of the salinity predictions throughout the Chesapeake Bay, we split up the in situ-satellite matchup dataset temporally and geographically. In both cases, the top three statistical models, determined by lowest mean absolute errors (MAE), were used in a cross-validation on different spatial and temporal periods of the Bay.

There is great seasonal and geographic variability in in situ salinity, fresh water discharge (Fig. 3), as well as in cloud cover that interferes with satellite retrieval of surface reflectances in the Bay. Therefore, although we developed and tested the statistical models on year-round in situ-satellite matchups, it is important to train the models on one season and predict for another to reflect the variations in fresh water inflow into the Chesapeake Bay. To do so, we divided the entire matchup dataset into two discharge datasets: high (December through May) and low (June through November). As described above, there are various salinity regimes throughout the Bay, which exhibit certain characteristics dependent on geographic location and biophysical processes in that location. For example, cold saline seawater is characteristic of water lying along the mid-eastern shore due to estuarine circulation patterns in the Chesapeake Bay. To cross-check this spatial variability, we also split in situ-satellite matched datapoints spatially, into North versus South and East versus West. Geographic divisions were performed separately from seasonal divisions. For both, the top three statistical models were trained on one database (low/South/East) and then tested on the other (high/North/West), then vice-versa.

3.3. One-to-one comparison of remotely sensed versus in situ salinity

To assess the functionality of our empirical salinity model, we tested the top-performing model on a separate set of remotely sensed independent variables. To guarantee a one-to-one comparison with in situ salinity measurements, we chose a daily MODIS image with good spatial coverage from a day (September 18, 2006) when the Chesapeake Bay Monitoring Program conducted in situ salinity measurements. Overlap in MODIS and station measurements from that day allowed for 13 in situ-satellite comparison points.

Table 5
Comparison of holdout MAE, RMSE, and MSE values.

	GAM	ANN	GLM	CART	BCART	RF	MEAN	BART	MARS
MAE	1.82	1.85	1.93	2.39	2.38	2.06	3.72	2.04	1.98
RMSE	2.38	2.50	2.53	3.03	3.01	2.67	4.69	2.60	2.52
MSE	5.67	6.28	6.40	9.17	9.08	7.14	22.07	6.77	6.33

Table 6
Comparison of mean predicted salinity based on 120 random holdout samples.

	Mean salinity	p-Value GLM	p-Value GAM	p-Value ANN
In situ	16.73	0.476	0.495	0.394
GLM	16.31		0.986	0.875
GAM	16.32			0.864
ANN	16.22			

4. Results and discussion

4.1. Model comparisons

The in situ-satellite dataset was fit with the eight statistical models outlined above using a repeated holdout validation test. Each of the statistical models was compared to the mean prediction model in the holdout test to determine how well each model performed assuming the dataset mean salinity value. This results in 36 pair-wise tests with a mark of statistical significance if the p-value[5] on a given test is less than 0.00014 in accordance with the needed Bonferroni correction (Devore, 1995). As shown in Table 3, all eight statistical models outperform the mean model by a statistically significant amount ($p<2.2e-16$). The generalized additive model has the best prediction accuracy with the lowest MAE of 1.82 followed by the 45-node ANN model with a MAE of 1.85, and the GLM with a MAE of 1.93. The one to one regressions of the matched in situ salinity vs. the model predicted salinity for the GAM, and the ANN models (Fig. 4) show that there are approximately ten data points in which the prediction model clearly under predicts the true salinity value. The locality of these data accounts for the large error as it was found that each outlying predictor was nearby the mouth of a fresh water tributary. Not only do we see increased freshwater flow, but also variability in the discharge of sediments, terrigenous organic matter, detritus, and chlorophyll concentrations in these regions. These changes can complicate the bio-optical properties of the water due to the absorptive properties of CDOM, phytoplankton mass, and detritus, which further affect the shape of the remote sensing signal at each location. Further model development and variable specification need to be carried out to understand the effects of these environmental conditions on model prediction.

GAM, followed by ANN, also has the highest predictive accuracy when judged by mean square error (MSE) and root mean square error (RMSE) (Table 4). The difference in MSE values between GAM and ANN is not significant at a 95% confidence level. All empirical models outperform the mean model with respect to MSE. MSE and RMSE are useful metrics for identifying outliers in the model fit. A RMS error of equal or higher value than the MAE (see Table 5) indicates that there are outlier salinity outputs in the top three salinity

[5] p-Value indicates the probability that the result obtained in a statistical test is due to chance rather than a truce relationship between measures (Brownlee, 1960).

Table 7
Approximate significance of GAM smoothed terms.
p-Values in **bold** represent statistical significance ($p<0.05$).

Smoothed term	p-Value
Lat	**2.20e−16**
Lon	**2.20e−16**
Rrs_678	**4.95e−05**
Rrs_667	**4.27e−08**
Rrs_645	**0.007**
Rrs_555	0.118
Rrs_547	0.293
Rrs_531	**1.16e−06**
Rrs_488	**1.11e−11**
Rrs_469	0.289
Rrs_443	**2.94e−14**
Rrs_412	**3.18e−11**

Table 8
LAT–LON only model comparison of holdout MAE and RMSE values.
Values in **bold** represent the models that are significantly different ($p<0.05$) than the original eight models.

	GAM	ANN	GLM	CART	BCART	RF	BART	MARS
MAE	**2.36**	**2.38**	**2.55**	2.41	2.42	**2.40**	**2.36**	**2.35**
RMSE	**2.98**	**2.98**	**3.21**	3.05	3.05	**3.01**	**2.96**	**2.98**

prediction models. It is important to note that there is no statistically significant difference in the salinity prediction for the GAM, the ANN, or the GLM (see Table 6).

Table 9
MAE and RMSE values for cross-validation tests.
Naming convention for cross-validation is as follows: "East for West" translates to model trained on East dataset and tested on West dataset.

	MAE				RMSE			
	GLM	GAM	ANN	MEAN	GLM	GAM	ANN	MEAN
East for West	2.1	1.8	3.7	3.3	2.6	2.3	4.7	4.0
West for East	2.6	2.8	4.0	4.1	3.3	3.5	5.2	5.3
North for South	3.4	2.1	5.9	5.7	4.2	2.8	7.0	6.8
South for North	3.0	6.4	6.1	5.7	4.2	9.9	7.1	6.5
High for Low	2.3	2.3	2.6	4.2	3.0	3.0	3.3	5.3
Low for High	2.5	2.3	2.8	3.9	3.0	2.7	4.3	4.8

For GAM, it is also possible to examine the specific importance and influence of each of the reflectance bands in the prediction of salinity. Table 7 lists the p-values associated with each smoothed term in the GAM. Nine of 12 variables included in the GAM are statistically significant (p-value<0.05). Though model results show that latitude and longitude are the most significant predictor variables in the GAM model, a holdout run using only latitude and longitude shows a significant (p-value<0.05) decrease in prediction accuracy (Table 8), and thus value added in using the remotely sensed reflectance values. While all but three of the predictor variables are statistically significant and thus important in predicting the response variable y, not all of the variables that have high importance are highly influential

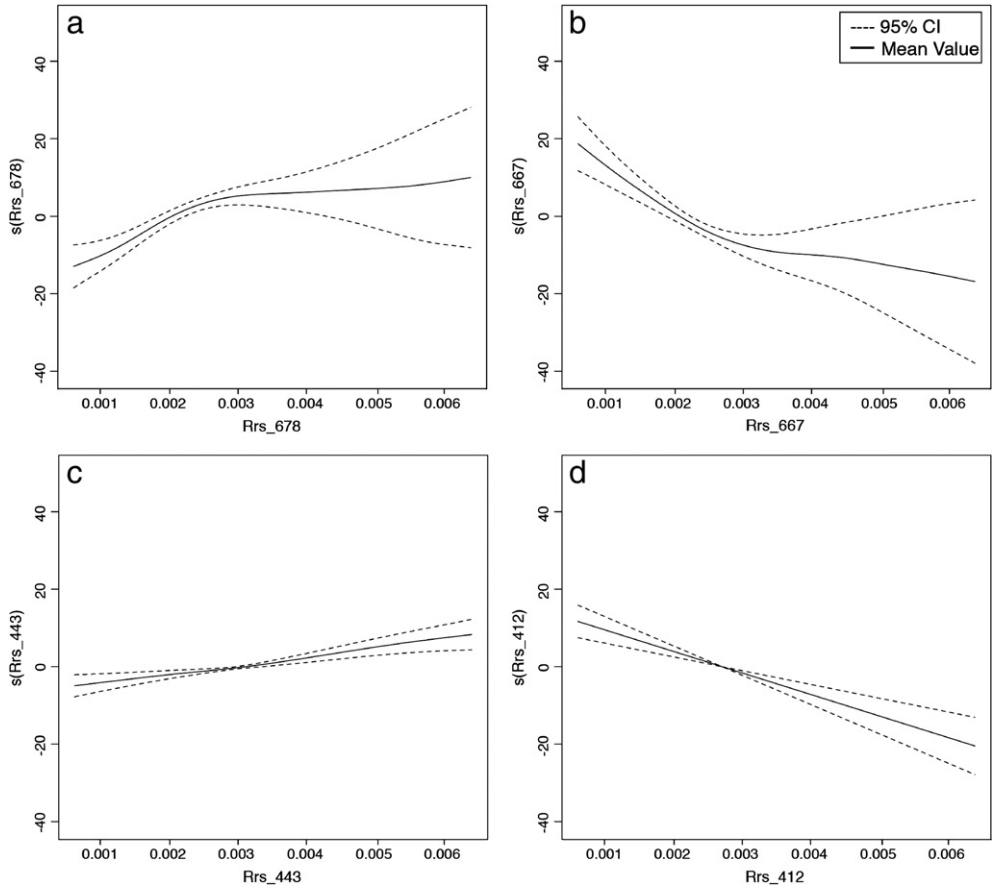

Fig. 5. Variable plots for GAM model; a) Rrs_678, b) Rrs_667, c) Rrs_443, and d) Rrs_412. y-Axis value is the transformed spline value, x-axis shows Rrs value at each unique data point.

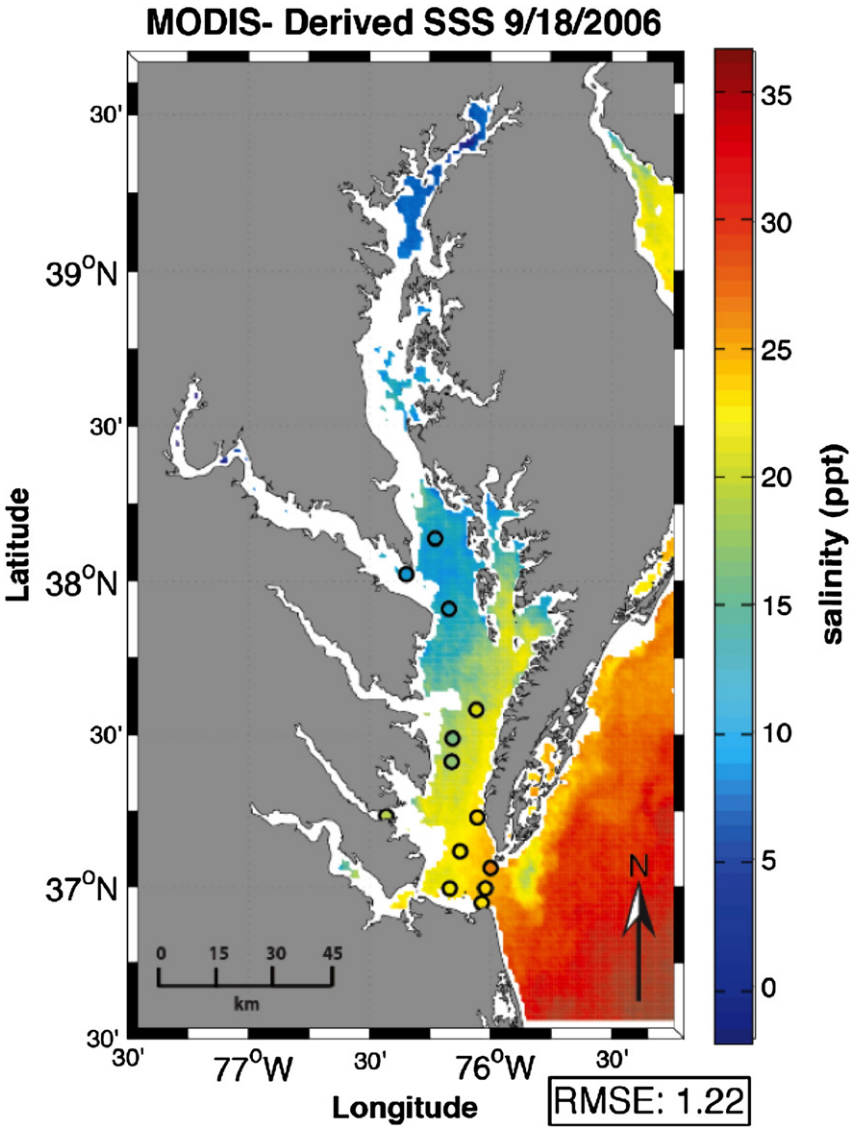

Fig. 6. GAM predicted salinity for September 18, 2006 with in situ station locations and actual salinity values marked by color-filled black circles.

to the model outcome[6]. Of the twelve smoothed terms included in the GAM, half show high influential behavior on the predicted response. Indicative examples of variable responses are shown in Fig. 5. Remote sensing reflectance (Rrs) at 488 nm is positively associated with salinity (Fig. 5a), while Rrs at 667 nm and at 443 nm is negatively associated (Fig. 5b, c). Other predictor variables, such as Rrs at 412 nm, are statistically significant in the GAM but show no particularly strong independent influence on salinity (Fig. 5d).

4.2. Cross-validation of models

To test the generalizability of our remotely sensed salinity product in the Chesapeake Bay, we ran six seasonal and regional cross validation tests using the top three performing salinity models. In these cross-validation analyses, the GAM and the GLM perform with better error accuracy than the ANN in all cases but one (Table 9). The first two of the six cross-validation tests evaluated the generalizability of salinity models from east to west in the Bay. In training the three models on the eastern Bay portion and testing on the West and vice-versa, GAM performs the best with a MAE of 1.8 in the first case, and GLM the best when trained on the West and tested on the East (note that differences between GAM and GLM were not statistically significant). When the same tests were conducted for low and high, all three models perform well—when trained on high for low testing, both GLM and GAM have a MAE of 2.3. While the generalizability of the models for East versus West and low versus high performs well in terms of low MAE and RMSE values, the cross-validation tests for North versus South are not as consistent in their prediction results. From Table 9, we can see that although the GAM MAE for "North for South" performs equally as well as the previous tests, the model trained on the South and tested on the North underperforms relative to the mean model. This is the only generalizability test for which either GAM or GLM was outperformed by the mean model. This result is likely a product of systematic differences between the relatively fresh North and the saltier South, and is the subject of continued investigation.

4.3. One-to-one daily GAM predicted, in situ comparison

The comparison of in situ salinity to GAM predicted salinity for September 18, 2006 results in improved prediction accuracy over

[6] Variable reduction was performed on both the GAM and the GLM, but this did not improve the prediction accuracy over the final non-reduced models.

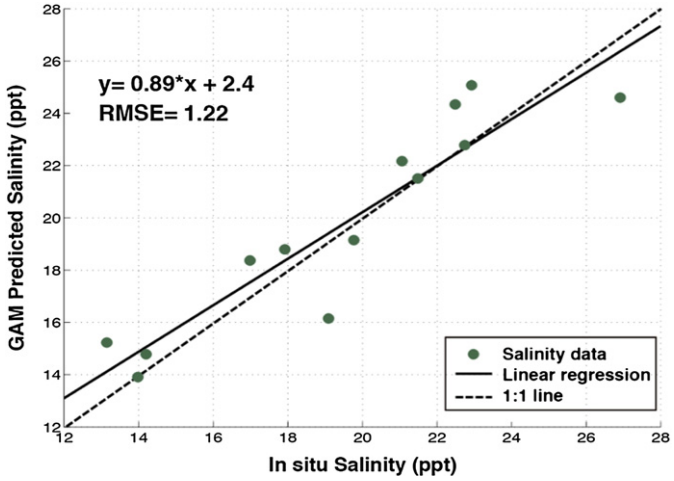

Fig. 7. Regression between in situ salinity and GAM prediction salinity for September 18, 2006. The RMSE is 1.22.

the holdout validation data sets. Five of the 18 in situ stations were removed from the one-to-one comparison because they fell outside remote sensing coverage for the given day. Fig. 6 shows the predicted GAM salinity for the entire Bay, as well as the actual in situ salinity at the stations marked by filled circles. The RMS error improved from 2.38 in the holdout validation tests to 1.22 for the daily prediction versus in situ. Fig. 7 shows the regression of the in situ versus GAM predicted salinity with a slope of 0.89. In addition to the improved RMS error between actual and predicted salinity, predicted salinity from the GAM follows a believable salinity regime for the Bay. Not only do the predicted values fall within the natural range for the Bay, but also the prediction actually exhibits the spatial gradients explained earlier in this paper.

5. Conclusions

The eight statistical models presented above show that remotely sensed products can be used to accurately estimate sea surface salinity in the Chesapeake Bay. While predicting salinity via remote sensing for the Bay is still in its beginning stages, the results of applying these models to remotely sensed measurements can provide the imperative missing block to many biological and physical marine applications. Three models that perform particularly well in estimating salinity were the generalized additive model, the generalized linear model, and the artificial neural network.

Additionally, six cross-validation tests were run to evaluate the generalizability of our salinity estimates across various temporal and spatial regimes in the Chesapeake Bay. Table 8 summarizes the MAE and RSME results from the six cross-validation models. From the prediction results we can conclude that for the Chesapeake Bay, the GAM and GLM outperform the artificial neural network; further supporting our original hypothesis that a more transparent model can estimate sea surface salinity with equal or better accuracy than an ANN. We can assume that the tendency of the more complicated neural network was to over fit the data, resulting in the poor prediction accuracy, showing that the transparent models like the GLM and GAM are more generalizable to the Chesapeake Bay region.

The empirical models presented in this study are particularly good at estimating sea surface salinity in the Chesapeake Bay. We do note, however, that salinity estimates were found to be highly dependent on geographic location. Results show that latitude and longitude are the most significant predictor variables in the nine surface salinity estimation models. While this locality issue was anticipated for the Chesapeake Bay and thus accounted for, it indicates that attention to mixing processes, fresh water inflow, and seasonality will be required when applying these statistical salinity models to other coastal regions. A second limitation of the study is in the data itself. The in situ salinity measurements presented in the paper were taken at a water depth of approximately 0.5 m. Satellite remote sensing is useful in detecting sea surface reflectance signals, but the inability to penetrate below the ocean's surface and clouds often limits the availability of data. Therefore lies a discrepancy between the depth of the in situ measurement and the remotely sensed surface reflectance. Further work will focus on interpolation methods to understand salinity changes as a function of water column depth. A third limitation of the model's training data is the temporal and spatial scarcity of in situ salinity measurements. As presumed, the availability of remotely sensed reflectance data far exceeds the number of environmental surface measurements.

In order to obtain full temporal and spatial coverage of Chesapeake Bay, the satellite remote sensing data and in situ observations can be combined with a fluid dynamical model through data assimilation. In this way, the observations are utilized when they are available, but model dynamics will drive accurate forecasts in the absence of observations. Data merging of in situ and RS observations through the use of a numerical model will provide a full 3 dimensional coverage of the Bay that will therefore allow us to propagate the satellite sea surface information deeper into the water column. Such a data assimilation system is being developed for the Chesapeake Bay (Hoffman et al., in review) and in future work we hope to leverage that system and create more complete sea surface salinity estimations for the Bay.

Acknowledgments

Special thanks to Carlos del Castillo at the Johns Hopkins University Applied Physics Laboratory and Bruce Monger at Cornell University. This research was supported by the Earth and Planetary Sciences Department, the Glenadore and Howard L. Pim Postdoctoral Fellowship, and the Global Water Program of Johns Hopkins University.

References

Baird, D., & Ulanowicz, E. (1989). The seasonal dynamics of the Chesapeake Bay ecosystem. *Ecological Monographs*, *59*, 329–364.
Blough, N., Zafiriou, O., & Bonilla, J. (1993). Optical Absorption Spectra of Waters From the Orinoco River Outflow: Terrestrial Input of Colored Organic Matter to the Caribbean. *Journal of Geophysical Research*, *98*, 2271–2278.
Blume, H., & Fedors, J. (1978). Measurement of ocean temperature and salinity via microwave radiometry. *Bound-Layer Meteorology.*, *13*, 295–308.
Bowers, D., & Brett, H. (2008). The relationship between CDOM and salinity in estuaries: An analytical and graphical solution. *Journal of Marine Systems*, *73*(1–2), 1–7.
Breiman, L., Freidman, J., Olshen, R., & Stone, C. (1998). *Classification and regression trees.* Boca Raton: Wadsworth.
Brownlee, K. (1960). *Statistical theory and methodology in science and engineering.* New York: John Wiley & Sons, Inc..
Cameron, A., & Trivedi, P. (1998). *Regression analysis of count data.* New York: Cambridge University Press.
Chipman, H., George, I., & McCulloch, R. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics.*, *4*, 266–298.
D'Sa, E., & Miller, R. (2003). Bio-optical properties in waters influenced by the Mississippi River during low flood conditions. *Remote Sensing of Environment*, *84*, 538–549.
Del Castillo, C., Coble, P., Morell, J., Lopez, J., & Corredor, J. (1999). Analysis of the optical properties of the Orinoco River Plume by absorption and fluorescence spectroscopy. *Marine Chemistry*, *66*, 35–51.
Del Vecchio, R., & Blough, N. (2004). Spatial and seasonal distribution of chromophoric dissolved organic matter and dissolved organic carbon in the Middle Atlantic Bight. *Marine Chemistry*, *89*(1–4), 169–187.
Devore, J. L. (1995). *Probability and statistics for engineering and the sciences* (3 rd ed.). Pacific Grove: Brooks/Cole.
Friedman, J. (1991). Multivariate adaptive regression spline. *The Annals of Statistics*, *19*, 1–141.
Geiger, E., Grossi, M., Trembanis, A., Kohut, J., Oliver, M. (in press). Satellite-Derived Coastal Ocean and Estuarine Salinity in the Mid-Atlantic. *Continental Shelf Research.*
Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, *1*, 297–310.
Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). : Springer-Verlag.
Hoffman, M., Haine, T., Miyoshi, T., Ide, K., Brown, C., Murtugudde, R. (in review). An Advanced Data Assimilation System for the Chesapeake Bay: Performance Evaluation. *Journal of Atmospheric and Oceanic Technology.*

Kachan, M., & Pimenov, S. (1997). Remote sensing of water salinity at decameter wavelengths. *IEEE Transactions on Geoscience and Remote Sensing*, *35*, 302–306.
Khorram, S. (1982). Remote Sensing of Salinity in the San Francisco Bay Delta. *Remote Sensing of Environment*, *12*, 15–22.
Lee, K., & Park, J. (1992). Short-term load forecasting using an artificial neural network. *Transactions on Power Systems.*, *7*(1).
Lerner, R., & Hollinger, J. (1977). Analysis of 1.4 GHz radiometric measurements from Skylab. *Remote Sensing of Environment*, *6*, 251–269.
Maisonet, V., Wesson, J., Burrage, D., & Howden, S. (2009). *Measuring coastal sea-surface salinity of the Louisiana shelf from aerially observed ocean color*. Conference proceedings: Oceans 2009 MTS/IEEE. Biloxi, Mississippi.
Marlyand Department of Natural Resources (2011). *Chesapeake Bay monitoring.* Web access: October 4, 2011. http://www.dnr.state.md.us/bay/monitoring/
McKeon, J., & Rogers, R. (1976). *Water quality map of Saginaw Bay from computer processing of Landsat-2 data.* Spec. report to Goddard Space Flight Center, Greenbelt, Maryland.
Morel, A., & Gentili, B. (2009). A simple band ratio technique to quantify the colored dissolved and detrital organic material ocean color remotely sensed data. *Remote Sensing of Environment*, *113*, 998–1011.
National Oceanic and Atmospheric Administration (NOAA) (2010). *Remote sensing for coastal management. Sea nettle forecast. CoastWatch program, Chesapeake Bay office. Web.* Web access: October 4, 2011. http://chesapeakebay.noaa.gov/forecasting-sea-nettles
National Aeronautical Space Administration (NASA) (2011). *MODIS information page. Web.* Web access: May 1, 2011. http://modis.gsfc.nasa.gov/
Nelder, J., & Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A*, *135*, 370–384 (General).
Pritchard, D. (1952). Salinity distribution and circulation in the Chesapeake estuarine system. *Journal of Marine Research*, *11*, 106–123.
R Development Core Team (2011). *R: A language and environment for statistical computing.* R foundation for statistical computing http://www.R-project.org
Sutton, C. (2005). Classification and regression trees, bagging, and boosting. *Handbook of statistics. Data mining and data visualization, Vol. 24*, . PA: Penn State University.
United States Geological Survey (2012). *USGS real-time water data for the nation.* USGS 01578310 Susquehanna River at Conowingo, MD. Web Web access: February 10, 2012. http://waterdata.usgs.gov/usa/nwis/uv?01578310