EPA/600/R-15/003 | September 2015 | www2.epa.gov/research

United States
Environmental Protection
Agency

Predictive Seagrass Habitat Model

Office of Research and Development
National Health and Environmental Effects Research Laboratory

EPA/600/R-15/003|September 2015

Predictive Seagrass Habitat Model

Prepared by:

Naomi E. Detenbeck and Steven Rego

Atlantic Ecology Division

National Health and Environmental Effects Research Laboratory

Narragansett, RI 02882

National Health and Environmental Effects Research Laboratory

Office of Research and Development

U.S. Environmental Protection Agency

Washington, DC 20460

Notice

The development of the model described in this document has been funded by the U.S. Environmental Protection Agency (EPA), with GIS support for input data preprocessing received under SES3 BPA GS-35F-4594G, Task Order 1520 Remote Sensing and Geographical Information Systems Support (RSGIS) TDD 2-1 Narragansett. This report has been subjected to the Agency's peer and administrative review and has been approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. James Hagy, Giancarlo Cicchetti, and Michael McManus provided comments on a preliminary version of this report.

Table of Contents

Appendices

Figures

Tables

Appendices Figures and Tables

Acronyms

AIC	Aikake's Information Criterion
ALU	Aquatic life use
ANN	Artificial neural network model
AUC	Area Under the Curve
BART	Bay Assessment and Response Team
BRT	Boosted Regression Tree
C	Centigrade
CBP	Chesapeake Bay Program
C-CAP	Coastal Change Analysis Program
CMAQ	Community Multi-Scale Air Quality (Model)
CRM	Coastal Relief Model
CVI	Coastal vulnerability index
DEM	Digital elevation model
DIN	Dissolved inorganic nitrogen
EDM	Estuary Data Mapper
EPA	Environmental Protection Agency
ESI	Environmental Sensitivity Index
ESRI	Environmental Systems Research Institute, Inc
GAM	Generalized additive model
GAMM	Generalized additive mixed model
GIS	Geographic information system
GLM	Generalized linear model
GLMM	Generalized linear mixed model
K_d	Vertical light attenuation coefficient
LIDAR	Light Detection And Ranging
MA	Massachusetts
MARS	Multi-adaptive regression spline
MCMC	Markov Chain Monte Carlo
MGET	Marine Geospatial Ecology Tools
MHHW	Mean higher high water
MHW	Mean high water
MLLW	Mean lower low water
MLW	Mean low water
MSL	Mean sea level
MTL	Mean tide level
N	Nitrogen
NASA	National Aeronautical Space Administration
NAVD	North American Vertical Datum
NDBC	National Data Buoy Center
NGDC	National Geophysical Data Center
NGVD	National Geodetic Vertical Datum
NOAA	National Oceanic and Atmospheric Administration

ORD	Office of Research and Development
P/A	Presence/Absence
POC	Particulate organic carbon
PSU	Practical salinity unit
PTSI	Preliminary transplant suitability index
R	Software package
RIDEM	Rhode Island Department of Environmental Management
RIGIS	Rhode Island Geographic Information System
RMSE	Root mean square error
ROC	Receiver operating characteristic
SAC	Spatial autocorrelation
SAV	Submerged aquatic vegetation
SD	Secchi depth
SS	Suspended solids
SST	Sea Surface Temperature
STD	Standard deviation
TN	Total nitrogen
TOC	Total organic carbon
TSI	Transplant suitability index
US	United States
US EPA	U.S. Environmental Protection Agency
USGS	U.S. Geological Survey
VIF	Variance inflation factor
WAVES	USGS program to calculate wave statistics
WBID	Water body identification code
WEMo	Wave Energy Model
WGS	World Geodetic System
WWTP	Wastewater treatment plant
Z_c	Light compensation depth

Executive Summary for Coastal Managers

We developed an approach for creating statistical models to predict seagrass presence/absence at the scale of individual grid cells (10 m x 10 m) as well as a number of endpoints relative to seagrass presence/absence along transects perpendicular to the shoreline: presence/absence, relative frequency along transects, and minimum and maximum depth of occurrence. Unlike many of the statistical models that have been developed to predict seagrass occurrence, ours take into account the nonrandom distribution of seagrass (patchiness), nonlinear effects of light availability, salinity (as an indicator of total N gradients), sediment organic carbon, and parameter interactions.

We tested this modeling approach using data for distribution of seagrass (*Zostera marina*) in Narragansett Bay, RI. Models developed were most robust for predictions of shoreline occurrence rather than at the 10 m x 10 m grid scale. Based on model results, multiple factors affect seagrass success. The minimum depth of occurrence is influenced by both wave energy (wave mixing depth) as well as sediment particle size, with finer sediments more susceptible to disturbance by wave action. The maximum depth of occurrence is influenced by both transparency and sediment organic carbon, an indicator of past eutrophication. Depth limits decrease as sediment organic carbon increases due to increased energetic demands for seagrass to counteract effects of increased toxicity of anaerobic sediments (e.g., sulfide toxicity). Our models allowed us to distinguish multiple modes of action for nitrogen effects on seagrass distribution: 1) shading by phytoplankton affecting Secchi depths, 2) shading and/or competition by periphyton and macroalgae mediated by nitrogen concentrations, and 3) effects of sediment organic carbon on minimum light requirements. Incorporation of historic distribution of seagrass patches did not improve model predictions, suggesting that patches may exist in a state of dynamic equilibrium with a lag time for recovery of disturbed sites.

Our models detected significant differences in the probability of seagrass occurrence by shoreline even after we factored out the effect of site characteristics. This could be explained by hysteresis effects related to tidal currents. Tidal currents in Narragansett Bay are strong enough to resuspend fine sediments, thus limiting establishment of new seagrass patches, but not strong enough to damage established seagrass beds. This suggests that better monitoring is needed for assessment of turbidity plumes near the bottom of the water column, not just water column transparency. If hysteresis exists, this has implications for recovery strategies as well. Specific restoration measures such as co-restoration of shellfish beds (to reduce suspended sediments and effects of tidal currents and wave action) or use of existing or constructed coastal barriers to limit effects of wave action and tides might improve probabilities of initial colonization success and the initiation of positive feedback effects.

Our predictive models have multiple potential applications: identification of aquatic life use zones for setting nutrient criteria for areas of potential seagrass habitat, prioritization of areas and strategies for seagrass restoration, and projection of potential benefits of management actions. We applied our model to predict the potential recovery of seagrass given a 40% decrease in total N loading from wastewater treatment plants and atmospheric deposition (assuming an equivalent reduction in water column

concentrations). Based on the current model, the colonized area for all shorelines combined following a 40% reduction in TN loads (and concentration) would increase from 12% of area in the 0 to 5 meter depth zone to about 63% of area in the short term and slightly more in the long term (as sediment organic carbon levels recover). Adaptive management will need to take into account different projections for short-term versus long-term recovery due to the multi-decadal persistence of organic carbon in sediments and effects on minimum light requirements for seagrass.

Abstract

Restoration of ecosystem services provided by seagrass habitats in estuaries requires a firm understanding of the modes of action of multiple interacting stressors including nutrients, climate change, coastal land-use change, and habitat modification. Often, managers have used the reported historic depth limits of seagrass to project the future distribution of seagrass in response to nitrogen load reductions. In general, these predictions are based on empirical or modeled estimates of the influence of phytoplankton production in the water column on the light environment, and do not account for the interaction of multiple factors. We explored the application of generalized linear mixed models (GLMMs) and generalized additive mixed models (GAMMs) to describe the simple and interactive effects of environmental factors on the distribution of a common seagrass, *Zostera marina*, in Narragansett Bay, Rhode Island. We used a random shoreline effect to account for "founder" (random colonization or extinction) effects. We provide several strategies to overcome three challenges in developing empirical species distribution models to describe and predict seagrass distribution in estuaries: the fine-scale patchiness of seagrass distributions with attendant problems of spatial autocorrelation; the large areas of interest for model development and application entailing significant memory demands for modeling; and the potential co-variance of multiple interacting factors affecting seagrass. We developed a spatial framework describing the coordinates of spatial autocorrelation in estuarine systems, with the main axis parallel to the shoreline and a secondary axis perpendicular to the shoreline. We demonstrated an approach to incorporate a term for residual autocorrelation in GLMMs first introduced by Crase (Crase, B., Liedloff A.C., and B.A. Wintle. 2012). To account for anisotropy in the system, we calculated zonal averages of residual errors within rectangular boxes oriented parallel to the shoreline along the longer main axis. We successfully dealt with covariance of influential factors by centering variables, by using multiple strategies to describe the interaction of the light environment and wave energy with depth, and by excluding correlated variables where necessary. We predicted seagrass distribution at the scale of 10-meter grid cells, as presence/absence or average presence/absence associated with shoreline locations spaced at 10-meter intervals, and minimum or maximum depth of distributions at those locations. Prediction of seagrass absolute or average presence/absence at shoreline locations was very robust, with area-under-the-curve (AUC) values associated with Receiver Operating Characteristic (ROC) curves of 0.95 – 0.98 following 10-fold cross-validation of models. Random shoreline effects varied over several orders of magnitude, probably tied to the distribution of tidal currents. Tidal currents are

weak enough to allow persistence of existing seagrass beds, but strong enough to interfere with successful recolonization through resuspension of fine sediments. For the model predicting seagrass presence/absence at the grid cell scale, the most influential predictor is Secchi depth, followed by (in order): shoreline isolation, sediment percent total organic carbon, sediment type, and salinity. The least influential variable is water depth greater than average wave mixing depth. For the model predicting presence of seagrass at shoreline locations, the most influential predictor is sediment type, followed by sediment percent total organic carbon (at low Secchi depth), then salinity (as an indicator of downstream gradients in water column total nitrogen). As demonstrated in other recent studies, sediment total organic carbon interacts with light availability by increasing energy requirements and the light compensation point for seagrass. For all shorelines combined, our model predicts that following a 40% reduction in TN loads (and concentration) the colonized area would increase from 12% of area in the 0 to 5 meter depth zone to about 63% of area in the short term and slightly more over subsequent decades as sediment organic carbon recovers. Finally, we provide data sources for application of this approach to other U.S. estuaries, with much of the data available through EPA's Estuary Data Mapper application (www.epa.gov/edm).

Keywords: eelgrass; seagrass; *Zostera marina*; estuary; Narragansett Bay; species distribution model; generalized linear mixed model; spatial autocorrelation

Chapter 1. Introduction

1.1 Purpose

Seagrasses are essential in providing valuable ecosystem services but are in decline globally (Orth et al. 2006). Compton et al. (2011) identified loss of seagrass habitat as one of the most costly impacts related to nitrogen loading based on the relationship between loss of submerged aquatic vegetation and fishery declines in estuaries. Restoration of ecosystem services provided by seagrass habitats in estuaries requires a firm understanding of the modes of action of multiple interacting stressors including nutrients, climate change, coastal land-use change, and habitat modification. Managers often use the reported historic depth limits of seagrass to project the future distribution of seagrass in response to nitrogen load reductions based on empirical or modeled estimates of the influence of phytoplankton production in the water column on the light environment (Dennison et al. 1993). However, this approach does not account for the interaction of multiple factors, so more comprehensive models are needed.

Recent discussions among Environmental Protection Agency (EPA) Office of Water, EPA Regions, EPA Office of Research and Development (ORD), and states have highlighted the need for states to refine aquatic life use (ALU) definitions. The Chesapeake Bay Program (CBP) has provided one example of tailoring ALUs to reflect different expectations and habitat support functions for different zones within an estuary (US EPA 2003). The Maryland Coastal Bays National Estuary Program has used seagrass potential habitat as a target for seagrass distribution and as a normalizing factor to describe seagrass coverage by estuarine segment (Wazniak and Hall 2005). Their mapping of potential habitat considers only two factors: depth and % silt/clay in sediment. An extension and refinement of the CBP conceptual model for ALU definitions by region and habitat zone for different estuary types could help foster state efforts to refine ALUs. Defining appropriate ALUs provides the foundation for setting water quality criteria (including nutrient criteria) by establishing targets or expectations for ecosystem condition in the absence of pollution.

Support for submerged aquatic vegetation (SAV) defines one of the potential habitat uses in the CBP model for aquatic life use (US EPA 2003). To extend this model to estuaries in other regions, we need to define habitat constraints (essentially a habitat suitability index) for seagrass species dominant over different geographic ranges along the US coast using readily available data. The regulatory objective of interest is to provide a scientific framework to support nutrient criteria and restoration plans by defining spatially-explicit targets for expected condition (seagrass habitat presence/absence) in estuaries. Many water quality criteria are based on an initial definition of the use (including ALU) that should be supported. Once areas suitable for seagrass habitat are identified, the state agencies responsible for setting criteria can determine appropriate targets for seagrass coverage within a system.

The purpose of this report is to:

1) assess predictive modeling approaches for seagrass presence-absence,
2) discriminate the effects of nutrient enrichment from effects of other stressors and co-factors on seagrass occurrence to foster improved management of estuarine systems, and
3) provide users with easy access to input data for predictive models of seagrass habitat on regional and national scales.

Our modelling approach can be applied to other estuaries, although specific data sources, variables included, and best models selected may differ across systems. We illustrate the approach outlined for predictive seagrass habitat models with a case study using data for the Narragansett Bay estuary in Rhode Island. The resultant models include both variables related to different mechanisms for nutrient effects on seagrass as well as other cofactors and stressors which limit the distribution of seagrass. We show how the model can be applied to assess the potential improvement in seagrass coverage following nutrient load reductions. We designed the approach to be generic and adaptable to other estuaries. To facilitate this process, we provide information on data sources and examples of R programming code in the appendices to this report.

1.2 Conceptual Model

Some of the main factors affecting seagrass presence/absence are water depth, light availability, temperature, salinity, energy regime, substrate type, sediment sulfide content, and macro-algal coverage (Burkholder et al. 2007, de Boer 2007, Koch et al. 2007, Lee et al. 2007, Ralph et al. 2007, Touchette 2007). Many of these factors can be readily monitored in the field or calculated from measured parameters, mapped, and recorded as digital geographic information system (GIS) coverages, while others (sulfide concentration, macro-algal coverage) are not generally available from monitoring programs. Site-specific disturbances (wasting disease, wetland dredge/fill activities, boat anchors, grazing, hurricanes) can also limit the distribution of seagrass species but are not predictable, and so must be accounted for on a case-by-case basis (Short et al. 1987, 1988, Neckles et al. 2005, Rivers and Short 2007, Oakley et al. 2013). We focus our habitat suitability model on factors for which georeferenced data are readily available, and for which optima or thresholds for seagrass growth and survival are available from the literature: wave energy, light availability at the sediment surface, substrate type (particle size, organic matter content), salinity, and temperature.

1.3 Existing Approaches

1.3.1 Preliminary Transplant Suitability Index:

Short and Burdick (2005) developed a software application based on their report on a preliminary transplant suitability index (PTSI) for screening potential seagrass transplant sites. This modeling software takes data availability into account, and requires some on-site field measurements to assess light and bioturbation. The modeling process was performed in three stages: the PTSI was developed using readily available data; field data were collected to assess light conditions and bioturbation; and then the final Transplant Suitability Index (TSI) score was developed. The PTSI development process uses a set of

parameters (Table 1) and is calculated by multiplying the ratings for each parameter. The higher the score, the better the likelihood that seagrass will survive transplantation at that site. The TSI does not include sites that were rated zero. Sites close to existing seagrass beds receive lower scores not because they have a low likelihood of success but because transplantation to these sites will not increase the geographic extent of existing seagrass.

Table 1. Potential Transplant Suitability Index (PTSI) data listing.

Parameter	PTSI rating	Reference
Historical eelgrass distribution	1: previously unvegetated	Fonseca et al. (1998)
	2: previously vegetated	
Current eelgrass distribution	0: currently vegetated	
	1: currently unvegetated	
Proximity to natural eelgrass bed	0: < 100m	Orth et al. (1994)
	1: > 100m	
Sediment type	0: rock or cobble	Kenworthy & Fonseca (1977) Short et al. (1987, 1993)
	1: > 70% silt/clay	
	2: cobble free with < 70% silt/clay	
Wave exposure (calculation)	0: > mean + 2 STD	Kopp et al. (1994), Murphey & Fonseca (1995), Fonseca et al. (1998)
	1: < mean + 2 STD	
Water depth	0: shallow	Short et al. (1993)
	1: shallow edge of bed	
	2: average of bed	
	1: deep edged bed	
Water quality	0: poor	Batiuk et al. (1992), Dennison et al. (1993), Costa et al. (1996)
	1: fair	
	2: good	

1.3.2 Predictive modeling approaches

Predictive modeling approaches for seagrass habitat range from simple empirical approaches to more detailed mechanistic modeling approaches. The simplest approach involves predicting presence/absence of seagrass using generalized linear models (GLM; van der Heide et al. 2009) or generalized additive models (GAM) (Downie et al. 2013). The latter allow nonlinear effects to be incorporated into models. One variant of these models

has been proposed to predict potential habitat based on presence only, e.g., using the Maxent software (Downie et al. 2013). More complex statistical approaches include Bayesian models which can incorporate prior knowledge of parameter distributions (March et al. 2013), and boosted regression trees (BRTs), which can handle both nonlinear effects and parameter interactions for a large number of predictive variables (Crase et al. 2012). Some investigators have incorporated not only the potential effects of the physical environment on individual plants, but also emergent properties influencing the accelerated growth of seagrass patches (Kendrick et al. 2005). The most complex process-based modeling approaches incorporate specific mechanisms underlying growth, loss rates, and interactions between seagrass and their physico-chemical environment. These processes can be incorporated into dynamic cellular automata models that can mimic the spread of seagrass patches with physical and biological feedback loops (Wortman et al. 1997, Fonseca et al. 2000b, Fonseca et al. 2004, Schonert and Milbradt 2005).

1.4 Study Site (Narragansett Bay, RI)

Narragansett Bay is one of the largest estuaries in southern New England, running from north to south along the state mid-line. The upper eastern portion of the estuary, including part of Mount Hope Bay, is in Massachusetts. The Bay's surface area is approximately 380 km^2 with 618 km of tidal shore (Figure 1). Much of the Bay is shallow and well mixed but depth varies, and descends to 40 m in some of the deepest channels.

Narragansett Bay's eutrophication problems began in the early to mid 1900s when increased development and sewering led to increased nitrogen loading to the Bay (Nixon and Pilson 1983). More recent data suggest, however, that the bay's nutrient concentrations are declining (Oviatt et al. 1995). The bay receives much of its nitrogen inputs from wastewater treatment plants near the head of the Bay, and nutrient concentrations exhibit a strong north to south gradient (Oczkowski et al. 2008). In general, Narragansett Bay remains well-mixed vertically under normal climatic conditions, and has a mean hydraulic residence time between 10 and 40 days depending on the time of year and on location in the Bay (Howarth 1988, Pilson 1985). Significant runoff events, particularly during neap tides, can induce the onset of stratification in the Bay by strengthening vertical salinity gradients.

Seagrasses were widespread in Narragansett Bay throughout the 1800s and early 1900s (Kopp et al., 1994). During the early to mid–1900s a precipitous decline in seagrasses was observed globally and in the New England Region. Many research studies suggest that decline may have been due to increases in environmental contaminants related to eutrophication, herbicide use, and increased turbidity. Reported instances of widespread infection (wasting disease) from *Labryinthula zosterae* (slime mold) in Narragansett Bay may have also contributed to seagrass decline (Short et al. 1987, 1988).

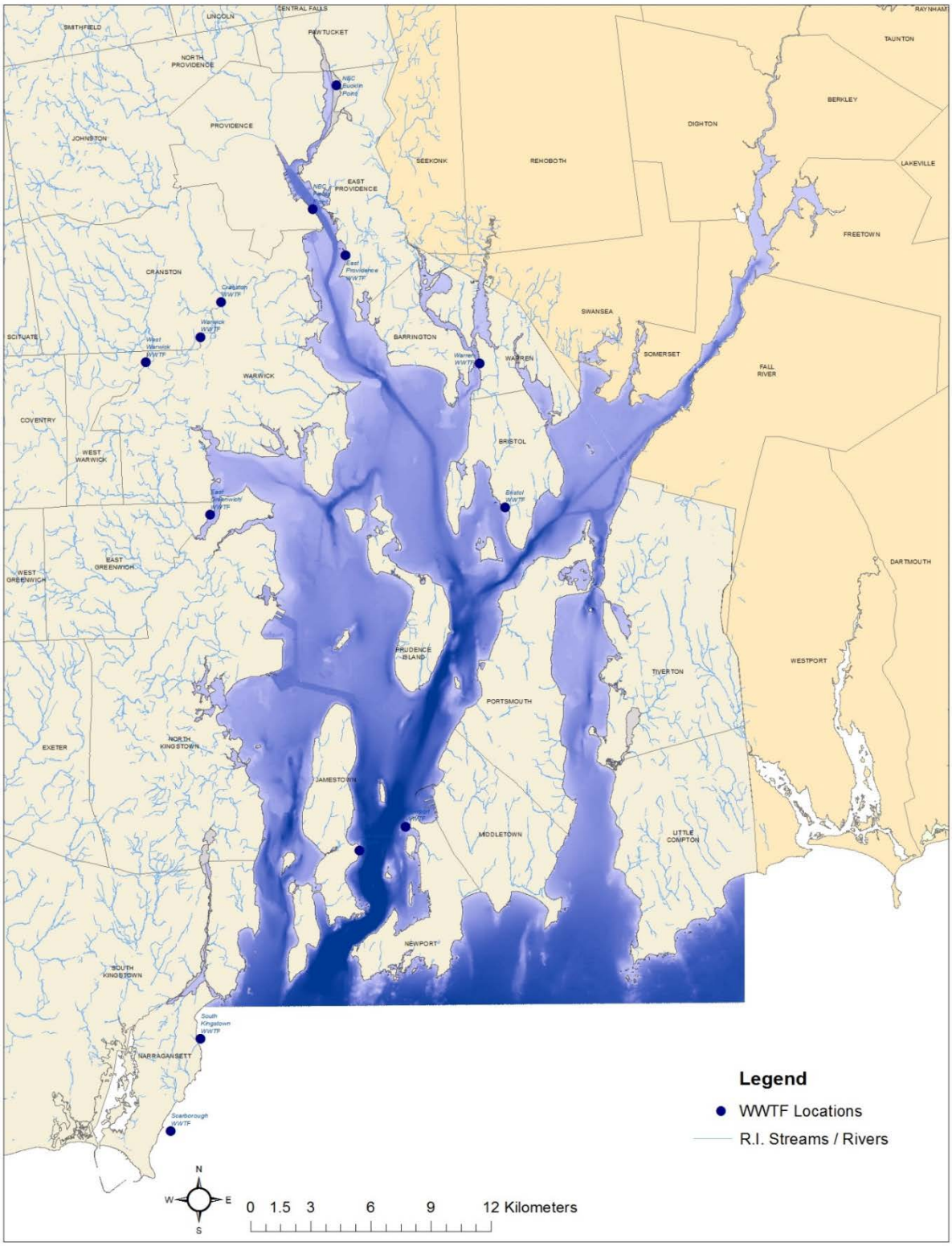

Figure 1. Map of Narragansett Bay and Rhode Island with color-coded bathymetry. WWTF = Wastewater treatment facility.

Chapter 2. Methods

2.1 Data Sources

We identified potential predictors of seagrass habitat based on information in the literature regarding previous modeling efforts, as well as the public availability of data (Appendix A). Detailed information on how to access similar data for other systems is provided in Appendix B. We selected data to be contemporaneous or within $\pm$ 5 years of reported seagrass field sampling whenever possible, focusing on data from the primary growing season, May through October. We reviewed data and metadata from each source for completeness and applied further geoprocessing if needed. We converted each data source to a grid with 10 x 10 meter cells, clipped the grid to the zero to five meter depth zone in Narragansett Bay. This depth zone represents the maximum potential extent of seagrass if no optically-active constituents are present in the water column (i.e., no light extinction due to chlorophyll, turbidity, or dissolved organic matter). Model points were sampled from the centroids of these grid cells within the 0 to 5 meter depth band (Figure 2a, b). This yielded 18,612 grid cells with seagrass present and 2,725,719 grid cells with seagrass absent.

2.2 Data Pre-processing

2.2.1 Salinity

We derived salinity data for Narragansett Bay from a RI Department of Environmental Management (RIDEM) dataset collected by the Bay Assessment and Response Team (BART) (www.ri.dem.gov/bart/netdata.htm). There are 12 BART sampling buoys that traverse the Bay from north to south. These data included salinity values collected at 15-minute intervals from a variety of stations between 2003 and 2012. Weekly salinity data from 2006 through 2012, were downloaded as excel spreadsheets, compiled, and averaged. We imported data into ArcGIS as points and created Thiessen polygons from original point data to fill in spatial data gaps. These data were sufficient to capture the gradient of salinity from north to south but not detailed enough to represent variation at the scale of smaller subembayments.

2.2.2 Temperature

We downloaded daily sea surface temperature (SST) data from NASA's multi-scale ultra-high resolution sea surface temperature remote sensing product (MUR SST; ftp://podaac-ftp.jpl.nasa.gov/allData/ghrsst/data/L4/GLOB/JPL/MUR/) using the EPA's Estuary Data Mapper (EDM) interface (www.epa.gov/edm), and calculated weekly mean temperature values for the years 2003 – 2012. In areas of the estuaries near the shoreline that were not covered by the SST data, we applied a Euclidean allocation method in ArcMap to fill in areas without data. Euclidean allocation was used rather than focal statistics to fill in data gaps because Euclidean allocation uses the closest cell data to the analysis point rather than using a large rectangle to calculate a mean value (as is the case with focal statistics). The resultant grid is a more realistic representation of sea surface temperature.

2.2.3 Sediment Type (Grain Size and Percent Total Organic Carbon)

We obtained sediment particle size classes as ArcGIS shapefiles from the RIDEM Narragansett Bay Estuary Program that were based on McMaster's collection of 493

surficial samples (1960) (http://www.narrbay.org/biological_data.htm). Sixteen sediment classes were available, including two missing data classes; 'not mapped' and 'not sampled'. Size classes included; gravel, sandy gravel, gravel sand silt, sand, gravelly sand, silty sand, silt, sandy silt, clay silt, sand silt clay, gravel silt clay and rock. No seagrass occurred in association with four of the original 14 sediment classes mapped; these four classes were relatively rare across the shallow water zone of Narragansett Bay.

a) b)

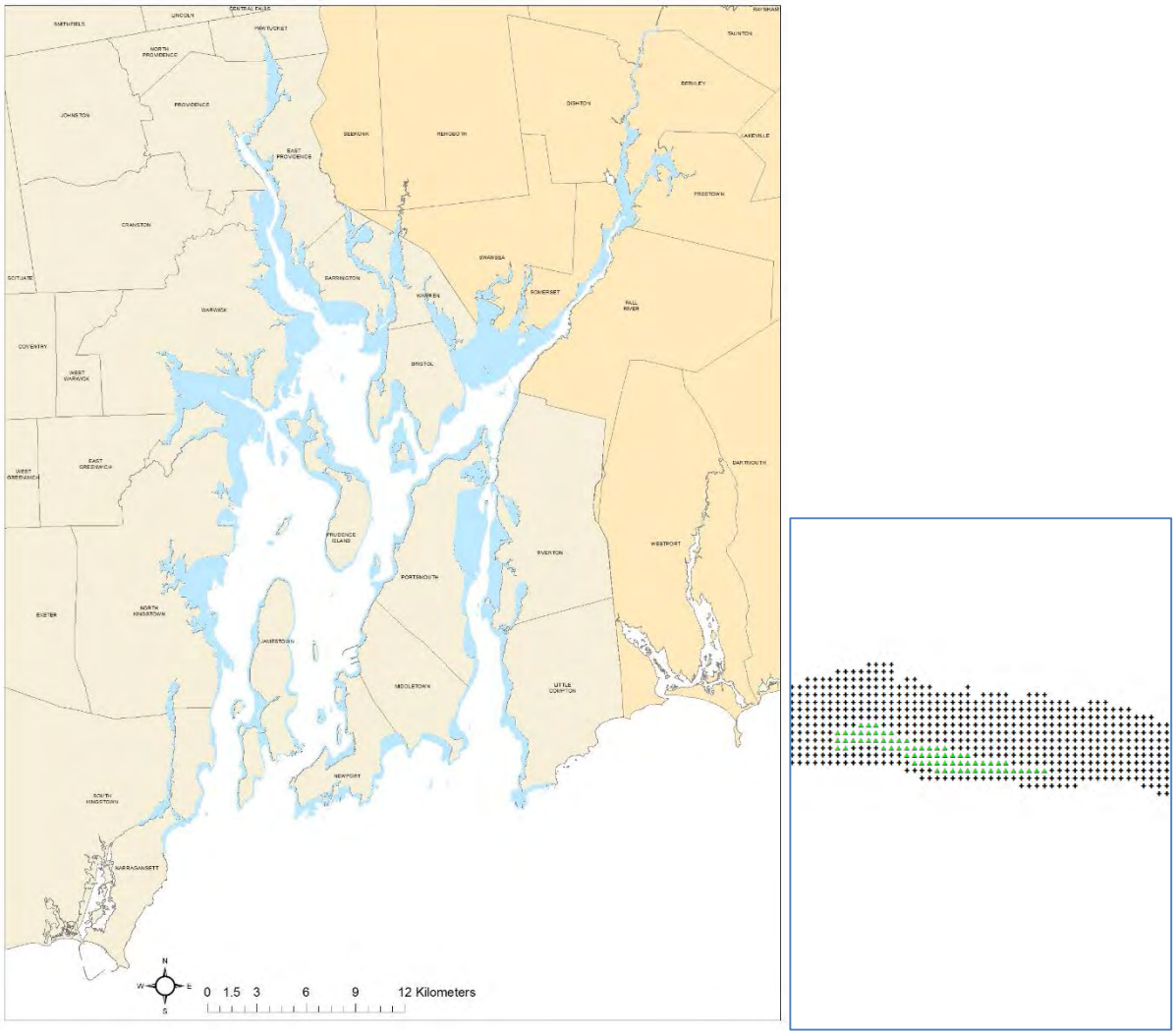

Figure 2a. Map of zone delineating boundaries of seagrass presence/absence grid (0 to 5 meter depth) in Narragansett Bay and b) close-up of 10-meter grid cells with seagrass presence (green) and absence (black). In Figure 2b, the shoreline is to the south of the grids.

We dropped the Rock class from the region of interest for modeling and combined the Sandy Gravel and Gravel-Sand-Silt classes (from which seagrass was entirely absent) with the Gravel class. We combined the Silt class (from which seagrass was absent) with the Sandy Silt class. Thus up to eight classes were used in the analysis.

We estimated total organic carbon was estimated across the Bay based on 119 surficial sediment grabs collected by the US EPA National Coastal Assessment (http://oaspub.epa.gov/coastal/coast.search). We interpolated values to create a complete grid within the shallow-water zone using inverse distance weighting in ArcMap 10.1 (©ESRI, Redlands, CA).

2.2.4 Seagrass (Current/Historical)

We obtained data layers for recent and historic eelgrass coverages for Narragansett Bay from Rhode Island Geographic Information System (RIGIS)(2000a, b, 2013) as ArcGIS shapefiles. Both the 1999 (RIGIS 2000, ab) and the 2006 RIGIS (Current seagrass) data were collected by the state of RI using NOAA Coastal Change Analysis Program (C-CAP) protocols (http://coast.noaa.gov/digitalcoast/_/pdf/ccap-products.pdf), so these datasets should be comparable. The true color imagery was recorded and analyzed at a scale of 1:12000 for both. The 1999 data were collected on July 6th and the 2006 data were collected on August 6th.

If eelgrass patch locations are stable from year to year, and patches progressively shrink or expand, then historic eelgrass locations should be a good predictor of current presence/absence. We intersected model points with historic seagrass coverage (1996-1998) to yield a presence/absence dummy variable (egPA99). To create a composite of patch boundaries, we converted seagrass polygons to lines and merged these with historic seagrass line coverages (representing seagrass patches less than 40 feet in width). Then we calculated distance to nearest historic seagrass patch boundary for each model point (DistToEG99). We also calculated the area of each nearest historic seagrass patch (AREA) as a potential explanatory variable.

We captured the effect of more distant historic events (wasting disease incidence in the 1930s and subsequent recolonization of seagrass up through the 1960s) in the models with a random "shoreline" effect. Because seagrass beds expand predominantly through vegetative growth rather than through reproduction by seeds, the growth of seagrass patches is most likely to occur along shorelines, areas of contiguous potential habitat. Each of 18 contiguous shorelines within Narragansett Bay was assigned a code, considering different potential current delivery vectors (Figure 3). We assigned a SHORLIN code of -99 to minor shorelines associated with small islands not connected to the mainland by suitable habitat (0 to 5 meters depth) as well as a dummy variable for ISOLATED status.

2.2.5 Transparency

We described transparency across the Bay using three data sources. We combined Secchi depths measured by the Narragansett Bay Commission between 2008 and 2012 (http://snapshot.narrabay.com/app/MonitoringInitiatives/WaterClarity) with measurements collected by the US EPA's National Coastal Assessment program between 2000 and 2006 (http://oaspub.epa.gov/coastal/coast.search). We averaged values by the Water Body Identification Code (WBID) estuarine segments used for assessment and listing by RI DEM. To avoid biasing segment averages, we removed Secchi depths greater than the maximum depth at the site of measurement from the records before averaging. We filled in gaps in transparency data along the southern shore of Conamicut Island (Sakonnett and Newport Bays) using K_d estimates from offshore MODIS satellite imagery. (Conamicut

Island is the large island near the mouth of the western arm of Narragansett Bay, Shoreline code 14.) We downloaded the latter data using the EDM tool and averaged over the growing season (May – October) for the years 2008-2012. We extracted average grid cell values for locations greater than 30 meters in water depth, and averaged values over a swath of offshore cells parallel to the south shore of Conamicutt Island[1]. We estimated Secchi depth from K_d values using an empirical relationship developed by Batuik et al. (2000).

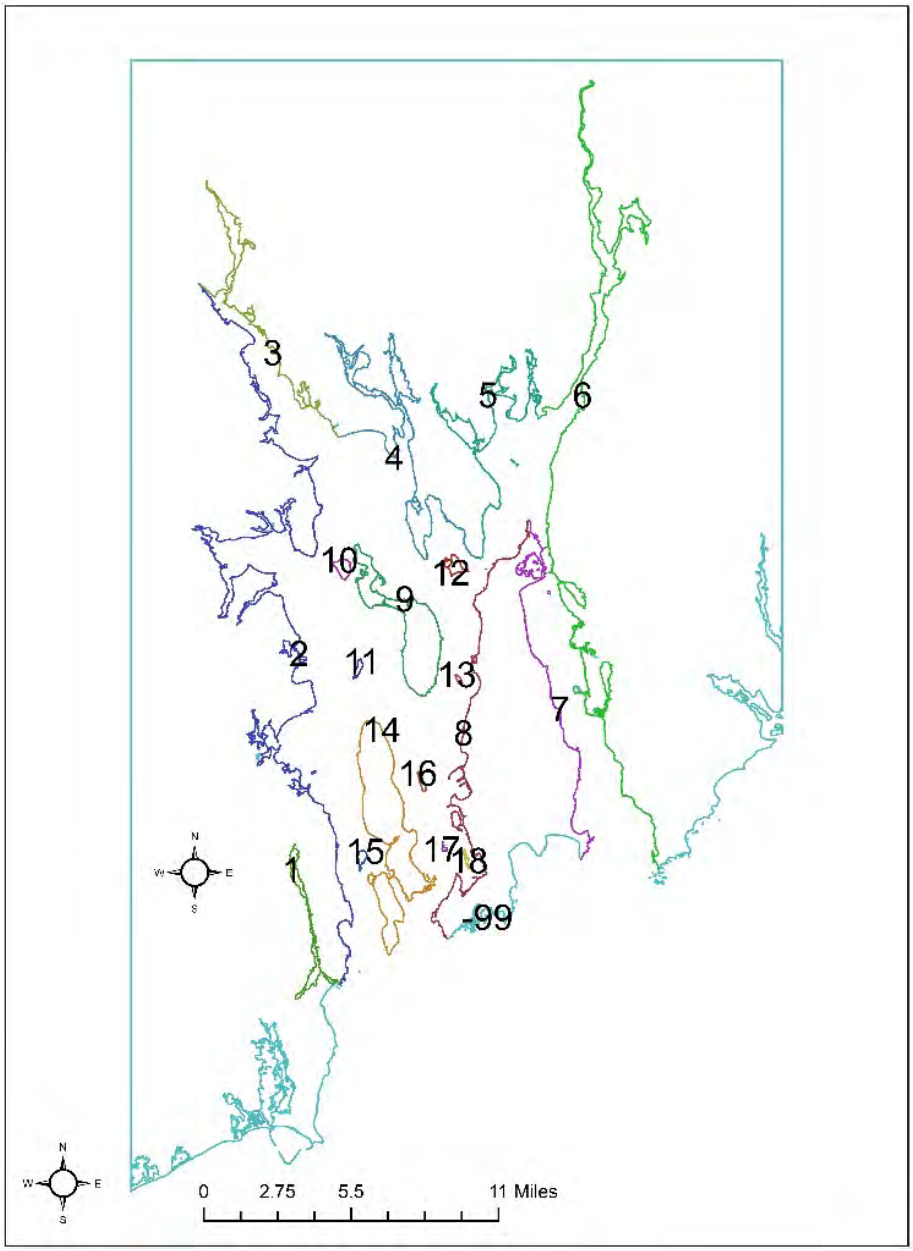

Figure 3. Shoreline code assignment for Narragansett Bay. A value of -99 was assigned to unused segments and isolated shorelines.

[1] Algorithms for K_d are not suitable for shallower water due to interference from bottom sediments.

2.2.6 Wave Exposure Data – WEMo and WAVES models

Assessment of estuarine wave energy is an important component of seagrass modeling. Wave energy relative to fetch and current can affect seagrass habitats in a number of ways, mainly by reducing the ability to establish and maintain effective beds, changing sediment grain size type and or loss, and physically damaging delicate sheaths and leaves. Other ecological processes (e.g., TOC accumulation, bioturbation etc.) are also linked to wave energy and velocities in seagrass habitats. Particle deposition and sediment resuspension are affected by wave energy, but these processes can be attenuated to some degree by established seagrass beds (Garcia et al. 1999).

A variety of tools can be utilized to spatially model fetch and wave energy inputs to estuarine shorelines (Howes et al. 1999, Fonseca and Malhotra 2010, Rohweder et al. 2012). We used two tools to spatially model fetch and wave energy inputs, the National Oceanographic and Atmospheric Administration (NOAA) Wave Exposure Model (WEMo; http://products.coastalscience.noaa.gov/wemo/) for calculation of wave energy, and the United States Geological Survey (USGS) WAVES tool for calculation of mixing depth. Each of these tools has varying input requirements relative to data and formatting, and all require additional pre-processing of input data through modeling applications. WEMo requires a variety of data inputs: closest weather station data (average wind speed and direction), bathymetry data (ESRI ArcGIS GRID format), shoreline edge (ESRI ArcGIS GRID format), and sampling points (generated by the user). We performed WEMo modeling using preset (default) values/conditions as specified in Fonseca and Malhotra (2010). No other WEMo modeling parameters were adjusted except for fetch interrogation distance, which was set to 5000 meters.

We calculated average and maximum wave mixing depths for the 2006-2007 growing seasons using the USGS WAVES extension to ArcMap 10.1 (Rohweder et al. 2012). Chambers (1987) has predicted minimum depth of occurrence of submerged aquatic vegetation in lakes based on calculation of mixing depths from wave action, as one half of the wave length (Chambers 1987). We calculated fetch and wave characteristics using time series of wind speed and direction from the nearest National Data Buoy Center (NDBC) station (Station PTCR1 - 8452951 - Potter Cove, Prudence Island, RI) merged with topobathymetry grids from NOAA's Coastal Relief Model (NOAA National Geophysical Data Center, NGDC Coastal Relief Model, Volume 1-8, http://www.ngdc.noaa.gov/mgg/coastal/coastal.html).

2.2.7 Distance to physical disturbance source: hardened shorelines and marinas

We evaluated two different indicators of physical disturbance related to anthropogenic activities as predictors in regression models: distance to hardened shorelines and distance to marinas. Structural shoreline hardening in Rhode Island includes the use of rock revetments, bulkheads and other types of walls or groins. Much of this hardening took place before coastal regulations existed. Many of the effects of shoreline hardening are very localized, with wave action reflecting off of hard structures causing scour (Shipman 2010). We obtained a GIS layer with locations of hardened shorelines from RIGIS (2003). Other potential sources of information on shoreline hardening (not used here) available for models in other estuaries include the Environmental Sensitivity Index (ESI)

(http://response.restoration.noaa.gov/maps-and-spatial-data/download-esi-maps-and-gis-data.html) and high resolution imagery with coastal LIDAR.

Although mooring beds are permitted throughout Narragansett Bay, there are no common GIS data layers available to document their location. As an indicator of the potential for damage from boat anchors, we calculated the distance to nearest marina based on a data layer obtained from RIGIS (1996) using the NEAR function in ArcGIS.

2.2.8 Unsewered residential development on high infiltration soils

Groundwater inputs to estuaries (with attendant nutrient loads) can be significant, but are generally poorly quantified. If groundwater nutrient inputs are taken up within seagrass beds, they may not be reflected in overlying phytoplankton concentrations. We used the density of unsewered residential development occurring on high infiltration soils in coastal catchments as an indicator of potential groundwater nitrogen inputs to seagrass habitats. The Narragansett Bay Sustainability Pilot Appendix A describes the derivation of data for this indicator (http://www2.epa.gov/sites/production/files/2013-12/documents/nbsp-phase-i-report-appendices.pdf) (data provided by Industrial Economics, Inc.). Grid centroid points were spatially "joined" to the polygon coverage to add this variable to point attributes.

2.2.9 Canada goose grazer density

Canada geese have been identified as a potential impact to seagrass growth and survival in Narragansett Bay (Rivers and Short 2007). The density of geese and their impact will vary by location and seagrass type. We estimated Canada goose density by zone in Narragansett Bay based on winter waterfowl surveys conducted in 2004 – 2006 (http://www.nbnerr.org/waterfowl.htm). Shapefiles of survey zones were provided by Rick McKinney, US EPA Atlantic Ecology Division, Narragansett, RI.

2.3 Statistical Model Development

2.3.1 Strategies to limit memory requirements

Due to the large geographic area covered and fine spatial resolution of seagrass presence/absence grids, we applied several strategies to limit memory requirements for statistical analyses. First, following the best practices outlined for species distribution models by Maggini et al. (2010), we restricted analyses to the geographic range and range of predictor variables associated with points at which seagrass was present in 2006 (Table 2). Second, we limited analyses to the 0 – 5 meter depth zone. Third, we generated models at multiple scales: a) predicting seagrass presence/absence at individual grid cells (grid presence/absence P/A; n = 518,890), b) predicting P/A at any point along transects perpendicular to the shoreline (shoreline segment P/A; n = 19,204), c) predicting relative frequency of occurrence along transects perpendicular to the shoreline (shoreline segment frequency; n =19,204), and d) predicting minimum and maximum depth of occurrence along transects perpendicular to shoreline where seagrass was present (seagrass minimum or maximum depth; n = 2,749; Figure 4).

2.3.2 Modeling approaches at different scales

Depending on the scale of dependent variables defined, we applied different model approaches, different initial sets of independent variables with associated interaction terms and R packages (Table 3, Figure C1 in Appendix C). We evaluated GLM and GAM for minimum and maximum depth endpoints. For all other seagrass endpoints, we developed mixed models (including both random and fixed effects, see below). For mixed models, we evaluated two approaches, the first a general linear mixed model which incorporated random effects related to potential shoreline-specific colonization or disease incidence using the glmmPQL function in the R MASS package (http://stat.ethz.ch/R-manual/R-patched/library/MASS/html/glmmPQL.html). Second, where memory demands were not too large to prevent application, we also analyzed general additive mixed models using the function gamm in the mgcv R package to account for potential nonlinearities in response (https://stat.ethz.ch/R-manual/R-devel/library/mgcv/html/gamm.html).

We modelled "shoreline" as a random effect because seagrass coverage varies significantly among Narragansett Bay shorelines. The "shoreline effect" can be interpreted as the combined result of random colonization (i.e., of wasting disease organisms) and re-colonization (seagrass recovery) effects. We evaluated two different types of mixed effect models with respect to their performance in predicting seagrass presence/absence in

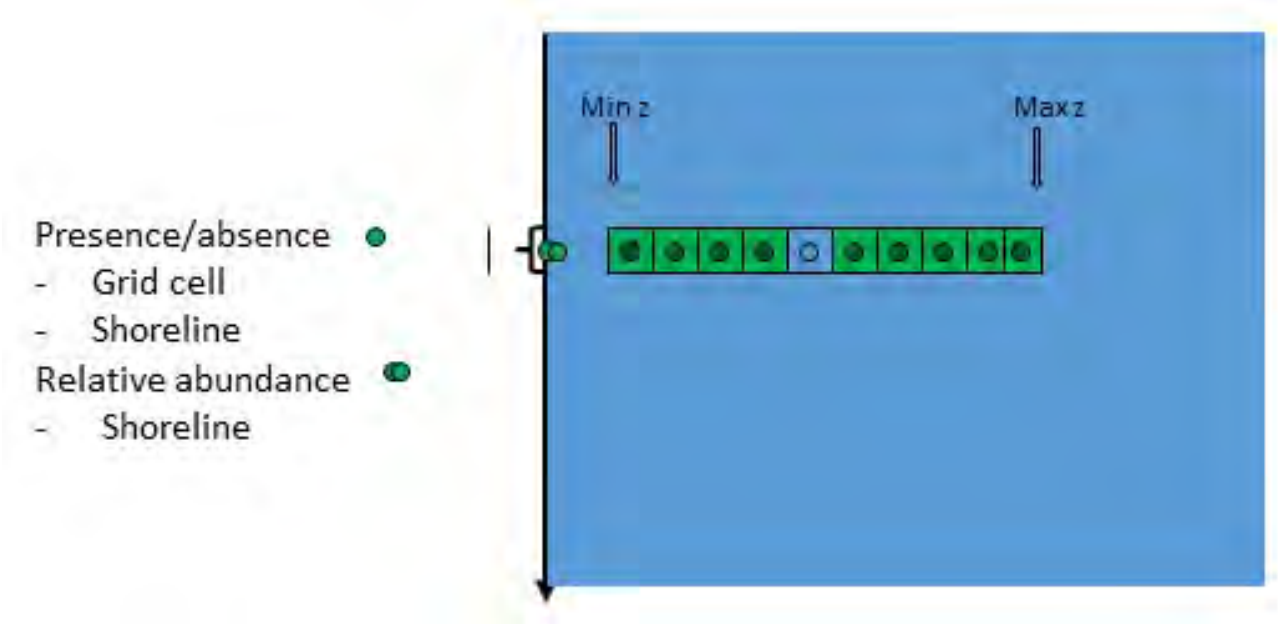

Figure 4. Seagrass endpoints defined at different scales: presence/absence (P/A) at individual grid cell centroids, P/A within transect perpendicular to shoreline (assessed after points were snapped to shoreline), relative frequency of points along perpendicular transect, minimum depth (z) of occurrence along transect, and maximum depth of occurrence along transect.

Table 2. Range of predictors of seagrass habitat for sites at which seagrass is present or absent within Narragansett Bay buffer zone. Shoreline event measure is an index assigned to each 10-meter shoreline segment based on distance from the beginning of a shoreline route.

Variable	Seagrass present (n = 14527)		Seagrass absent (n = 1309011)	
	min	max	Min	max
Depth	0.0	5.1	0.0	6.6
Salinity (PSU)	26.6	28.7	11.3	28.7
Avg Secchi depth (m)	1.3	3.8	0.8	3.8
Min Secchi depth (m)	0.7	3.7	0.5	3.7
Avg SD - water depth (m)	-2.7	3.7	-4.2	3.8
Min SD - water depth (m)	-3.7	3.7	-5.0	3.7
Sediment total organic carbon (%)	0.0	3.5	0.0	7.7
Isolated shoreline	0	1	0	1
Wind wave energy (Joules/m)	0	81724	0	91021
Avg wave mixing depth: water depth	0.002	11.2	0.002	12.4
Max wave mixing depth: water depth	0.002	145.4	0.002	206.6
Temperature (deg C)	14.7	15.0	14.5	15.0
Shoreline event measure (m)	125430	3094040	115340	3957430

Narragansett Bay: general linear mixed models (GLMM) and general additive mixed models (GAMM). We analyzed initial model residuals for evidence of spatial autocorrelation (see below).

2.3.3 Selection of best-fitting models

We applied different strategies to identify best-fitting models for simple GLMs and GAMs as compared to mixed models. For GLMs and GAMs we refined original models (Table 3) using the step function in R. We compared model fits for GLM or GAM using AIC values and prediction errors (Burnham and Anderson 2002, Zuur et al. 2009). For GLMMs, this approach was not possible because AIC values are not provided by the R packages we used. We fit models in a manner analogous to backward stepwise regressions, sequentially eliminating variables from initial full models (Table 3) based on lack of significance (i.e., $p > 0.01$). The glmmPQL package yields approximate p-values which should be used with caution when they are marginal (i.e., near $p = 0.05$), so p-values were interpreted conservatively using a p-value of 0.01. We evaluated final models using a 10-fold cross-validation procedure to test for robustness of results and as a check against overfitting. Sampling with replacement was conducted separately within each shoreline class to provide proportional representation.

Table 3. Initial full models evaluated for each seagrass endpoint. Second- and third-order terms were added to models in later stages only in cases where plots of residual error versus predictors showed evidence of nonlinearities.

				Seagrass model dependent variable				
				Grid cell P/A	Shore-line segment P/A[2]	Shore-line segment relative frequency[1]	Minimum depth of occurrence	Maximum depth of occurrence
Model type				GLMM	GLMM	GLMM	GLM, GAM	GLM, GAM
R package(s)				glmmPQL	glmmPQL	glmmPQL	glm, gam	glm, gam
Initial set of independent variables included in models prior to backward stepwise selection								
Variable	Definition	Units	Fixed or random effect					
Fshorlin	Shoreline code	-99, 1-18	R	x	x	x		
cSAL	Centered growing season salinity	PSU	F	x	x	x		
cTEMPER	Centered growing season average water temperature	Deg C	F	x	x	x		
fSEDn	Sediment type (n represents # classes after lumping)	1 to 13	F	x	x	x	x	
cWIND	Wind wave energy	Joules/m	F	x	x	x	x	
cPTTOC	Centered sediment percent total organic carbon	%	F	x	x	x		x
csecchiminMax	Centered Secchi Depth (max along transect)	Meters	F					x
cSDavggtrZ	Centered growing season average Secchi depth – water depth	Meters	F	x	x	x		
cSDmingtrZ	Centered growing season minimum Secchi depth – water depth	Meters	F	x	x	x		
fZgtrMXZ	Depth greater than wave mixing depth (0 = FALSE, 1 = TRUE)		F	x	x	x		
cZgtrMXZ	Centered water depth greater than wave mixing depth	Meters	F	x	x	x		
halflen0708gsaMax	Maximum growing season wave mixing depth	Meters	F				x	
fISOLATED	Isolated shoreline (0 = FALSE, 1=TRUE)		F	x	x	x		

[1] Average or optimum (minimum or maximum depending on variable) value for transect substituted for point values

				Seagrass model dependent variable				
				Grid cell P/A	Shore-line segment P/A[3]	Shore-line segment relative frequency[1]	Minimum depth of occurrence	Maximum depth of occurrence
Model type				GLMM	GLMM	GLMM	GLM, GAM	GLM, GAM
R package(s)				glmmPQL	glmmPQL	glmmPQL	glm, gam	glm, gam
Initial set of independent variables included in models prior to backward stepwise selection								
Variable	Definition	Units	Fixed or random effect					
cDistHdShor	Centered distance to hardened shoreline	Meters	F	x	x	x		
cDistMarina	Centered distance to nearest marina	Meters	F	x	x	x		
cUSRMARIkm2	Centered unsewered residences on high infiltration soils/catchment area	#/km^2	F	x	x	x		x
cCG046avkm2	Centered winter 2004-2006 Canada goose density	#/km^2	F	x	x	x		
fEGPA99	Historic (1999) eelgrass presence (0 = FALSE, 1 = TRUE)		F	x[4]	x	x		
cAREA	Centered area of 1999 eelgrass patch	Meters^2	F	x[1]	x	x		
cDistEG99	Centered distance to edge of nearest 1999 eelgrass patch	Meters	F	x[1]	x	x		
Interaction terms								
	cWIND x fsedn		F	x	x	x		
	halflen0708gsaMax x fSEDn		F				x	
	csecchiminMax x cpttocMax		F				x	
	csecchiminMax x cUSRMARIkm2Max		F				x	
	cpttocMax x cUSRMARIkm2Max		F				x	

[1] Average or optimum (minimum or maximum depending on variable) value for transect substituted for point values

[4] Models were run both with and without historic seagrass predictors

				Seagrass model dependent variable				
				Grid cell P/A	Shore-line segment P/A[5]	Shore-line segment relative frequency[1]	Minimum depth of occurrence	Maximum depth of occurrence
Model type				GLMM	GLMM	GLMM	GLM, GAM	GLM, GAM
R package(s)				glmmPQL	glmmPQL	glmmPQL	glm, gam	glm, gam
Initial set of independent variables included in models prior to backward stepwise selection								
Variable	Definition	Units	Fixed or random effect					
	csecchiavMax x cpttocMax x cUSRMARIkm2Max		F				x	
	cSAL x cSDavtomaxZ		F	x	x	x		
	cSAL x fZgtrMXZ		F	x	x	x		
	cSDavtomaxZ x fZgtrMXZ		F	x	x	x		
	cSAL x cptTOC		F	x	x	x		
	cSDavtomaxZ x cptTOC		F	x	x	x		
	fZgtrMXZ x cptTOC		F	x	x	x		
	cSAL x cSDavtomaxZ x fZgtrMXZ		F	x	x	x		
	cSAL x cSDavtomaxZ x cptTOC		F	x	x	x		
	cSAL x fZgtrMXZ x cptTOC		F	x	x	x		
	cSDavtomaxZ x fZgtrMXZ x cptTOC		F	x	x	x		
	cSAL x cSDavtomaxZ x fZgtrMXZ x cptTOC		F	x	x	x		

[1] Average or optimum (minimum or maximum depending on variable) value for transect substituted for point values

2.3.4 Model diagnostic tests

For both simple and mixed models, we used a series of diagnostic tests to check model assumptions. Methods and example codes are described in detail in Appendix C. We checked models for independence of predictor variables using correlation coefficients and variance inflation factors (VIFs) , for homogeneity of variance based on visual examination of residual versus predicted value plots, for linearity of response based on visual examination of conditional probability plots and of residuals plotted against predictors, and for spatial autocorrelation (SAC).

2.3.5 Development of coordinate framework to assess SAC

To assign coordinates to grid cell centroids, we first assigned a shoreline "event" measure for each observation. We edited the shoreline polyline to facilitate development of a shoreline "route" in ArcMap. Creation of a route in ArcMap allows one to assign distances to points along the line relative to the start of the shoreline route, termed "events". We eliminated areas of channel "braiding" in estuarine headwaters (and thus, divergent flowpaths), and "flipped" the direction of line segments in the polyline to ensure that all segments were oriented in a common direction. We visualized problem segments by changing the line symbol to an arrow and noting locations with adjacent arrowheads pointing in opposite directions. We "dissolved" the shoreline to minimize the number of features, then used the dissolved shoreline to create a shoreline route. We artificially created equidistant points along the shoreline route by generating an event table in Microsoft Excel, with MEAS values increasing in 10-meter increments and DISTANCE (from line) set to zero, then used the event table and shoreline route to create shoreline events in ArcMap. We saved the temporary shoreline events as permanent point features, then used the NEAR function in ArcMap to assign each observation to the nearest shoreline point with an associated shoreline distance (Figure 5). We provided variogram or correlation structure functions in R with the MEAS value as an "X-coordinate" and a constant of zero as the "Y-coordinate" to allow the program to calculate interpoint distances parallel to the shoreline.

2.3.6 Assessment of SAC ranges

Due to memory constraints with R packages and anisotropy we used two different R packages to evaluate SAC. Anisotropy exists when spatial autocorrelation varies by direction. Based on the elongated nature of seagrass patches and the different controls on spread along the shoreline as compared to along depth gradients, one would expect anisotropy in spatial autocorrelation of seagrass presence/absence. Vegetative spread of seagrasses predominates over long range seed dispersal as the mechanism for patch growth (Marba and Duarte 1998). Thus, seagrass is more likely to spread in a direction parallel to shorelines rather than perpendicular to shorelines over areas of deeper water.

To evaluate the range of spatial autocorrelation parallel to shorelines, we used spline correlograms based on shoreline distance. We evaluated SAC in a direction parallel to shorelines using nonparametric spline correlograms with the R spline.correlog function in the ncf package (Bjornstad 2013). Creation of nonparametric spline correlograms requires no assumptions about the distribution of errors or the shape of the correlogram plots. Thus, spline correlograms are more appropriate than standard correlograms for ecological

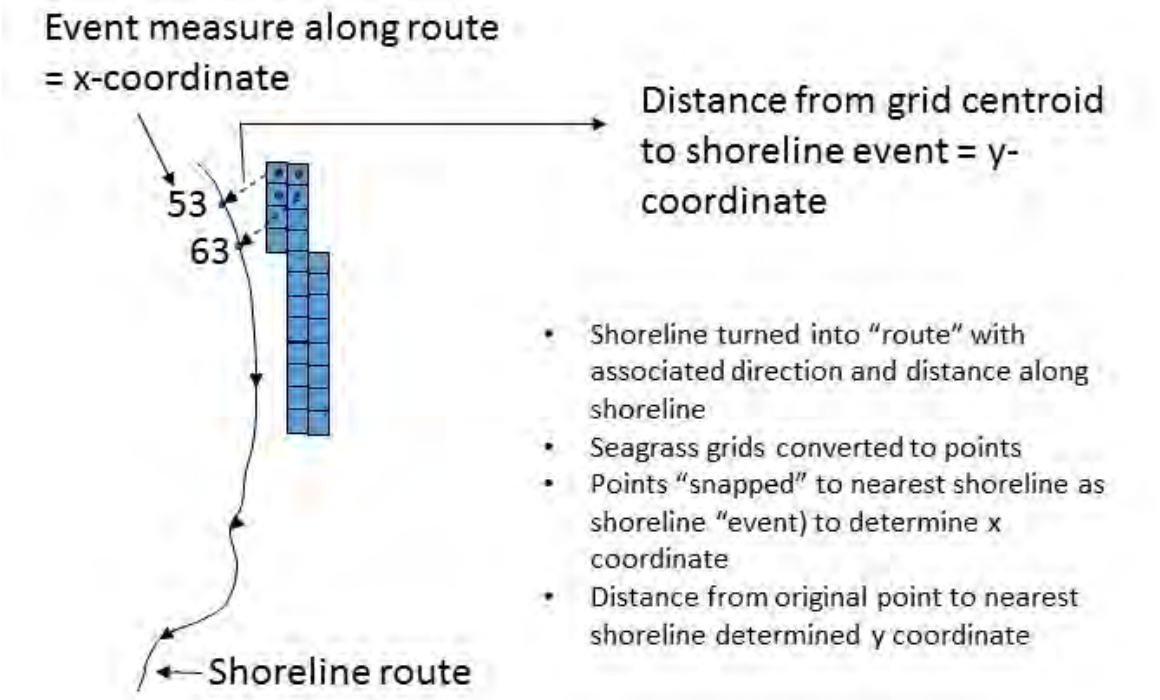

Figure 5. Process for assigning x-, and y-coordinates to seagrass grid cell centroids.

datasets which can exhibit a complex pattern of SAC at different scales resulting from different processes. We created coordinate inputs to the spline.correlog function by assigning shoreline distance as the x-coordinate and a fixed off-shore distance of zero as the y-coordinate. The spline function encounters memory limitations when more than ~9000 points are included in the analysis. Therefore, we had to develop the spline correlograms separately by shoreline, using a random subset of 9000 for each shoreline.

To evaluate the range of autocorrelation along depth gradients perpendicular to the shoreline we used the CommunityCorrelogram package in R to create Mantel correlograms (Andrus et al. 2014). The CommunityCorrelogram package adds functionality over existing Mantel correlogram functions by allowing directional (anisotropic) restrictions in both the xy (surface) plane and the z (depth) plane.

2.3.7 Eliminating or incorporating SAC into models

We evaluated two complementary methods to incorporate the effects of spatial autocorrelation into models. First, we incorporated local SAC into models using the approach of Crase et al. (2012). We initially fit models assuming spatial independence of errors. We then used zonal statistics in ArcMap to estimate a local average of original model residuals for each grid cell across the local range of SAC, using a moving window of 3 x 3 grid cells. We then incorporated the residual zonal averages into a modified predictive model to explicitly include the effect of SAC. We again checked adjusted model Pearson residuals for SAC patterns using the spline correlograms. The output includes confidence intervals for the spline fit and can be used to generate 95% confidence intervals for the first zero intercept, or the range at which spatial autocorrelation is no longer observed.

Attempts to use the Crase et al. approach in its original form failed because the range of SAC was much greater than the scale of the 3 x 3 moving window. Thus we reran the process using the empirically-derived range of SAC to adjust the size of the moving window for calculating residual zonal averages. We used the median value of SAC values across shorelines to calculate zonal averages, which also corresponded to the shoreline with the greatest abundance of seagrass.

Second, we eliminated some of the spatial autocorrelation in models by reducing dimensionality. We assigned the associated shoreline event measure (shoreline location) to each seagrass grid centroid, then calculated an average and maximum presence/absence value to estimate relative seagrass frequency (along a swath extending perpendicular to the shoreline) or simply seagrass presence/absence by 10-meter shoreline unit. For each shoreline event with seagrass present, we also calculated a minimum depth of occurrence and a maximum depth of occurrence. Predictive models for minimum and maximum depths were simpler because the random shoreline effect could be dropped, yielding a fixed effect model.

2.3.8 Strategies for eliminating multicollinearity

Development of GLM, GLMM and GAM models was constrained by the need to exclude predictor variables that exhibited collinearity (as evidenced by Spearman correlation coefficients of > 0.7) or multicollinearity (as evidenced by variance inflation factors (VIF) < 10; Zuur et al. 2009). Many of the variables reported in the literature as potential influences on seagrass growth and survival are potentially correlated with one another, making it difficult to determine precise model coefficients. We minimized the potential for collinearity by centering continuous variables (i.e., subtracting the mean) as per recommendations of Zuur et al. (2009). Centering variables also facilitates interpretation of coefficients in logistic models because the model intercept represents the predicted value when all independent variables are at their mean levels (and dummy variables equal zero). To prevent multi-collinearity, we also dropped sediment particle-size categories with no seagrass from the model, combined sediment particle-size categories representing a very small fraction of the model points into new classes, and adjusted the chosen "reference" level of the sediment type factor so that it represented one of the more common sediment types.

Effects of both transparency (Secchi depth) and wave energy (mixing depth) have to be evaluated with respect to water depth. The ratios of Secchi depth to water depth and mixing depth to water depth are highly correlated because they share a common denominator. Therefore we created one predictor based on the difference between Secchi depth and water depth and a second categorical predictor indicating whether or not mixing depth exceeded the actual water depth. The former variable had a more even distribution than the corresponding ratio and was less likely to create problems with outliers. Temperature and salinity were highly correlated, so we dropped temperature from our predictive models because it was not likely to reach critical levels in the subtidal zone of this system, and *Z. marina* does not occur in exposed locations in Narragansett Bay. We initially evaluated a subset of potential interaction terms in regression models based on probable mechanisms of action (Table 3). However, we eventually dropped almost all

potential interaction terms from the GLM, GLMM, and GAM models because they had extremely high VIF values.

2.3.9 Alternative approaches

We were limited in the range of R statistical packages available for mixed effects logistic models that could also incorporate spatial correlation structures (http://glmm.wikidot.com/pkg-comparison). The most commonly used functions for mixed modeling in R are: MASS::glmmPQL, lme4::glmer, MCMCglmm::MCMCglmm. Of the available R packages, glmmPQL is less memory-intensive as it applies a Penalized Quasi-Likelihood approach rather than producing true Maximum Likelihood Estimates (Venables and Ripley 2002). As a result, glmmPQL does not yield an estimate of log-likelihood or Aikake's Information Criteria (AIC), values commonly used in comparing model fits. Although we attempted to fit models using the glmer function in lme4, models often failed to converge, yielded "out of memory" errors before convergence, or proved impractical because model runs required more than 12 hours for a single iteration. Attempts to run the glmmADMB package in R also failed.

In summary, our approach yielded predictive models for five seagrass endpoints: grid-cell presence/absence, shoreline presence/absence, shoreline relative frequency, minimum depth of occurrence, and maximum depth of occurrence. We evaluated glmmPQL or GAMM models for the first three endpoints, and GLM or GAM models for the latter two. We incorporated SAC into predictive models using the Crase et al. (2012) approach after modifying the moving window to account for the actual range of SAC observed in model residuals. Finally, we minimized the potential for multi-collinearity across independent variables. See Appendix C, Figure C1, for a summary of the GIS and R processes necessary to develop the GLMMs.

Chapter 3. Results

For more detailed results, see Appendix D for exploratory analyses and description of intermediate models.

3.1 Final Models

3.1.1 Seagrass Grid Presence/Absence

The final model fit for seagrass grid cell occupancy was:

$$\text{fsavcode} = \text{cSAL} + \text{cSAL}^2 + \text{cSDavgrtrZ} + \text{cSDavgrtrZ}^2 + \text{cSDavgrtrZ}^3 + \text{cptTOC} + \text{cptTOC}^2 + \text{cptTOC}^3 + \text{fZgtMXZav} + \text{cCG046avkm} + \text{cDstoMarin} + \text{fISOLATED} + \text{fSED4} + \text{PResid14fa}, \quad (1)$$

where fsavcode = seagrass presence/absence code (1/0)

- cSAL = centered salinity
- cSDavgrtrZ = centered average Secchi depth minus water depth
- cptTOC = centered sediment percent total organic carbon
- fZgtMXZav = indicator of depth greater than average wave mixing depth (1/0)
- cCG046avkm2 = centered Canada goose density/square kilometer
- cDstoMarin = centered distance to nearest marina
- fISOLATED = indicator of isolated shoreline (1/0)
- fSED4 = sediment particle-size class
  - SED4_5 = Sand
  - SED4_7 = Silty sand
  - SED4_81011 = Silty, Sandy silt, Clay-Silt
  - SED4_6 = Gravelly sand
  - SED4_124 = Gravel, Sandy gravel, Gravel-sand-silt
- PResid14fa = Focal average of Pearson residual over zone of 1300 m shoreline length x 200 m offshore distance

We calculated the residual autocorrelation term as the focal average of Pearson residuals over a zone of 1300 meters (along shoreline direction) by 200 meters (offshore direction). After incorporation of the residual autocorrelation term, the final model showed minimal spatial autocorrelation (Figure 6a) and much reduced heterogeneity of variance (Figure 6b). Area under the ROC curve based on ten-fold cross-validation was 0.7144 (Figure 7). The model showed a small tendency to underpredict seagrass presence, with a mean

residual of -0.1285, and a median value of -0.001545 (interquartile range = -0.0739 to -0.000043). The root mean squared error was 0.34.

Colonization rates vary dramatically among shorelines (Table 4), with exponentiated coefficients (the associated odds ratio) ranging over nine orders of magnitude from 6.1 E-5 (Shoreline 2) to 93091 (Shoreline 17). The exponentiated intercept for the fixed effects portion of the mixed model is the odds ratio when all continuous variables are at mean values (centered value = 0), fZgtMXZav = 0 (depth is less than average wave mixing depth), fISOLATED = 0 (main shorelines), and the reference sediment class (5 = sand) is 1 (true) (Table 5). The corresponding probability is 0.22. The odds ratio for average conditions of seagrass on sand substrate for isolated shorelines is 0.279 + 0.004 = 0.283. The odds ratio increases by 1.43 (43%) for water depths greater than the wave mixing depth. The odds ratio for sediment types 8 and 10 combined (silty and sandy silt) is not significantly different than for the sand class, but is relatively lower for classes 6 (gravelly sand) or 12 (sand-silt-clay) and higher for classes 7 (silty sand) or 1+2+4 (gravel, sandy gravel, and gravel-sand-silt), respectively. The conditional odds ratio is 6.65 (or 565% greater) for the combined gravel classes. Predicted effects for unsewered residential density on high infiltration soils, Canada goose density, and distance to nearest marina are all opposite in direction to those expected but the magnitudes of predicted influences are negligible. The probability of seagrass presence increases at an accelerated rate as salinity increases.

Table 4. Random effects associated with shorelines for GLMM model 1 predicting seagrass presence/absence by grid cell. See Figure 3 for map of shoreline codes.

Shoreline	Random effect	Exp(R.E.)
-99	2.78	16.2
2	-9.70	6.1E-05
3	-5.67	0.003
6	-9.62	6.66E-05
7	-6.22	0.002
8	-4.14	0.016
9	0.42	1.5
10	3.00	20
11	-2.64	0.07
12	1.46	4.3
13	-4.33	0.01
14	0.28	1.3
15	6.65	771
16	9.70	16398
17	11.44	93091

The odds ratio for seagrass presence is negligible for water depths less than the Secchi depth, increases above 1.0 when Secchi depth exceeds water depth by one meter, and is predicted to peak at Secchi depths about 2.5 meters greater than water depth (Figure 8; note that these odds ratios must be adjusted for the shoreline of interest).

Table 5. Fixed effects for model 1 predicting seagrass presence/absence by grid cell.

	Coeff	Std. Error	DF	Exp(Coeff)	t-value	p-value
(Intercept)	0.42	1.66	518856	1.52	0.25	0.8001
cSDavgrtrZ	2.61	0.05	518856	13.63	48.69	<0.0001
$cSDavgrtrZ^2$	-0.23	0.03	518856	0.79	-7.11	<0.0001
$cSDavgrtrZ^3$	-0.06	0.010	518856	0.95	-5.86	<0.0001
cSAL	1.43	0.19	518856	4.20	7.60	<0.0001
$cSAL^2$	-0.57	0.16	518856	0.57	-3.65	0.0003
cptTOC	-2.44	0.09	518856	0.09	-28.45	<0.0001
$cptTOC^2$	1.32	0.07	518856	3.73	18.60	<0.0001
$cptTOC^3$	-0.32	0.05	518856	0.73	-5.78	<0.0001
cCG046avkm	0.01	0.004	518856	1.01	2.93	0.0034
cDstoMarin	-0.001	0.0001	518856	1.00	-27.87	<0.0001
fSED4124	3.66	0.33	518856	38.77	10.93	<0.0001
fSED47	0.63	0.07	518856	1.88	8.76	<0.0001
fSED4810	-0.38	0.15	518856	0.69	-2.51	0.0119
fSED412	-0.83	0.15	518856	0.43	-5.50	<0.0001
fSED46	-1.12	0.21	518856	0.33	-5.25	<0.0001
fZgtMXZav	0.53	0.14	518856	1.71	3.74	0.0002
fISOLATED1	-10.67	1.06	518856	0.00	-10.06	<0.0001
PResid14fa	3.17	0.06	518856	23.88	56.60	<0.0001

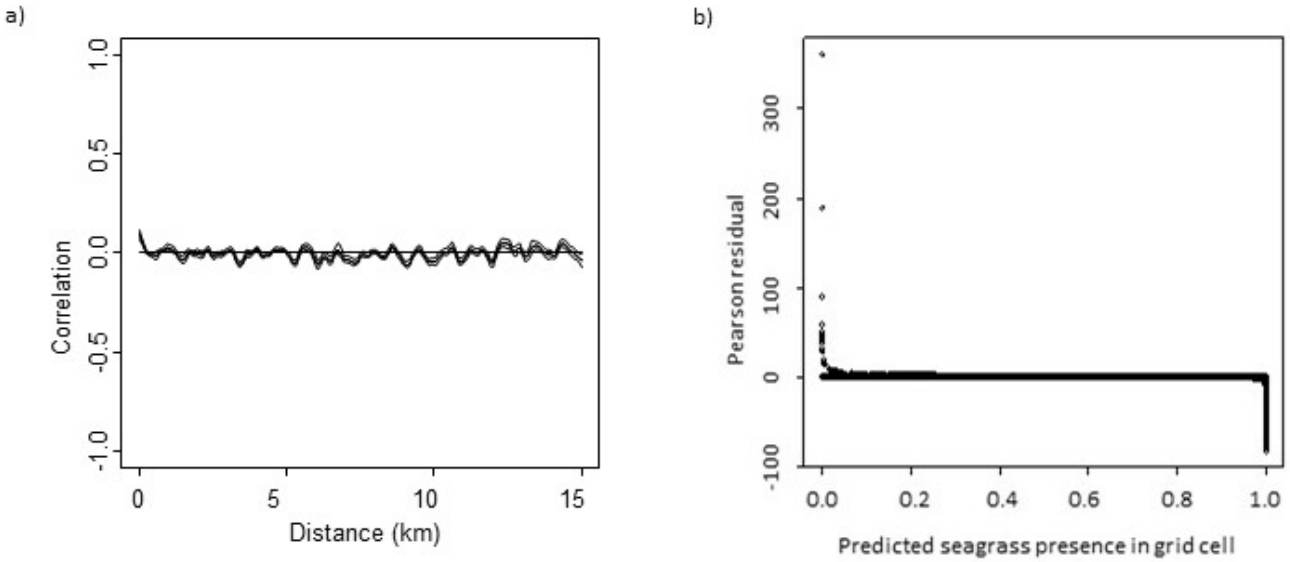

Figure 6. Diagnostic plots for model 1 predicting seagrass grid occupancy. Spatial autocorrelation of residuals is virtually eliminated. Heterogeneity of variance is greatly reduced. a) Spline correlogram of Pearson residuals for Shoreline 14. b) Plot of Pearson residuals versus predicted value.

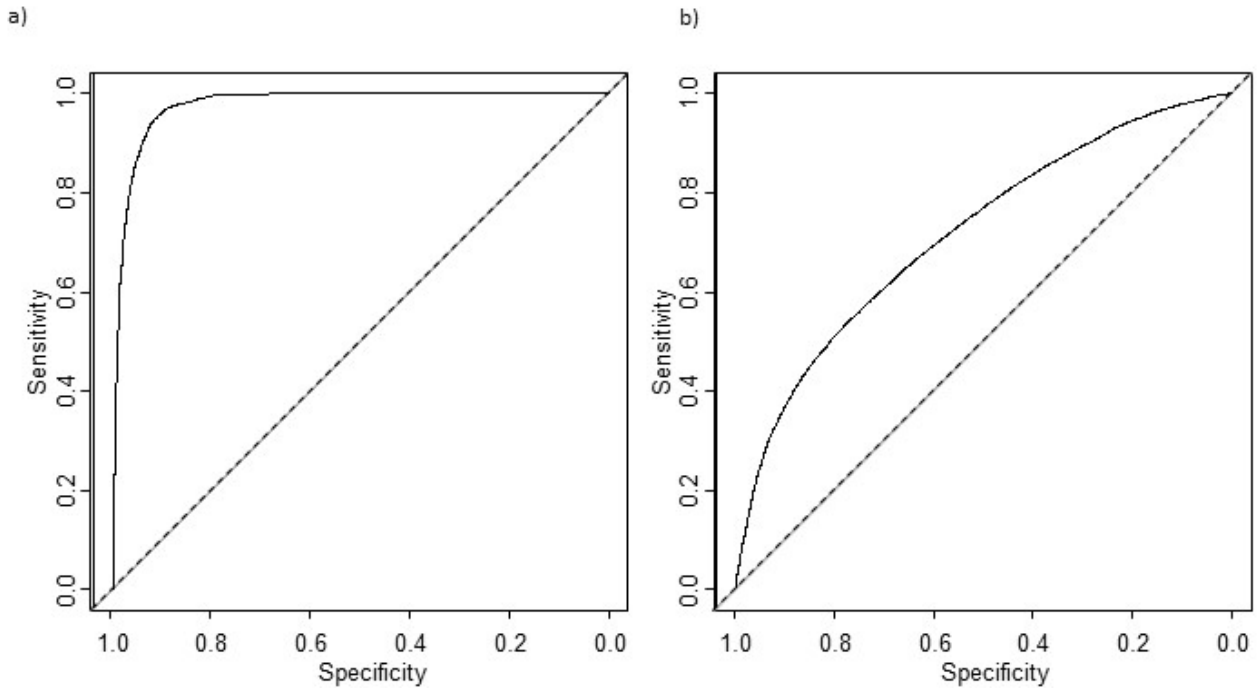

Figure 7. ROC curve for model 1 based on a) initial fit (area under curve = 0.9767) and b) 10-fold cross-validation (area under the curve is 0.7144).

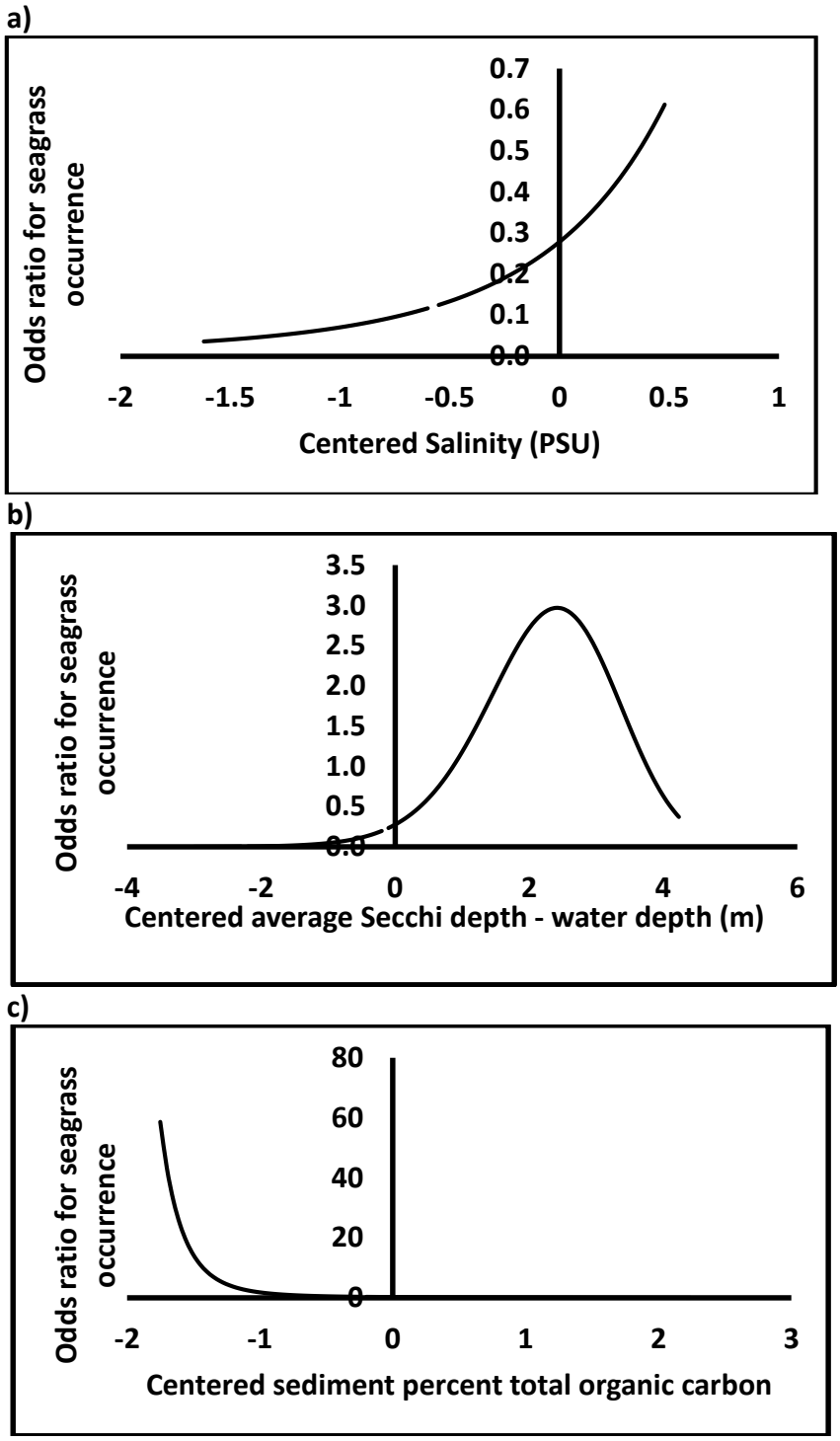

Figure 8. Nonlinear effects of a) centered Salinity (PSU), b) centered average Secchi depth – water depth (m), and c) centered sediment percent total organic carbon on odds ratio for seagrass occurrence in grid cell. Values are calculated assuming average values for all co-variables not represented in plot and across all shorelines with the reference sediment type for model 1.

3.1.2 Predictive models for shoreline segment P/A and minimum or maximum depth for seagrass

3.1.2.1 Shoreline segment P/A models without Serial Autocorrelation

The form of the final model for predicting seagrass presence/absence by shoreline distance was:

$$fsavcodeMax_i = csalAv_i + csecchimin_i * cpttocMin + fZgtMXZavM_i + cCG046avkm_i + cDistToHdS_i + cDstoMarin_i + cwindMin_i * fSED8_i + PR4wFA1370_i, \quad (2)$$

where i = shoreline event measure

$fsavcodeMax_i$ = Maximum seagrass presence/absence code (0/1) associated with shoreline event measure i

$csalAv_i$ = Average of centered average salinity at i

$csecchimin_i$ = Minimum of centered Secchi depth at i

$cpttocMin_i$ = Minimum of centered percent total organic carbon at i

$fZgtMXZavM_i$ = Maximum of 0/1 indicator of depths greater than average wave mixing depth at i

$cCG046avkm2_i$ = Minimum of centered Canada goose density (No./km2) at i

$cDistToHdS_i$ = Maximum of centered distance to hardened shoreline at i

$cDstoMarin_i$ = Maximum of centered distance to nearest marina at i

$cwindMin_i$ = Minimum of centered wind relative energy at i

$fSED8_i$ = Majority sediment class at i, including

fSED8_6 = Gravelly-sand

fSED8_7 = Silty sand

fSED8_124 = Gravel, Sandy Gravel, and Gravel-Sand-Silt

fSED8_512 = Sand and Sand-Silt-Clay

fSED8_8101113 = Silty, Sandy Silt, Clay-Silt, and Gravel-Silt-Clay

$PR4wFA1370_i$ = Focal average over 1370 meters distance of Pearson residual for model fit

The dummy variable for "isolated" shorelines, corresponding to small islands, was no longer significant after spatial autocorrelation was accounted for, and some of the sediment classes had to be collapsed from the original model to retain significance. After residual autocorrelation was incorporated into the model, heterogeneity of variance decreased

substantially (Figure 9a-c). The original model Area-Under-The-Curve (AUC) value for the Receiver Operating Characteristic Curve (ROC) was 0.95. Performance of the predictive model for seagrass shoreline presence was much more robust than performance of the model predicting seagrass presence at the grid cell scale. Ten-fold cross-validation of the model yielded a mean residual of -0.111, a Root Mean Squared Error (RMSE) of 0.29 (interquartile range = 0.014 – 0.155), and an ROC value of 0.9547 (see Figure 10). Again, the odds ratio varied by several orders of magnitude across different shorelines (Table 6). Higher order effects were not retained in the best GLMM model predicting shoreline presence. In this case, the odds ratio was greatest for the reference sediment class (6 = gravelly sand) and least for sediment classes 1 + 2 + 4 (gravel-dominated). There was an additional interaction term involving Secchi depth and sediment percent total organic carbon, which tends to decrease the positive effects of transparency at high sediment organic carbon levels (Table 7).

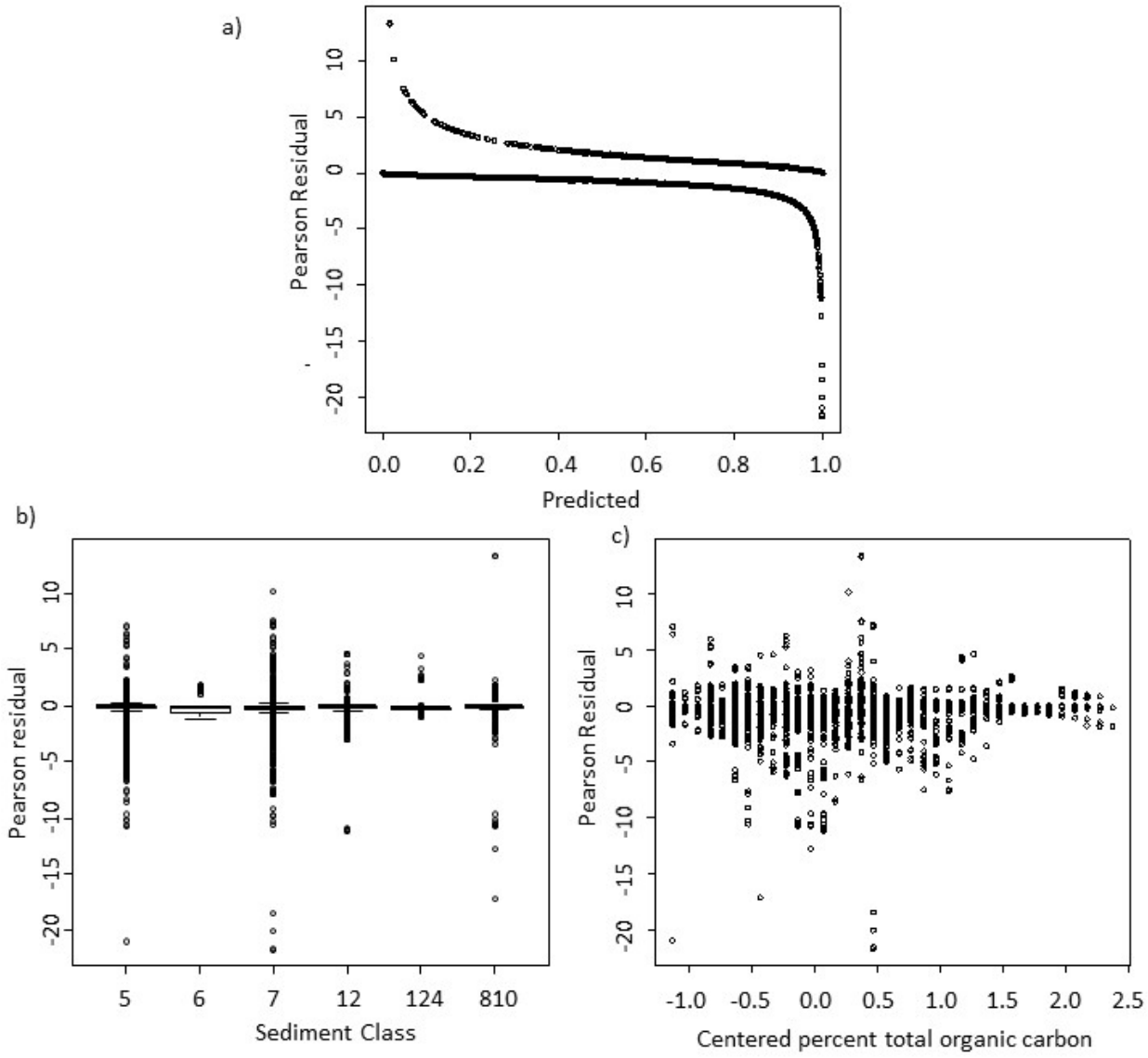

Figure 9. Diagnostic plots for model 2 predicting seagrass presence/absence by shoreline segment. a) Pearson residual versus predicted value, b) Pearson residual by sediment class, and c) Pearson residual versus centered percent total organic carbon.

Table 6. Random effects for GLMM model 2 predicting shoreline presence/absence for seagrass, not accounting for historic seagrass presence.

Shoreline	(Intercept)	exp(coeff)
-99	-4.30	0.01
10	6.55	698.81
12	4.91	136.15
17	3.34	28.33
9	2.78	16.14
14	2.47	11.78
13	1.58	4.85
16	1.37	3.94
18	0.67	1.95
15	0.21	1.24
8	-0.47	0.63
3	-2.11	0.12
2	-3.42	0.03
6	-3.47	0.03
7	-4.32	0.01
11	-5.79	0.00

Table 7. Fixed effects for GLMM model 2 predicting shoreline presence/absence for seagrass, not accounting for historic seagrass presence.

Parameter	Coefficient	Std.Error	exp(Coeff)	DF	t-value	p-value
(Intercept)	-1.16	0.99	0.3	19174	-1.17	0.2431
csalAv	1.57	0.09	4.8	19174	17.49	0
fZgtMXZav	1.55	0.32	4.7	19174	4.79	0
csecchimin	0.31	0.08	1.4	19174	3.68	0.0002
cCG046avkm	0.033	0.004	1.0	19174	8.80	0
cDistToHdS	-6.55E-04	7.56E-05	1.0	19174	-8.66	0
cDstoMarin	-2.55E-04	5.10E-05	1.0	19174	-4.99	0
cwindMin	-1.50E-05	2.70E-06	1.0	19174	-5.57	0
cpttocMin	-1.06	0.06	0.3	19174	-17.47	0
fSED87	-0.56	0.21	0.6	19174	-2.70	0.0069
fSED8124	-3.55	0.48	0.03	19174	-7.47	0
fSED8512	-0.79	0.22	0.5	19174	-3.65	0.0003
fSED88101113	-0.78	0.23	0.5	19174	-3.48	0.0005
PR9wFA1370	3.62	0.06	37.5	19174	57.62	0
Csecchimin x cpttocMin	0.27	0.10	1.3	19174	2.78	0.0054

3.1.2.2 Shoreline P/A models with Serial Autocorrelation

We created an alternative model to predict shoreline presence/absence using historic seagrass presence/absence as a predictor. The final model form was (Table 8):

$$\begin{aligned} fsavcodeMax_i = {} & fEGPA99Max_i + cAREAMax + cDistToEG99Min_i + csalAv_i + \\ & csecchiminMax_i * cpttocMin_i + fZgtMXZavMax_i + cCG046avkm2Min_i + \\ & cDstoMarinaMax_i + cwindMin_i * fSED6_I + PR6dwFA700, \end{aligned} \quad (3)$$

where terms are defined as above (Table 3) with the addition of:

$fSED6_i$ = Majority sediment class at i, including
- fSED6_6 = Gravelly-sand
- fSED6_124571213 = Gravel, Sandy Gravel, Gravel-Sand-Silt, Sand, Sand-Silt-Clay, and Silty-Sand
- fSED6_81011 = Silty, Sandy Silt, Clay-Silt, and Gravel-Silt-Clay

$PR6dwFA700_i$ = Focal average over 700 meters distance of Pearson residual for model fit

Table 8. Fixed effects model 3 components predicting shoreline presence and including historic 1999 seagrass predictors.

Parameter	Value	exp(coeff)	Std.Error	DF	t-value	p-value
(Intercept)	-2.09	0.12	1.09	19172	-1.9	0.0556
csalAv	0.83	2.28	0.13	19172	6.2	0
csecchiminMax	0.09	1.09	0.14	19172	0.6	0.5475
cCG046avkm2Min	-0.04	0.96	0.007092	19172	-5.9	0
cAREAMax	0.000006	1.00	2.4E-06	19172	2.6	0.0092
cDistToEG99Min	-0.00199	1.00	0.00011	19172	-18.1	0
cDstoMarinaMax	-0.00024	1.00	0.000088	19172	-2.8	0.0057
cwindMin	0.00015	1.00	3.53E-05	19172	4.3	0
cpttocMin	-0.43	0.65	0.11	19172	-3.9	0.0001
fEGPA99Max	1.84	6.32	0.37	19172	5.1	0
fZgtMXZavMax	1.49	4.44	0.50	19172	3.0	0.0028
fSED6124571213	-1.09	0.34	0.50	19172	-2.2	0.029
fSED681011	-1.93	0.15	0.54	19172	-3.6	0.0003
PR6dwFA700	3.57	35.39	0.10	19172	35.4	0
csecchiminMax x cpttocMin	0.75	2.12	0.18	19172	4.1	0
cwindMin x fSED681011	-0.0002	1.00	4.83E-05	19172	-4.2	0
cwindMin x fSED6124571213	-0.00016	1.00	3.55E-05	19172	-4.6	0

Model predictions based on the original fit were slightly better than the prior model which did not include historic eelgrass presence/absence, with a ROC value of 0.9753.

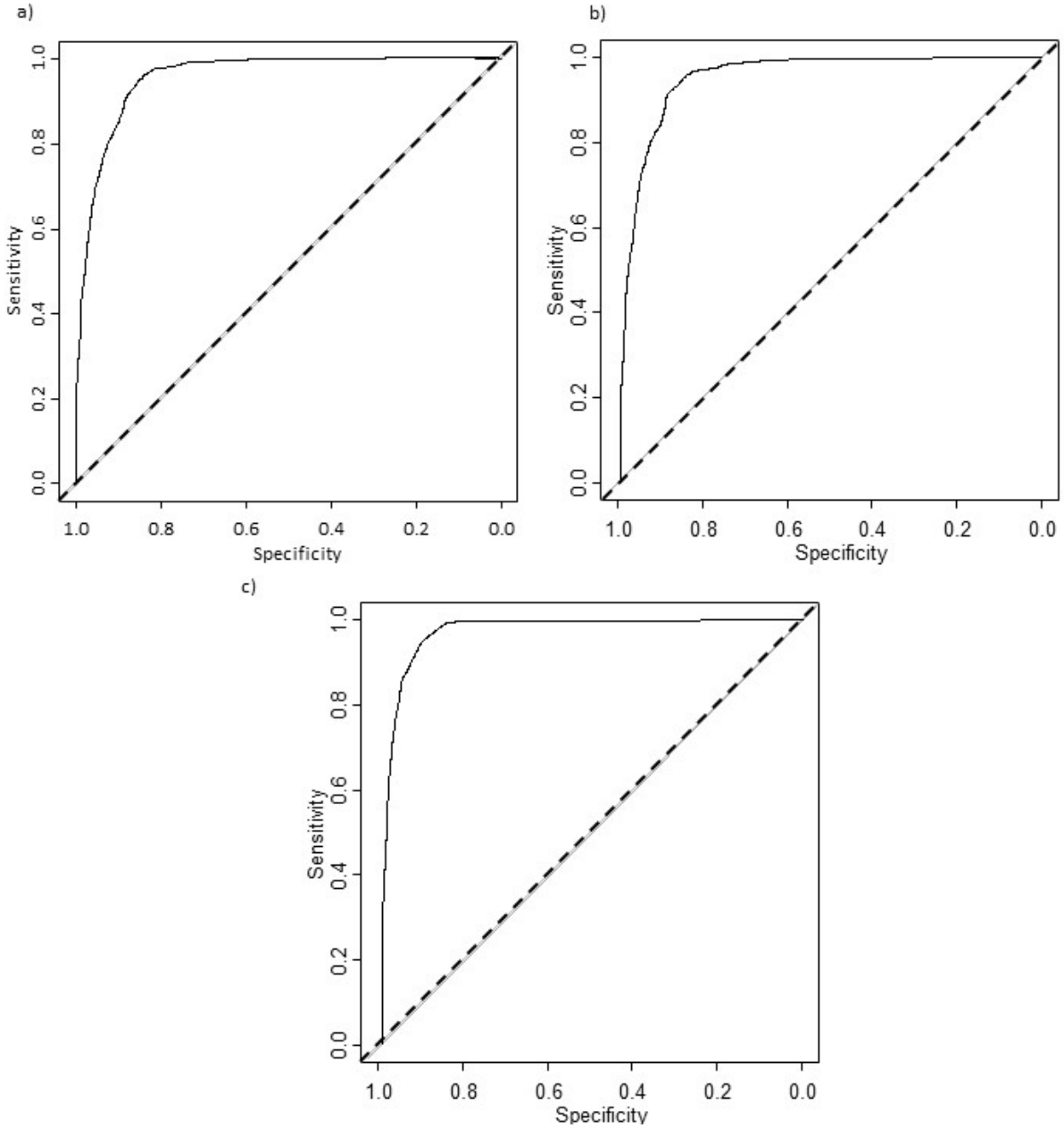

Figure 10. Receiver operating characteristic (ROC) curve for final model predicting shoreline presence/absence of seagrass based on a) original model 2 fit with full data set (ROC = 0.9546), b) same model with ten-fold model cross-validation (ROC = 0.9547), and c) original model 3 fit for model with full data set incorporating historic 1999 eelgrass presence/absence (ROC = 0.9753).

3.1.2.3 Shoreline Segment Relative Frequency Models without Serial Correlation

The final model predicting shoreline relative frequency (average shoreline seagrass occupancy along offshore transect) after incorporating residual autocorrelation was:

savcodeAv = csalAv + csalAv^2 + csecchiminAv + csecchiminAv^2 + cpttocAv + cpttocAv^2 + cZgtMXZavAv + cZgtMXZavAv^2 + cDistToHdShAv + cDstoMarinaAv + cUSRMARIkm2Av + fISOLATED + cwindAv * fSED5 + PR3xFA460, (4)

where terms are defined as above with the addition of

PR3xFA460 = focal average of Pearson residuals with zone length of 460 meters

After residual autocorrelation was incorporated into the model, the magnitude of the random effect term was very small compared to residual error, so the random shoreline effects were dropped and a simpler general linear model was run. Similar to the grid-cell occupancy model, several higher order effects were retained, but this time an additional second-order term for depth greater than average wave mixing depth was included. In this case the direction of effects for distance to nearest marina (positive influence) and unsewered residential density (negative influence) was as expected, but the magnitude of these effects was relatively small (Table 9). Diagnostics plots showed improvements over fits without the residual SAC term (Figures 11a-d). However, 3 outliers were apparent. The model was fit after removing the three outliers and the same terms were retained. Figure 12 illustrates the interaction between sediment class and wind-derived wave energy, with wave energy effects greatest for silt-dominated classes, as compared to sand or gravel-dominated classes. Performance of the model predicting average shoreline occupancy by seagrass was slightly lower than predictions of presence/absence, but still much more robust than model predictions at the grid cell scale. The area under the ROC curve was 0.889 for model 4, and slightly less (0.8377) for the model 4 fit without 3 outliers (Figure 13).

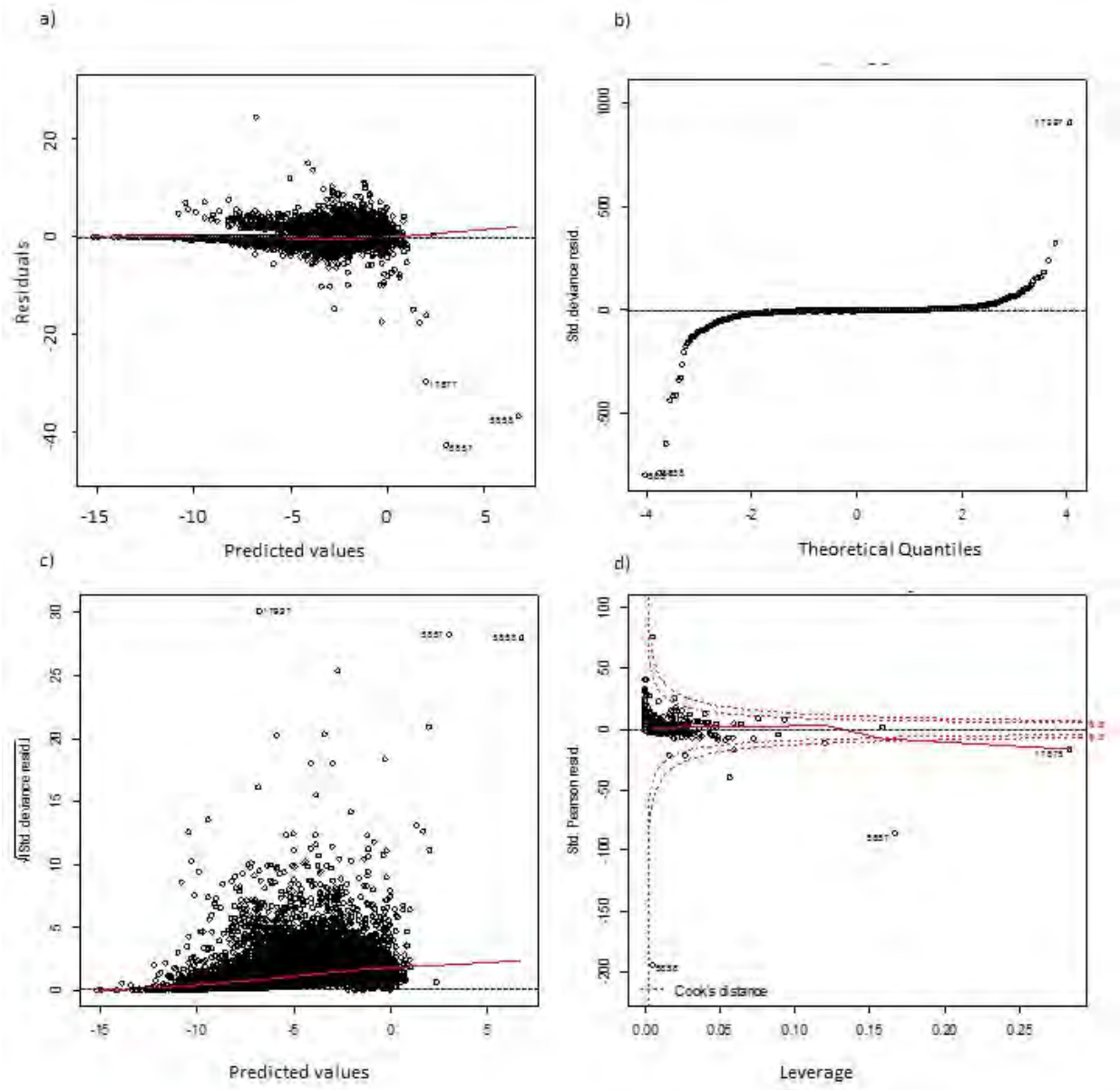

Figure 11. Diagnostic plots for model 4 predicting average shoreline occupancy. a) Pearson residuals vs predicted values, with loess curve superimposed showing no trend, b) Q-Q plot (standardized deviance residuals versus theoretical quantiles), c) Scale-Location plot (square root of standardized deviance residuals versus predicted values), and d) Residuals vs Leverage with three labeled outliers.

Table 9. Fixed effects in model 4 predicting average seagrass occurrence along transects perpendicular to shoreline, not incorporating historic seagrass presence.

	Model with full data set						Model without 3 outliers					
Coefficients:	Estimate	Std. Error	z value		Pr(>|z|)		Estimate	Std. Error	z value	Pr(>|z|)		
(Intercept)	-3.52	0.03	-103.6	<	2.00E-16	***	-3.30	0.03	-107.3	<2e-16	***	
csalAv	1.71	0.07	24.5	<	2.00E-16	***	1.45	0.06	22.6	<2e-16	***	
csalAv^2	0.65	0.05	12.0	<	2.00E-16	***	0.72	0.05	14.9	<2e-16	***	
csecchiminAv	-1.20	0.05	-23.9	<	2.00E-16	***	-1.04	0.05	-22.7	<2e-16	***	
csecchiminAv^2	1.07	0.03	39.8	<	2.00E-16	***	1.00	0.02	40.0	<2e-16	***	
cDstoMarinaAv	3.89E-04	1.56E-05	25.0	<	2.00E-16	***	3.38E-04	1.45E-05	23.3	<2e-16	***	
cwindAv	-6.53E-06	1.11E-06	-5.9		3.51E-09	***	-2.65E-06	1.22E-06	-2.2	0.0303	*	
cDistToHdShAv	-4.83E-04	3.11E-05	-15.5	<	2.00E-16	***	-4.88E-04	2.73E-05	-17.9	<2e-16	***	
cUSRMARIkm2Av	-0.013	0.000	-35.5	<	2.00E-16	***	-0.011	0.000	-30.5	<2e-16	***	
cpttocAv	-0.14	0.02	-6.8		8.62E-12	***	-0.43	0.02	-24.6	<2e-16	***	
cpttocAv^2	-0.58	0.03	-23.0	<	2.00E-16	***	-0.17	0.02	-9.0	<2e-16	***	
cZgtMXZavAv	-0.73	0.01	-51.4	<	2.00E-16	***	-0.57	0.01	-47.3	<2e-16	***	
cZgtMXZavAv^2	-0.70	0.01	-51.1	<	2.00E-16	***	-0.53	0.01	-47.6	<2e-16	***	
fISOLATED	1.46	0.04	37.0	<	2.00E-16	***	1.28	0.04	34.7	<2e-16	***	
fSED581011	0.30	0.05	6.3		2.35E-10	***	-0.007	0.046	-0.2	0.8738		
fSED5124561213	-0.27	0.02	-11.4	<	2.00E-16	***	-0.47	0.02	-21.4	<2e-16	***	
PR3xFA460	3.444E-01	2.954E-03	116.6	<	2.00E-16	***						
cwindAv x fSED5124561213	-2.81E-05	1.97E-06	-14.3	<	2.00E-16	***	-3.63E-05	1.87E-06	-19.4	<2e-16	***	
cwindAv x fSED581011	-7.40E-05	5.02E-06	-14.7	<	2.00E-16	***	-9.11E-05	5.27E-06	-17.3	<2e-16	***	

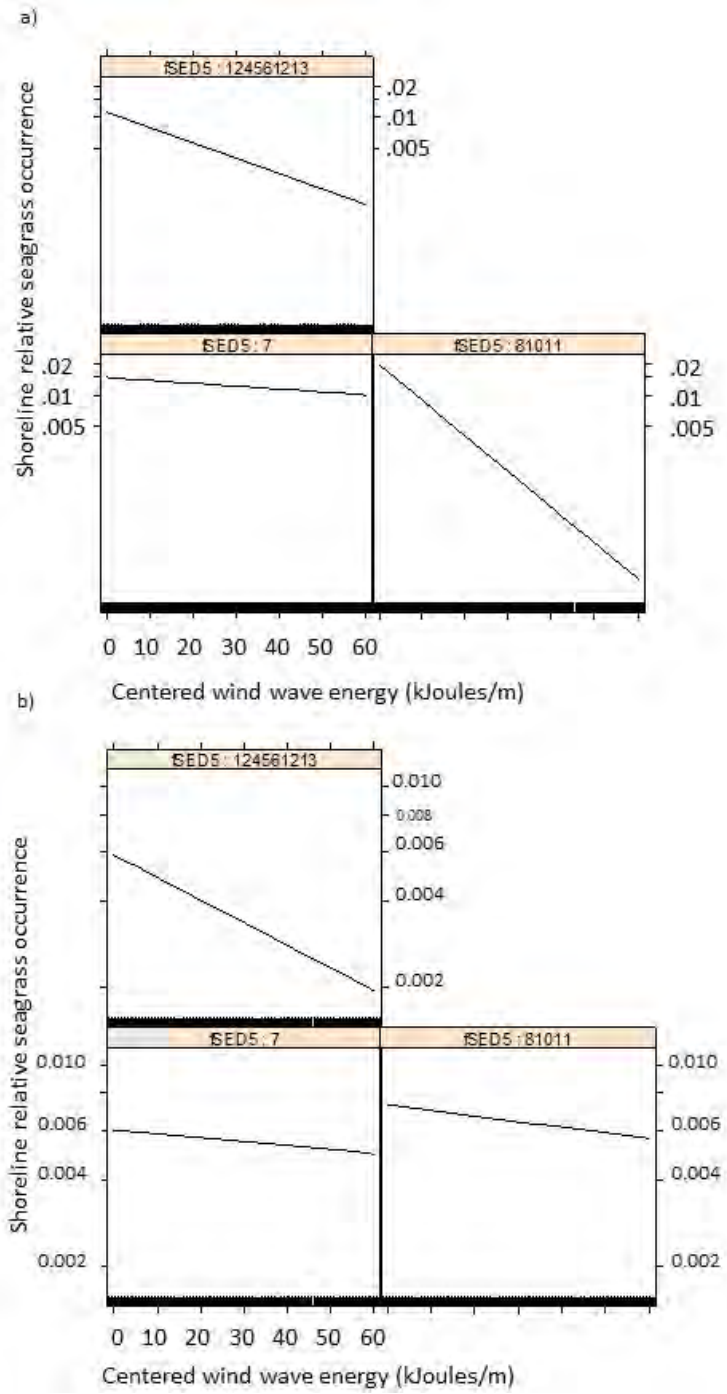

Figure 12. Interaction plots showing effect of shoreline wind-derived wave energy by sediment class for models 4 and 5 predicting average shoreline occupancy a) without (4) and b) with historic seagrass presence (5) as predictors. 124561213 = Gravel, Sandy gravel, Gravel-sand-silt, Sand, Gravelly sand, Sand-silt-clay, and Gravel-silt-clay, 7 = Silty sand, and 81011 = Silty, Sandy silt, and Clay silt.

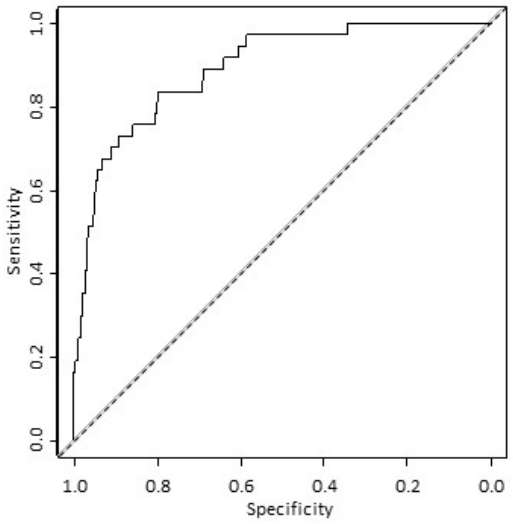

Figure 13. ROC curve showing fit of model 4 prediction of average shoreline occupancy after removal of three outliers

3.1.2.4 Shoreline Segment Relative Frequency Models with Serial Correlation

With the effect of historic 1999 occupancy incorporated, the final model predicting average shoreline occupancy was:

savcodeAv ~ cEGPA99Av + cAREAAv + cDistToEG99Av + csalAv + csalAv^2 + csecchiminAv + csecchiminAv^2 + cpttocAv + cZgtMXZavAv + cZgtMXZavAv^2 + cDistToHdShAv + cDstoMarinaAv + cUSRMARIkm2Av + cwindAv * fSED5 + PR4gFA460, (5)

where terms are defined as above with the addition of

cEGPA99Av = centered average shoreline occupancy in 1999

cAREAAv = centered average 1999 eelgrass patch size by shoreline index i

cDistToEG99Av = centered shoreline average distance to edge of nearest 1999 eelgrass patch, and

PR4gFA460 = zonal average over 460 meters of Pearson residual for model 4h

Relative seagrass presence increased in areas of historic 1999 seagrass, more so as historic patch size increased and less so with distance from historic patch edge (Table 10). As above, diagnostic plots showed the presence of three outliers (Figure 14), but re-analysis of model 4h without these outliers yielded the same set of predictors. Performance and

Table 10. Fixed effects for model 5 predicting average seagrass presence/absence along transects perpendicular to shoreline, including effects of historic seagrass presence. Model coefficients are compared for model fit with full data set and with data set minus three outliers.

	Full data set						Data set minus three outliers					
Coefficients	Estimate	Std. Error	z value		Pr(>|z|)		Estimate	Std. Error	z value		Pr(>|z|)	
(Intercept)	-5.10	0.05	-111	<	2.00E-16	***	-6.70	0.06	-112	<	2.00E-16	***
csalAv	2.48	0.08	31.9	<	2.00E-16	***	3.47	0.08	41.7	<	2.00E-16	***
csalAv^2	1.34	0.06	23.0	<	2.00E-16	***	2.04	0.06	32.4	<	2.00E-16	***
cEGPA99Av	1.69	0.16	10.8	<	2.00E-16	***	0.98	0.16	6.2		5.03E-10	***
cpttocAv	0.16	0.02	7.8		6.31E-15	***	0.16	0.02	7.3		2.75E-13	***
cAREAAv	7.32E-06	6.68E-07	11.0	<	2.00E-16	***	1.02E-05	6.68E-07	15.2	<	2.00E-16	***
cwindAv	-3.18E-06	9.98E-07	-3.2		0.00144	**	-5.74E-06	1.01E-06	-5.7		1.34E-08	***
cDstoMarinaAv	-1.79E-04	1.76E-05	-10.2	<	2.00E-16	***	-2.71E-04	1.82E-05	-14.9	<	2.00E-16	***
cDistToHdShAv	-3.50E-04	3.36E-05	-10.4	<	2.00E-16	***	-2.67E-04	3.62E-05	-7.4		1.55E-13	***
cUSRMARIkm2Av	-8.40E-03	2.89E-04	-29.0	<	2.00E-16	***	-8.00E-03	2.91E-04	-27.5	<	2.00E-16	***
cDistToEG99Av	-1.03E-03	1.42E-05	-72.8	<	2.00E-16	***	-1.52E-03	1.86E-05	-81.7	<	2.00E-16	***
cZgtMXZavAv	-0.54	0.01	-37.1	<	2.00E-16	***	-0.47	0.01	-32.9	<	2.00E-16	***
cZgtMXZavAv^2	-0.56	0.01	-40.9	<	2.00E-16	***	-0.34	0.01	-27.7	<	2.00E-16	***
csecchiminAv	-1.38	0.05	-28.6	<	2.00E-16	***	-1.60	0.05	-31.2	<	2.00E-16	***
csecchiminAv^2	0.90	0.03	34.3	<	2.00E-16	***	0.95	0.03	34.8	<	2.00E-16	***
fSED581011	0.20	0.05	4.1		4.93E-05	***	0.29	0.05	5.8		6.23E-09	***
fSED5124561213	-0.03	0.03	-1.1		0.27671		0.23	0.03	7.8		4.56E-15	***
PR4gFA460	0.457	0.004	114.2	<	2.00E-16	***	0.61	0.01	117.9	<	2.00E-16	***
cwindAv x fSED581011	-1.29E-06	5.58E-06	-0.2		0.81674		1.66E-05	5.28E-06	3.1		0.00167	**
cwindAv x fSED5124561213	-1.50E-05	2.20E-06	-6.8		8.90E-12	***	-2.64E-05	2.53E-06	-10.4	<	2.00E-16	***

robustness of the model were actually degraded by including historic seagrass presence as a predictor. The area under the ROC curve was 0.615 for model 4h but only 0.4887 for the same model with three outliers removed (Figure 15).

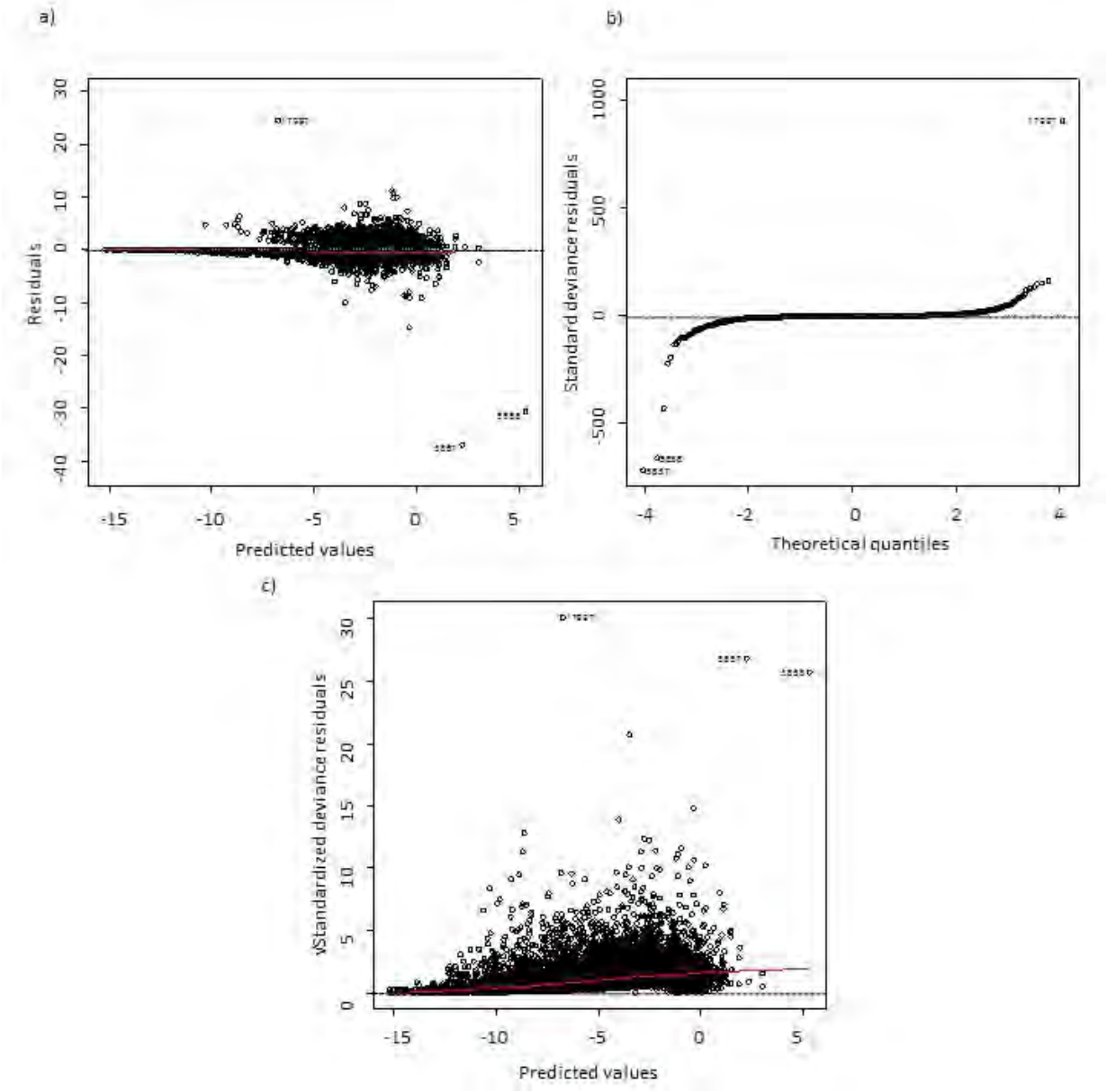

Figure 14. Diagnostic plots for model 5 predicting average shoreline occupancy including historic 1999 eelgrass predictors. a) Residuals vs predicted with loess plot overlaid. b) Normal Q-Q plot, and c) Scale-Location plot. Note the presence of three outliers.

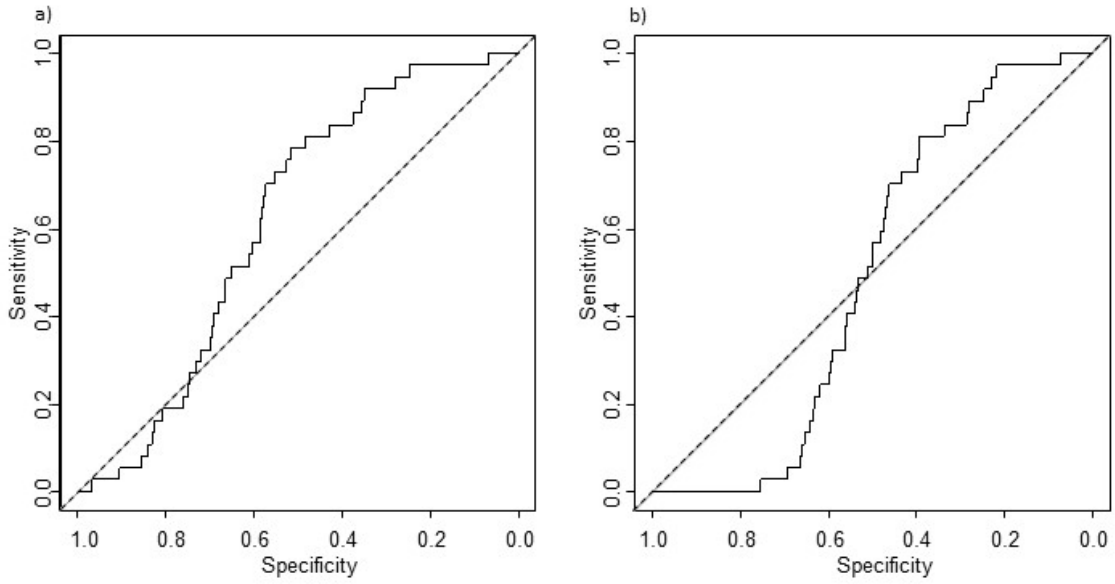

Figure 15. ROC curve for model 5 a) before and b) after removal of three outliers.

3.1.2.5 Shoreline Segment Minimum Depth of Occurrence

Based on a comparison of AIC values, the final GLM model for predicting minimum depth of seagrass occurrence is:

(6)

bathymmin = fSED4 * halflen0708gsmMax,

where bathymmin = minimum depth of seagrass occurrence (m)

fSED4 = sediment class

halflen0708gsmMax = maximum wave mixing depth over growing season

Maximum wave mixing depth proved to be a better predictor of minimum seagrass depth of occurrence than the average value or average + 2SD (Table 11). Seagrass occurring on sediments of class 12 (sand-silt-clay) were most sensitive to wave mixing depth (Figure 16). Residuals showed no evidence of nonlinearities, a slight tendency for increasing variance with the mean, and were close to a normal distribution but with three outliers identified (Figure 17).

Table 11. Fixed effects for model 6 predicting seagrass minimum depth of occurrence by shoreline distance.

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.30	0.09	3.4	8.07E-04	***
halflen0708gsmMax	0.03	0.01	4.7	2.89E-06	***
fSED412	5.53	1.30	4.3	2.34E-05	***
fSED47	0.74	0.11	6.5	1.50E-10	***
fSED4810	0.25	0.25	1.0	0.32	
fSED4810 x halflen0708gsmMax	0.004	0.019	0.2	0.82	
fSED47 x halflen0708gsmMax	-0.04	0.01	-4.3	1.65E-05	***
fSED412 x halflen0708gsmMax	-0.37	0.10	-3.8	0.000158	***

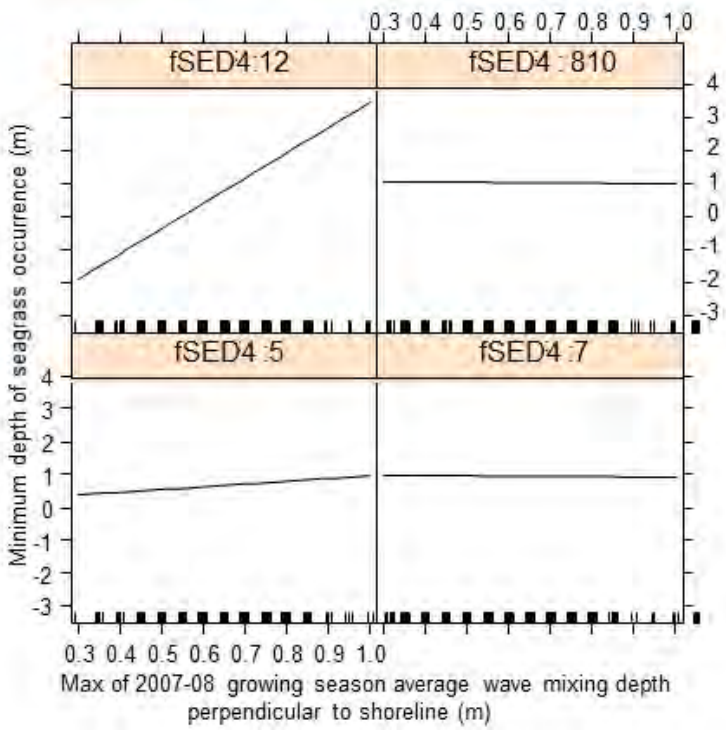

Figure 16. Interaction of sediment class and maximum wave mixing depth on minimum depth of seagrass occurrence (model 6). SED4 class definitions are: 12 = sand-slit-clay; 810 = silty and sandy-silt; 5 = sand; 7 = silty sand.

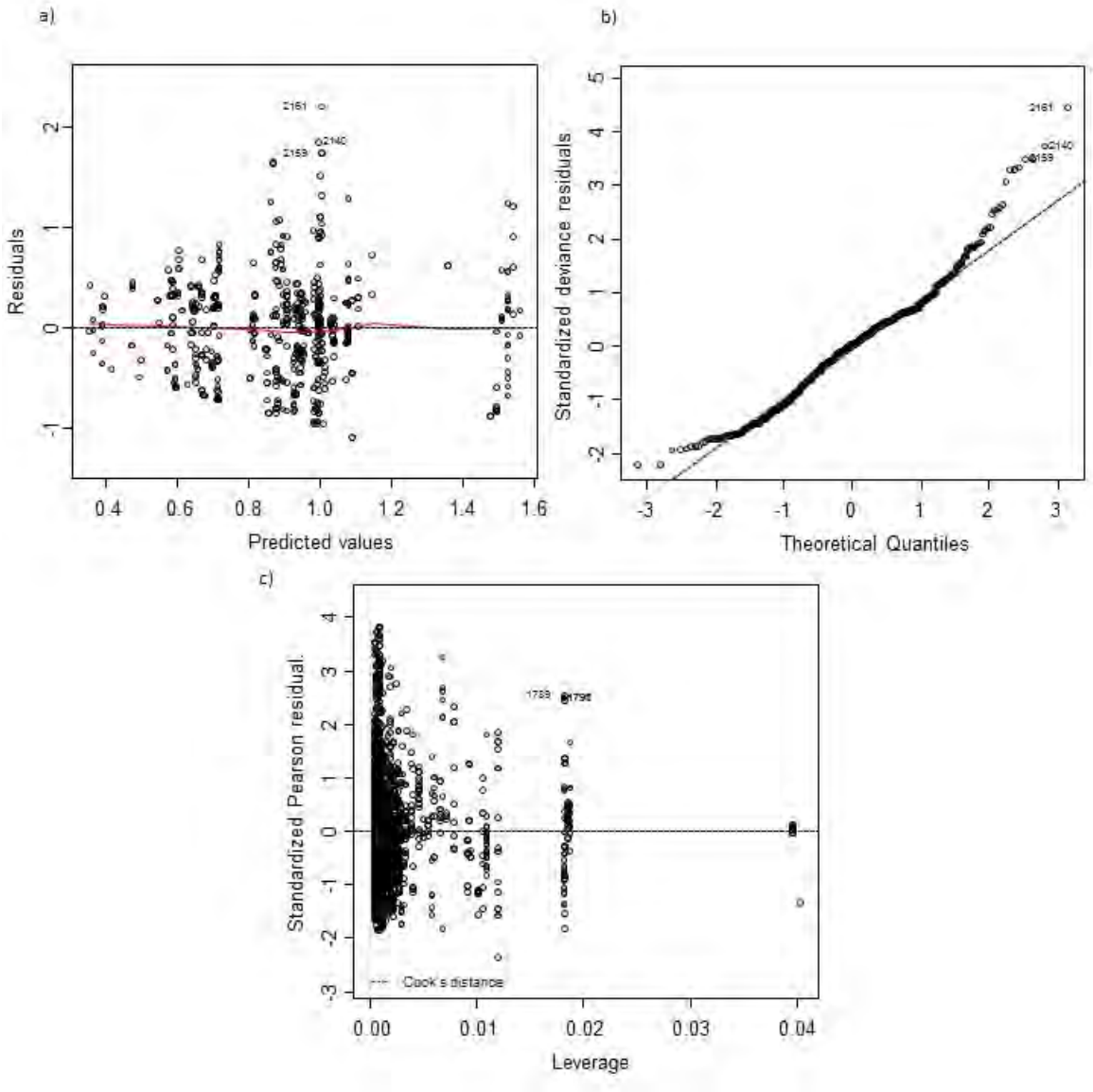

Figure 17. Diagnostic plots for model predicting minimum seagrass depth of occurrence (model 6) a) Residuals versus predicted values, b) Q-Q plot of residuals(dashed line shows expectation for normal distribution), and c) residuals versus leverage, highlighting presence of three outliers (labelled).

3.1.2.6 Shoreline Segment Maximum Depth of Occurrence

Based on a comparison of AIC values, a GAM model provided a superior fit over GLM models in predicting maximum depth of seagrass occurrence (Figure 18):

bathymmax = s(cL10secchiavMax) + s(cL10pttocMax) + s(cL10USRMARIkm2Max) r^2_{adj}= 0.2 (7)

where s = smoothing function

cL10secchiavMax = centered $\log_{10}$ maximum of seasonal average Secchi depth (m)

cL10pttocMax = centered $\log_{10}$ maximum of sediment total organic carbon (%), and

cL10USRMARIkm2Max = centered $\log_{10}$ maximum of unsewered residential units on high infiltration soils/km^2

Although the GAM model provided the best fit from a statistical viewpoint (lower AIC value of 6838.7 as compared to 7110.6), the best GLM model may provide a more practical and realistic option, i.e., without overfitting the data. Therefore, we are presenting the best GLM model fit as well, determined using the step option in GLM, which had an AIC of 7110.6 as compared to the next best model with an AIC of 7112.2:

bathymmax = csecchiminMax + cpttocMax + cUSRMARIkm2Max + csecchiminMax:cpttocMax + csecchiminMax:cUSRMARIkm2Max + cpttocMax:cUSRMARIkm2Max, (8)

where bathymmax = maximum depth of seagrass occurrence at shoreline index i

csecchiminMax = centered maximum of Secchi depth seasonal minimum at shoreline index I, and

cpttocMax = centered maximum of sediment percent total organic carbon at shoreline index i

Although main effects for minimum Secchi depth and density of unsewered residences on high infiltration soils were not significant, they were retained in the model because these terms were included in significant interaction terms. Maximum depth of seagrass

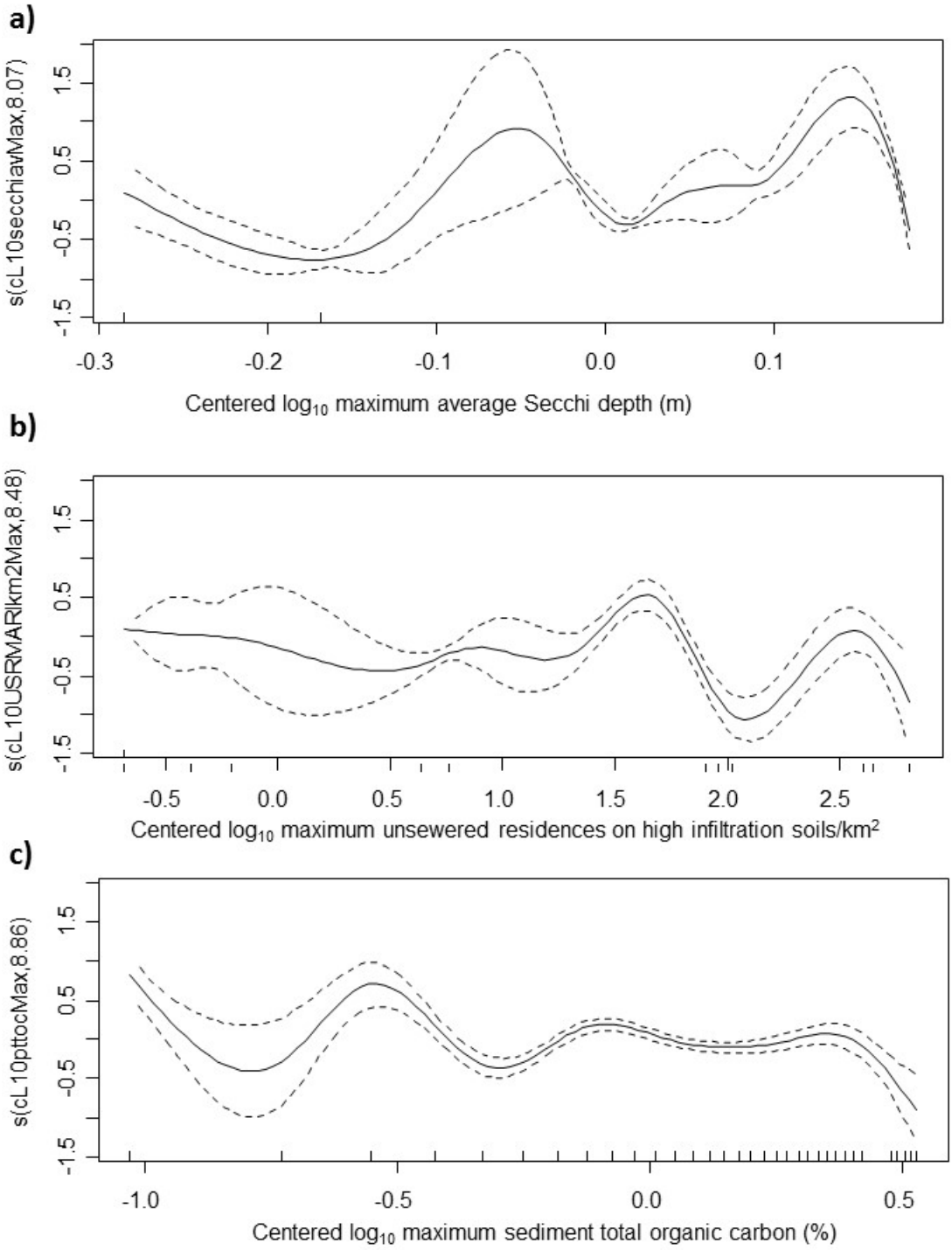

Figure 18. Smoothing functions in generalized additive model 7 to predict maximum depth of seagrass occurrence by shoreline position. a) Centered log_{10} shoreline maximum average Secchi depth (m), b) Centered log_{10} shoreline maximum unsewered residences on high infiltration soils/km^2, and c) Centered log_{10} shoreline maximum sediment total organic carbon (%).

occurrence tended to increase with sediment total organic carbon at high Secchi depths but decrease with TOC at low Secchi depths (Figure 19). Interaction effects involving density of unsewered residences were significant but weak (Table 12).

Table 12. Fixed effects in generalized linear model 8 predicting maximum depth of occurrence for seagrass at shoreline index i.

Coefficients:	Estimate	Std. Error	t-value		Pr(>|t|)	
(Intercept)	1.43	0.02	79.3	<	2.00E-16	***
cUSRMARIkm2Max	0.0003	0.0008	0.4		0.72197	
csecchiminMax	-0.02	0.03	-0.6		0.53662	
cpttocMax	-0.26	0.03	-8.7	<	2.00E-16	***
csecchiminMax x cpttocMax	0.63	0.05	13.7	<	2.00E-16	***
csecchiminMax x cUSRMARIkm2Max	0.004	0.001	3.1		0.00166	**
cpttocMax x cUSRMARIkm2Max	0.003	0.001	1.9		0.05699	.

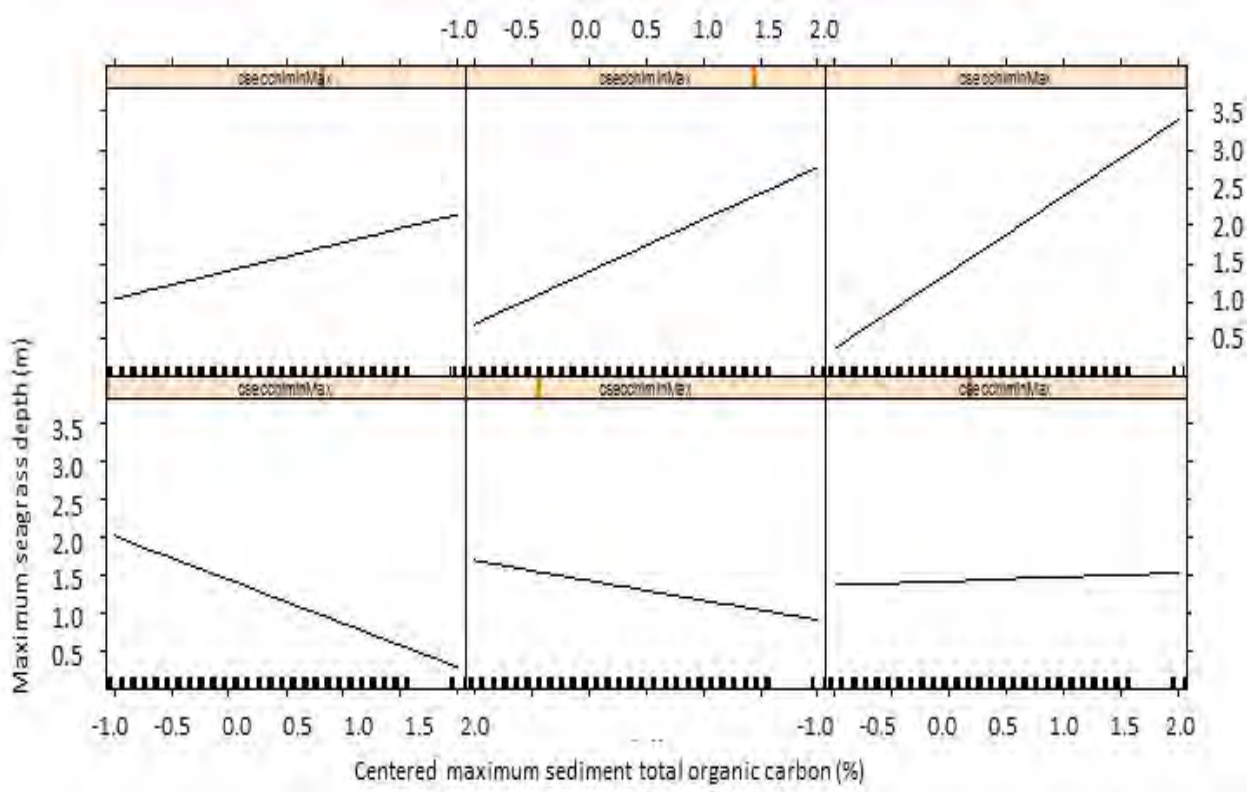

Figure 19. Interactive effects of centered sediment total organic carbon and shoreline maximum of seasonal minimum Secchi depth on maximum seagrass depth at shoreline index I (model 8). Sediment TOC has a stimulatory effect at high Secchi depths but an inhibitory effect at low Secchi depths. Secchi depth for each plot is indicated by brown vertical line in each tan horizontal header.

Chapter 4. Discussion

4.1 Data and modeling challenges and constraints for statistical predictive seagrass models

Development of predictive statistical models for seagrass occurrence is inherently challenging, but the problems can be addressed through application of a framework to describe spatial autocorrelation of seagrass patches, new techniques for incorporating SAC terms into predictive models, and careful selection of independent variables to avoid cross-correlations. Problems to be solved include the need to predict at fine spatial resolutions over large areas with associated memory constraints, spatial autocorrelation of model residuals with anisotropy, serial correlations in time series, and potential cross-correlation of predictive variables that co-vary with water depth or with salinity gradients from head-of-tide to estuarine mouth. We successfully applied strategies to limit difficulties with parameter cross-correlations: using centered variables, carefully selecting the reference sediment class against which to compare other sediment factors, choosing alternative strategies to relate transparency and mixing depth to water depth as a difference versus binary variable, and dropping correlated variables (e.g., temperature) of lesser interest. After quantifying the actual range of SAC both parallel and perpendicular to the shoreline, we were successful in incorporating terms for spatial autocorrelation into models using a modified version of the Crase et al. (2012) approach. Although memory constraints limited the number of R packages available for analysis, we were successful in applying the glmmPQL package.

Our modelling approach and national data sources should be applicable to other estuaries, with substitution of localized data sources for finer resolution data or more site-specific information on seagrass disturbances where appropriate. With some exceptions, the significant components of the predictive models developed here for seagrass presence are available for estuaries across the United States: bathymetry, salinity, grain size, wave mixing depth (derived from bathymetry and wind speed and direction), sediment percent total organic carbon, density of unsewered residences on high infiltration soils, and historic seagrass extent. For some variables, we used more localized sources of data because of their improved spatial resolution, e.g., sediment grain size. While satellite imagery is available to predict optical properties of seawater, most algorithms developed to date are not appropriate for coastal waters (< 30 meters depth) and so we used local monitoring data for Secchi depth. As coastal algorithms for optical parameters are improved, appropriate satellite data will become more readily available (Keith et al. 2014). We were able to use NOAA's WEMo model to predict wave energy (Fonseca and Malhotra 2010); however, this model has not been updated for versions of ArcMap beyond 9.3 so will not be readily available to other users. It is possible that coarser estimates of relative wave energy such as metrics within the USGS Coastal Vulnerability Index (Hammar-Klose and Thieler, 2001) could serve as a proxy for WEMo model values; that option has not been tested here. Other researchers will have access to an updated version of the USGS WAVES extension for ArcMap, which provides estimates of wave mixing depth. We relied on some local maps of disturbance indicators (distance to hardened shoreline and marinas, Canada goose density) but these effects were marginal in models.

It is possible to incorporate spatial autocorrelation in predictive models for seagrass, rather than avoiding it by sampling at coarser scales to avoid SAC, and ignoring the inherent patchiness of seagrass distributions. Seagrass currently occurs in narrow linear patches within Narragansett Bay and thus we modeled seagrass occurrence at a fine spatial scale (10 meter grid cells). Even after restricting model boundaries to the zero to five meter depth zone and to the parameter space within which seagrass is currently found in Narragansett Bay, we were still faced with the challenge of developing models with a very large data set (518856 10 x 10 meter grid cells or 19172 10-meter length shoreline locations). The magnitude of our data set restricted the statistical packages and approaches we could use for binomial GLMMs and GAMMs due to memory and processing time constraints. However, incorporation of a residual autocorrelation term in models allowed us to account for spatial autocorrelation in models in a less memory-intensive manner than would have been required by simultaneous fitting of spatial autocorrelation model distributions (with anisotropy). We were able to create realistic representations of spatial patterns (patchiness) in seagrass distribution by 1) incorporating shorelines as a random "founder" effect, 2) defining spatial coordinates relative to longshore and offshore distance, and 3) calculating moving averages of residual autocorrelation based on zones reflecting different ranges of spatial autocorrelation in longshore versus offshore directions.

Unexplained variation in our models could have been related to factors we did not include in predictive models due to lack of available data, including: biotic effects (epiphyte/grazing interactions, macroalgae competition, bioturbation and bioirrigation effects by fauna [Nelson 2009]), sediment sulfide/redox (Goodman et al. 1995), tidal range effects on minimum depth potentially interacting with light availability (Koch et al. 2001), wave energy associated with winter storms (Kelly et al. 2001), historic hurricanes or tropical storms, tidal currents, anthropogenic disturbance (shellfish dredges or rakes, anchor beds, propeller scars; see Neckles et al. 2005, Oakley et al. 2013), and restoration activity. We also failed to include all potential interaction terms in the models due to problems with variance inflation factors. Any of these factors could have contributed to unexplained variation; yet, despite this and the other challenges inherent in predicting seagrass distributions, our prediction of seagrass absolute or average presence/absence at shoreline locations was very robust, with area-under-the-curve (AUC) values associated with Receiver Operating Characteristic (ROC) curves of 0.95 – 0.98

4.2 Comparative performance of alternative modeling approaches

This was the first study to explicitly address spatial autocorrelation in eelgrass models using mixed models. Failure to account for spatial autocorrelation in species distribution models can inflate the significance of explanatory variables (Crase et al. 2012). Most researchers developing predictive models for eelgrass habitat have either ignored the potential for spatial autocorrelation and effects on variable selection (e.g., Krause-Jensen et al. 2003, Grech and Coles 2010) or have selected sample points at a minimum distance apart to avoid spatial autocorrelation (Downie et al. 2013). Although the latter strategy takes care of the issue from a statistical standpoint, it fails to acknowledge the positive feedback effects of seagrass on sediment stability and the light environment, and cannot successfully predict the patchy nature of seagrass distribution. However, patch dynamics or

cellular automata models (Wortman et al. 1997, Fonseca et al. 2000b, Fonseca et al. 2004, Schonert and Milbradt 2005) do predict this patchiness. Incorporation of a term for spatial autocorrelation allowed us to avoid overestimation of percent variance explained and to reduce the number of significant predictors in our final regression models to an appropriate number. Examples of terms dropped following addition of SAC error terms include the shoreline isolation term in some models and the number of sediment classes among which we could detect differences in either main effects or interactions with wind energy or wave mixing depth.

Models predicting seagrass presence were more robust at the shoreline scale than at the individual grid cell (Table 13). Between original tests and 10-fold cross validation results, model performance measurements declined considerably based on 10-fold cross-validation at the grid-cell scale but stayed constant at the shoreline scale. Model performance was slightly better for prediction of presence/absence as compared to average presence/absence at the shoreline segment. Model performance for prediction of average shoreline presence/absence was significantly degraded after information on historic eelgrass presence/absence was included in models.

Table 13. Summary of model performance at different scales, based on resubstitution or 10-fold cross-validation and with or without incorporation of data on historic seagrass coverage.

Dependent variable	Historic P/A included?	Resubstitution or 10-fold cv?	AUC
Grid cell P/A	No	R	0.98
Grid cell P/A	No	10xcv	0.71
Shoreline max P/A	No	R	0.95
Shoreline max P/A	No	10xcv	0.95
Shoreline max P/A	Yes	R	0.98
Shoreline max P/A	Yes	10xcv	0.98
Shoreline avg P/A	No	R	0.89
Shoreline avg P/A	No	10xcv	0.72
Shoreline avg P/A	Yes	R	0.62
Shoreline avg P/A	Yes	10xcv	0.62

We have chosen the AUC statistic to compare performance across our models and between our models and those of others because it is insensitive to the probability of occurrence of the species of interest, unlike overall prediction accuracy and some other measures (Fielding and Bell 1997). Some researchers present only an overall prediction accuracy, so we are unable to compare their results with ours. The AUC statistic represents the area under a ROC curve, and varies between 0 and 1, with a value of 0 representing zero predictive power, 0.5 representing predictions no better than chance, and 1 representing perfect predictions. The ROC curve is a plot that demonstrates the performance of a model predicting a binary variable. It is created by plotting the true positive rate (sensitivity) against the false positive rate (1 – specificity) as the discrimination threshold is varied.

Our models predicting seagrass P/A at the 10 x 10 meter grid cell scale did not perform as well as some existing models predicting P/A at 25 x 25 meter or 50 x 50 meter grid cell scales, but our model performance for predicting shoreline occurrence exceeded earlier reported AUC statistics. Valle et al. (2013) calculated AUC values with 5-fold cross-validation to compare regression approaches (GLM, ANN, MARS, GAM) with machine learning methods (BRT, RF) to predict seagrass presence/absence at the (50m x 50m) grid cell scale. At a grid-cell scale 25 times coarser than our model predictions, their best GLM model performance had a mean AUC of 0.84, while their best models based on boosted regression tree methods yielded a mean AUC of 0.94. Using an independent test set comprised of 30% of available data, Downie et al. (2013) compared the model performance using GAM and maximum entropy modeling (Maxent) and found model AUC values of 0.84 (GAM) and 0.80 (Maxent) at a grid cell scale of 25 x 25 meters. It is possible that our model performance at the grid cell scale is lower than that obtained by other researchers because of the finer scale of our grid cells, or because of differences in study area characteristics and the predominant form of *Z. marina* (intertidal annual versus subtidal perennial) found in each setting. Our models predicting shoreline presence were robust and yielded predictive accuracy (AUC values) equal to or better than the results of Downing et al. (2013) or Valle et al. (2013). It is important to note, however, that our models incorporated a random shoreline effect which explained much of the background variability within the estuary (see discussion below).

Although GAM models can perform better than GLM models based on AIC and AUC values, they may overfit the data and can be replaced with GLM models incorporating higher-order terms to capture nonlinearities. In the one case where we could derive a GAM model for comparison with a GLM to predict maximum depth of eelgrass occurrence, the statistical fit of the GAM model was better based on AIC values. Likewise, Valle et al. (2013) found better performance for GAM models (mean AUC = 0.93) as compared to GLM models (mean AUC = 0.84) in predicting eelgrass presence/absence. However, the multiple peaks we observed in response curves had no mechanistic basis, suggesting that GAM models were probably overfitting the data. It is likely that higher order terms in GLM models are sufficient to capture nonlinear responses in logistic models. Addition of a second order term can yield an S-shaped probability plot, while addition of a third-order term can describe a unimodal response to an environmental gradient.

4.3 Relative importance of different factors in predicting seagrass P/A

4.3.1 Seagrass sensitivity to different environmental variables in Narragansett Bay

We can compare the sensitivity of seagrass survival and persistence to different environmental variables by comparing the potential change in ln(odds ratio) over the range of each predictor in the Narragansett Bay area of interest. For predictors involving higher order terms, we can predict minimum and maximum potential contributions. For interaction terms, we can examine sensitivity at high and low ends of modifying factors (Table 14). For the model predicting seagrass presence/absence at the grid cell scale, the most influential predictor was Secchi depth, followed by, in order: shoreline isolation, sediment percent total organic carbon, sediment type, and salinity. The least influential

variable was water depth greater than average wave mixing depth. For the model predicting presence of seagrass at shoreline locations, the most influential predictor is sediment type, followed by percent total organic carbon (at low Secchi depth), then salinity.

4.3.2 Seagrass sensitivity to different environmental variables in other regions

Managers need to consider the energy environment of different regions and estuarine hydrogeomorphic types and different growth forms in comparing the predictions of seagrass models and the relative influence of different predictor variables. Most of the regression or machine-learning models developed for predicting presence/absence of *Zostera marina* are based on analyses of western Europe datasets, e.g., the Baltic Sea and Wadden Sea (Krause-Jensen et al. 2003, van der Heide et al. 2009, Downie et al. 2013, Valle et al. 2013,) or of the Great Barrier Reef (Grech and Coles 2010). In many of these cases, the energy regime (relative wave exposure, current velocity) and/or substrate are the predominant variables predicting eelgrass presence (Grech and Coles 2010, Downie et al. 2013, Valle et al. 2013). However, investigators predicting eelgrass presence over larger regions have found that light availability (Krause-Jensen et al. 2003, van der Heide et al. 2009), total N in surface water (van der Heide et al. 2009), and, to a lesser extent, salinity (Krause-Jensen et al. 2003) are the driving factors predicting eelgrass presence.

Results from the models developed for western European regions may not be comparable to our results for Narragansett Bay because they focus on higher energy environments (or those with more extreme physico-chemical gradients). In addition, *Zostera marina* populations in the Wadden Sea are comprised of annual forms which occur in intertidal zones dominated by frequent disturbance and less subject to light limitation. Both the robust perennial form of *Z. marina* and the more flexible annual form occurred in the Wadden Sea prior to the incidence of wasting disease in the 1930s, but the perennial form never recovered (van Katwijk et al. 2009). Narragansett Bay populations of *Z. marina* are currently subtidal and are presumably composed of the perennial form. Most studies of the factors affecting eelgrass presence/absence have ignored the distinction between annual and perennial forms of Z marina. The annual form of *Z. marina* has been reported in Nova Scotia and Maine estuaries (Keddy and Patriquin 1978) and in Ninigret Pond, RI (Thorne-Miller et al. 1983) but may be present in other New England estuaries where eelgrass occurs in intertidal zones (e.g., Great Bay).

Table 14. Potential effect of independent variables across range of predictors. For sediment types, the range of coefficients is given. Effects are expressed in terms of ln(odds ratio). Maximum effect for Secchi depth is at an intermediate value of the predictor, not the maximum.

Model	Predictor	Min	Max	Coeff	Min effect	Max effect	Effect range	Rank
7	cSDavgrtrZ, cSDavgrtrZ^2, cSDavgrtrZ^3	-3.00	4.23	1.65, -0.149, -0.053	-6.14	1.09	7.23	1
7	fISOLATED1	0	1	-5.42	0.00	-5.42	-5.42	2
7	cptTOC, cptTOC^2, cptTOC^3	-1.26	2.24	-0.91, 0.65, -0.33	-3.76	1.56	5.33	3
7	fSED4	0	1	-1.28 to 1.89	-1.28	1.89	3.17	4
7	cSAL, cSAL^2	-1.62	0.48	1.56, 0.18	-3.33	-0.49	2.84	5
7	cUSRMARIkm	-31.02	285.18	0.004	-0.12	1.14	1.26	6
7	cDstoMarin	-2242.6	252.71	-0.0005	1.12	-0.13	-1.25	7
7	cCG046avkm	-3.60	28.60	0.013	-0.05	0.37	0.42	8
7	fZgtMXZavT1	0	1	0.35	0.00	0.35	0.35	9
8	fSED8	0	1	-3.55 to 3.62	-3.55	3.62	7.17	1
8	cpttocMin at min csecchimin				1.46	-3.05	-4.51	2
8	csalAv	-1.72	0.38	1.57	-2.70	0.60	3.30	3
8	cpttocMin at max csecchimin				0.54	-1.13	-1.67	4
8	cDistToHdS	-537	1965	-6.55E-04	0.35	-1.29	-1.64	5
8	fZgtMXZavMTRUE	0	1	1.55	0.00	1.55	1.55	6
8	cwindMin	-8572	73428	-1.50E-05	0.13	-1.10	-1.23	7
8	cCG046avkm	-3.9	28.3	0.033	-0.13	0.93	1.06	8
8	cDstoMarin	-328	2174	-2.55E-04	0.08	-0.55	-0.64	9
8	csecchimin at max cpttocMin				-0.801	-0.055	0.75	9
8	csecchimin at min cpttocMin				-0.004	0.010	0.01	10
8	csecchimin	-0.84	2.16	0.31				
8	cpttocMin	-1.13	2.37	-1.06				
8	csecchimin:cpttocMin			0.27				

4.3.3 Limitations to existing regression models to predict seagrass in U.S. estuaries

In contrast to our approach, most regression models predicting eelgrass presence and/or cover in U.S. estuaries include a smaller subset of variables at a time, generally focusing either on nutrients and/or light availability (Duarte 1991, Latimer and Rego 2010, Benson et al. 2013, Kenworthy et al. 2014) or on the energy regime (Kelly et al. 2001) but not considering additive or interaction effects (Koch 2001). Developers of habitat or transplant suitability indices for eelgrass have considered a combination of eelgrass colonization sources, substrate characteristics, wave exposure, and water quality/light

environment based on data from empirical or experimental studies but tend to give these factors equal weight and do not consider potential factor interactions. In more focused studies, researchers have provided empirical and/or experimental evidence for interactive effects of light availability with sediment organic matter content (Wicks et al. 2009, Kenworthy et al. 2014) or of salinity with nutrients (van Katwijk et al. 1999).

4.3.4 Comparison of estimates for light compensation depth and optimum light levels

Interpretation of predictions of light compensation points (maximum depth occurrence) and optimum light levels depends on the vertical datum of the merged topobathymetric grid used to estimate water depth. We used NOAA's Coastal Relief Model (CRM), in which source bathymetric data retained their original vertical datum of either mean lower low water (MLLW) or mean low water (MLW), while source topographic data remained in either North American Vertical Datum (NAVD) 88 or National Geodetic Vertical Datum (NGVD) 29 prior to merger. Given the semi-diurnal nature of tides in Narragansett Bay, the MLLW value is the tidal datum of interest. Mean tidal range between MLLW and Mean Higher High Water (MHHW) varies from 0.876 to 1.55 meters at NOAA tidal stations in Narragansett Bay having a recorded datum (http://tidesandcurrents.noaa.gov/stations.html?type=Datums), with an average value of 1.27 meters.

Our model for seagrass presence/absence at the grid cell scale yields comparable information on effects of light availability on seagrass maximum depths to values in the literature. Our model for seagrass presence at the grid scale predicts a maximum probability of occurrence at a centered Secchi depth (average Secchi depth – water depth) value of 2.4 meters, which corresponds to a precentered value of 2.4 + (-0.433) = 1.97 meters. Our model also predicts an odds ratio of 1 (p = 0.5) at a centered difference value of 0.985 meters, corresponding to a pre-centered value of 0.985 - 0.433 = 0.552 meters. If we make the assumption that eelgrass light limitation depends on the mean tidal level, then the model-predicted compensation depth would be 0.552 - (1.27/2) = -0.083 meters greater (0.083 meters less) than the Secchi depth, or (2.44-0.083)/2.44 = 0.96 times average Secchi depth. Duarte's compilation of compensation points for *Z. marina* from the literature yields a prediction of $Z_c = 1.86/K$, as compared to Dennison's (1987) value of 1.62/K for northeastern U.S. estuaries, and Nielsen and colleague's (1989) value of 1.53/K. Using Poole and Atkin's (1929) relationship between the light attenuation coefficient, K, and Secchi depth (K = 1.7/S.D.), these correspond to a range of 0.9 – 1.11S.D. for Zc. Our estimate of 0.96 times average Secchi depth at an odds ratio of 1 (probability of 0.5) falls well within the range of values reported by Duarte (1991) for the light compensation depth.

Although our model predicted an optimum light level for seagrass growth, the shoreward limits to seagrass are more likely related to physical disturbance. Most researchers have only evaluated maximum depth limits for eelgrass related to light limitation, with no estimation of optimum light levels. Krause-Jensen et al. (2003) found that the probability of eelgrass occurrence in Danish coastal waters increased up to 60% of surface irradiance (at about 4 meters depth), and then cover declined at higher values. Although Krause-Jensen suggested that although this could have been related to photoinhibition at higher levels, our models suggest it was more likely related to the increased probability of

physical disturbance at shallower depths. Optimum light values for eelgrass in our model 7 occurred at depths of 1.97 meters less than average Secchi depth, or approximately 2.44 – 1.97 + (1.27/2) = 1.105 meters depth relative to Mean Tidal Level (MTL) for average Secchi. A Secchi depth of 2.44 meters corresponds to a K_d of 0.70, which would yield a light intensity of 1% incident light at 6.6 meters depth.

Our model for seagrass shoreline occupancy incorporated an interaction term between water depth relative to Secchi depth and sediment percent total organic carbon which is consistent with the literature. This interaction has also been observed by Kenworthy et al. (2014) for southeastern MA embayments, with a decrease in eelgrass compensation depth as sediment organic matter increases. This is likely due to the increase in sulfide content of sediments with increased organic matter content and associated toxicity (Goodman et al. 1995, Holmer and Bondgaard 2001).

4.3.5 Seagrass sensitivity to factors other than transparency

Our models showed mixed evidence for the effect of unsewered residences on high infiltration soils, an indicator of potential groundwater N inputs. In two cases, model coefficients were negative, in one case positive, and in one case the variable showed weak interaction effects with other nutrient-related variables. Valiella's model incorporates the potential effect of unsewered residences on nitrogen loads (Valiella et al. 2004), but applications of his model to predict load effects on seagrass have not attempted to separate the effects of reduced transparency related to phytoplankton biomass with other effects related to groundwater inputs (Latimer and Rego 2010). It is possible that groundwater DIN inputs could favor the growth of macroalgae and/or periphyton at the expense of *Z. marina* (Harlin and Thornemiller 1981, Costa 1988, Teichberg et al. 2010). Unsewered residence density might have shown a stronger and more consistent effect in our models if we had accounted for variability in residence time in various subembayments within Narragansett Bay as was done by Latimer and Rego (2010).

The range of salinity values encompassed by effects predicted by our model (26.6 – 28.7) is well within the tolerance ranges reported for *Zostera marina* populations in the literature: 5 to 35 Practical Salinity Units (PSU) in the northern hemisphere (den Hartog 1970) but slightly greater than ranges of 14 to 22 PSU reported for the Chesapeake Bay (Wetzel and Penhale 1983). Salinity effects predicted by our models could have been related to either direct negative effects of salinity and energetic costs of osmoregulation, the interactive effects of salinity and DIN, or the correlation of the salinity gradient with the gradient of dissolved inorganic nitrogen (DIN) and total nitrogen (TN) in Narragansett Bay (Krumholz 2012). Salinity and nitrate can have synergistic interactions, possibly related to the tendency to incorporate low C (high N) amino acids as tissue nitrogen levels increase in contrast to the need to generate the high C/low N amino acid proline involved in osmoregulation (van Katwijk et al. 1999). Van Katwijk et al. (1999) observed a decreased tolerance of *Z. marina* to eutrophication at salinity levels of 26–30 ppt. It is likely that the negative coefficient for salinity in our models reflects the negative impacts of water column TN, and possibly an interaction between salinity and TN effects.

We found a strong interactive effect between wave mixing depth and sediment type on minimum depth of seagrass occurrence, with the effect strongest on gravel and sandy

gravel sediments, which probably also represent high energy environments. However, wave mixing depth was the least influential predictor in our eelgrass model predicting seagrass presence at the grid cell scale. Numerous researchers have reported a negative effect of relative wave energy on seagrass presence, both in western Europe systems (Downie et al. 2013, Valle et al. 2013) and in the United States (Kelly et al. 2001), but these cases represent systems with more extreme energy gradients. Fewer researchers have tried to test the relationship of wave mixing depth with minimum depth extent of eelgrass. Chiscano (2000 in Koch 2001) found a poor relationship between minimum depth of occurrence and wave mixing depth in shallow portions of the Chesapeake Bay with gentle slopes. Our models indicated that wave mixing depth in combination with sediment particle size was a significant factor determining minimum depth of seagrass occurrence.

The effects of sediment characteristics on habitat suitability for eelgrass have historically been examined in isolation. Most evaluations of particle size effects have been correlative in nature (Nelson 2009), but two studies did demonstrate greater growth rates on finer sediments than on coarse sand or sand plus gravel, possibly related to nutrient availability (Short 1987, Thom et al. 2001). Earlier evaluations of the effect of sediment organic matter on eelgrass presence tended to be correlative as well, but more recent studies have elucidated the same interaction between light compensation point (maximum depth of occurrence) and percent organic matter in sediment types that we observed (Kenworthy et al. 2014).

Somewhat surprisingly, our shoreline model for seagrass relative abundance was not as robust after we incorporated historic seagrass presence, distance, and patch size as predictors in the model, suggesting that seagrass patch distribution may represent a dynamic equilibrium with significant transition probabilities. In examining a 14-year time series of intertidal eelgrass distributions in the Ems estuary in the Netherlands, Valle et al. (2013) found transition probabilities of 12.7% for colonized areas and 12.9% for areas that had disappeared, suggesting that patches are relatively mobile in this system even though total area is relatively stable. Valle et al. (2013) characterize these intertidal populations as r-selected in response to the high disturbance environment they are found in. Likewise, Kelly et al. (2001) predicted that 16% of the seagrass area in their system was highly susceptible to acute storm events. These studies suggest that predictive models will have an inherent error rate and possibly a tendency to overpredict seagrass presence, i.e., a significant proportion of suitable habitat may be unoccupied during any given year. Moreover, recolonization rates may be slower in the Northeast, leading to less patch predictability from year to year. Based on observed trajectories, Neckles et al. (2005) projected recovery times of over 10 years for dragged eelgrass patches in Casco Bay, Maine. Recovery rates may also depend on the availability of seed sources and the presence of annual versus perennial forms of *Z. marina* (Jarvis and Moore 2010) which will determine the rate of vegetative spread versus sexual reproduction. The poor contribution of location of historic seagrass patches to prediction of current seagrass patch locations suggests that Narragansett Bay is not in a state of equilibrium and that patch locations may vary from year to year, perhaps due to lags in recovery from disturbance.

Random effects of shoreline code showed extreme differences in probability of seagrass colonization, possibly due to hysteresis effects. This was surprising, given the widespread historic occurrence of seagrass in Narragansett Bay prior to the incidence of wasting disease in the 1930's (Figure 20). It is possible that this discrepancy in historic and current habitat extent is due, in part, to effects of hysteresis related to positive feedbacks of seagrass on sediment stabilization (van derHeide et al. 2007). Van derHeide et al. (2007) used a simple model relating *Z. marina* growth rates in the Wadden Sea to light availability which was affected by background levels of suspended solids (SS) in the water column, an increase in SS towards the sediment surface reflecting resuspension, and the reduction in resuspension of SS as *Z. marina* biomass increases and dampens tidal currents and wave action. This simple model predicts that seagrass and bare sediment represent alternative stable states in coastal waters. Seagrass can persist until dramatic reductions due to a disturbance or disease, after which recovery may be unlikely or impossible in some zones because of the resuspension of fine sediments in the shallow nearshore zone which reduce light availability.

Tidal currents estimated for Narragansett Bay are sufficiently strong to resuspend fine sediments and reduce the probabilities of eelgrass re-establishment following disturbance, but not strong enough to damage existing eelgrass beds. Oviatt and Nixon (1975) measured sediment resuspension rates in Narragansett Bay 8–20 times greater than sediment deposition rates, with resuspension rates greatest 1 meter above the sediment and decreasing towards the water surface. Likewise, Collins (1976) found an increase in transparency in upper waters of the Bay from north to south, but a corresponding increase in a bottom turbidity plume which extended 1 meter above the sediment in shallower stations to up to 4 meters above the sediments at deeper stations. Tidal currents or maximum orbital velocities associated with wave action greater than 15 to 30 cm/sec are sufficient to resuspend the fine sediments (clay, silt and find sand) predominant in much of Narragansett Bay (Oviatt and Nixon 1975). Maximum tidal currents during ebb flow in Narragansett Bay are typically above the 15-30 cm/s range in the mid to lower Bay, except for selected protected areas (Figure 21). Note that most thresholds for physical damage to mature eelgrass by currents are somewhat higher than those required for fine sediment resuspension, in the range of 40 – 180 cm/sec, beyond the maximum tidal currents predicted for the bay (Nelson 2009). Therefore, once established, eelgrass would be expected to survive throughout much of the bay given adequate light, sediment quality and protection from wave energy. Maximum wave orbital velocities calculated for the 2009 growing season were much lower, with areas of potential resuspension very close to the shore and average probability of suspension for fine sand equal to 6.5 % for those areas (Figure 22).

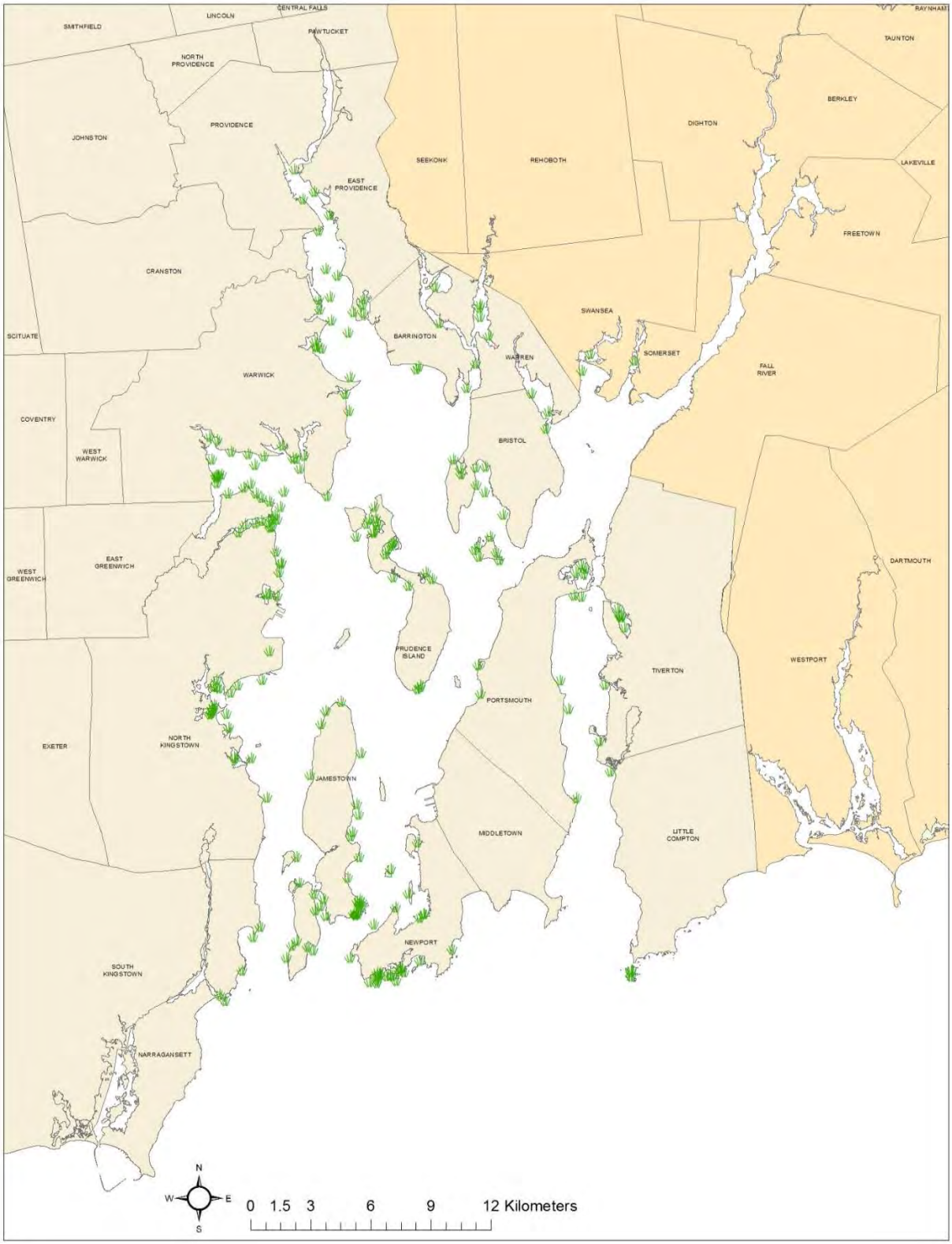

Figure 20. Historic occurrences of seagrass in Narragansett Bay compiled by Kopp (1995).

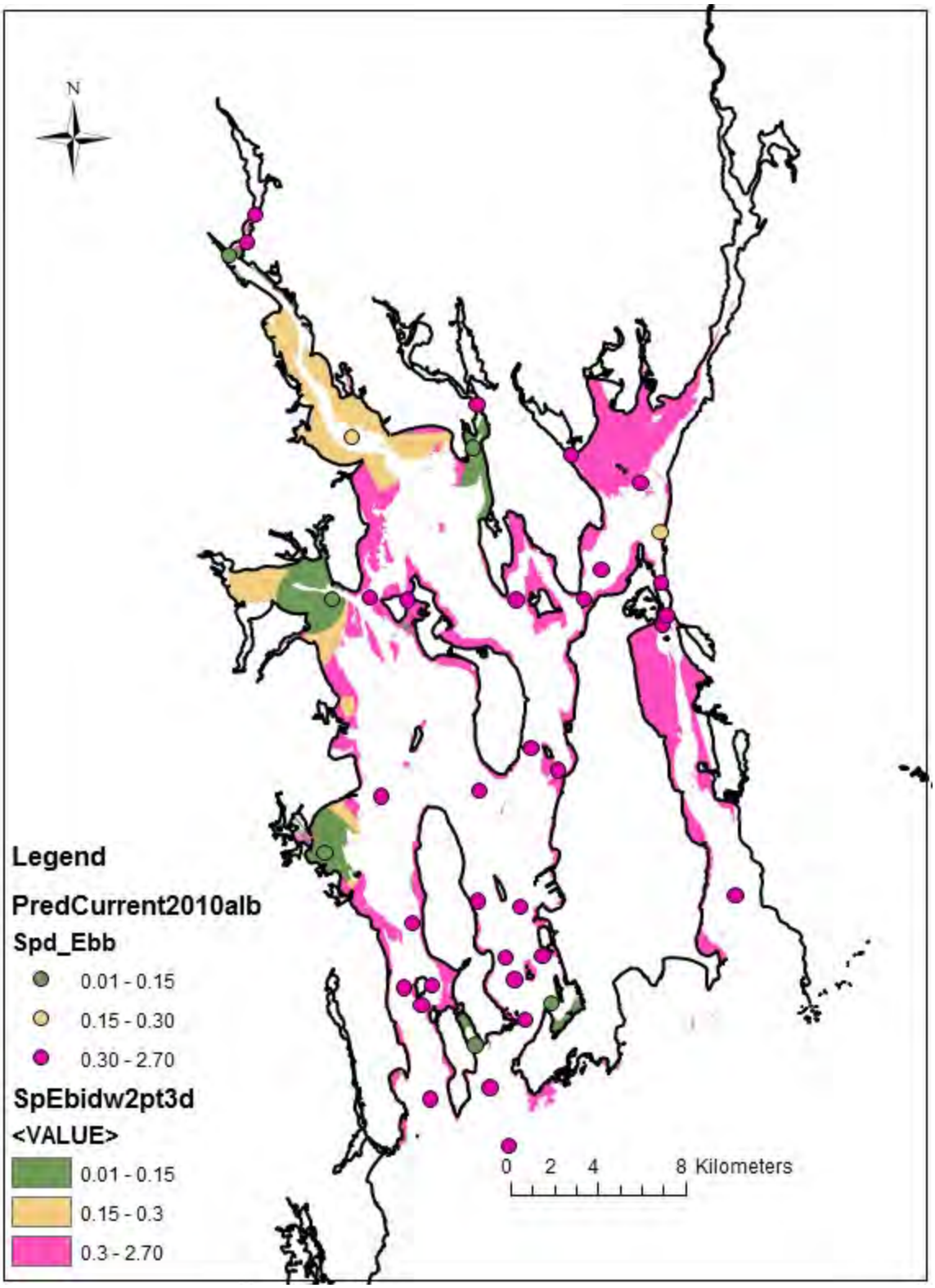

Figure 21. Predicted 2010 maximum tidal currents (m/sec) during ebb tides (from NOAA tidal currents web site: http://tidesandcurrents.noaa.gov/curr_pred.html**). Points represent predicted values at tidal stations. Shaded areas were nterpolated within the 0-5 meter depth zone by inverse distance weighting (2 point interpolation with shorelines as barriers to interpolation).**

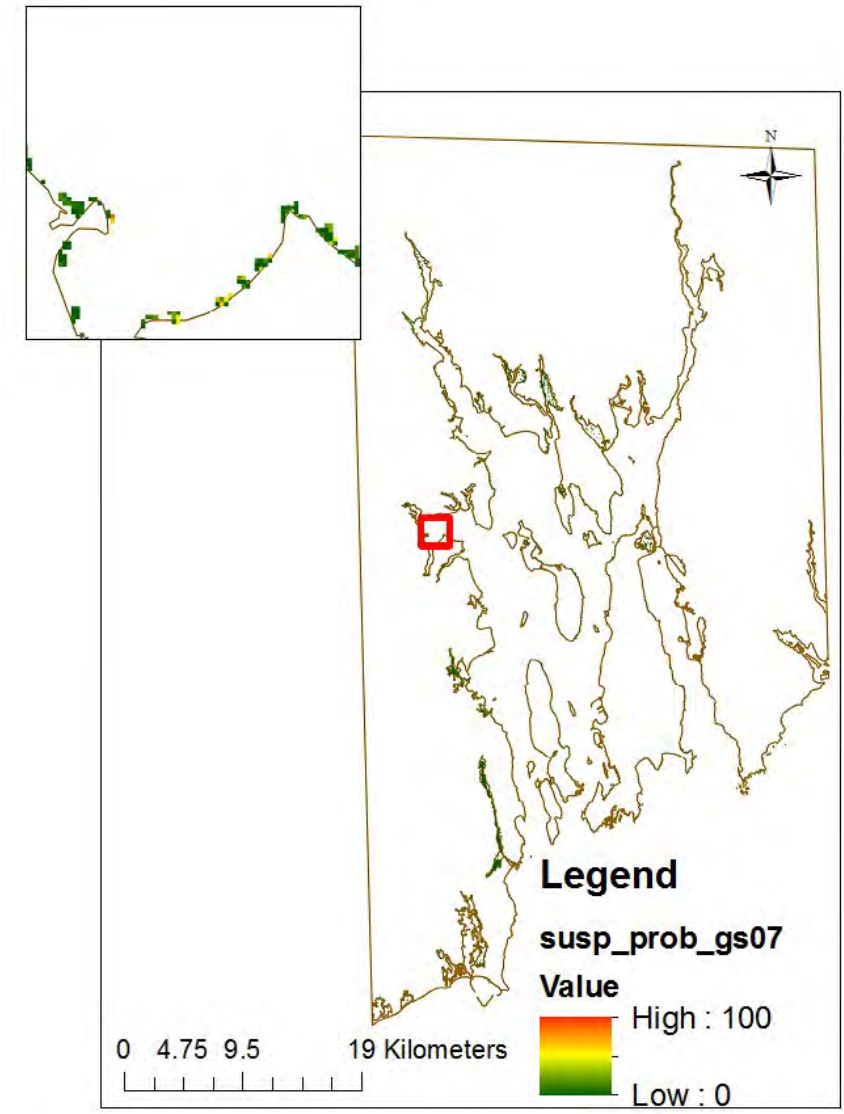

Figure 22. Probability of fine sediment resuspension based on USGS WAVES model for growing season. Average probability of resuspension for limited nearshore areas where the average probability greater than zero is 4.9%.

4.4 Model application for assessing management scenarios

Here we demonstrate an application of our model to support adaptive management, with prediction of the effects of planned reductions in total nitrogen loadings from wastewater treatment plants (WWTPs) and from atmospheric deposition. Our predictive models contain four terms related to direct or indirect nitrogen effects on seagrass presence: density of unsewered residences on high infiltration soils (indicator of groundwater N inputs), salinity (indicator of N gradient), Secchi depth, and sediment percent total organic carbon. The density of unsewered residences on high infiltration soils is likely correlated with groundwater N concentrations (IEC 2012), although there may be a lag in response due to groundwater travel time. The range of salinity values encompassed by this model is well within the tolerance ranges reported for *Zostera marina* populations: 5 to 35 PSU in the northern hemisphere (den Hartog 1970) and historically, seagrass was found in

Narragansett Bay from the lower Providence River south to the mouth of the estuary (Figure 20). Thus, it is probable that the modeled positive response to salinity over the range of 26 to 28.5 PSU in Narragansett Bay represents a response to the inversely correlated gradient in surface water nitrogen concentrations from head-of-tide to the mouth of the estuary (Krumholz 2012). Nitrogen can exert negative effects on seagrass independent of light reductions from phytoplankton shading, e.g., through stimulatory effects on epiphytes (Costa 1988) and macroalgae (Teichberg et al 2010) which reduce light availability and increase anoxia in the sediments. Costa estimated that the depth of eelgrass growth in Buttermilk Bay (a well-flushed sub embayment within Buzzards Bay) decreased by 9 cm for every 1 μM increase in dissolved inorganic nitrogen due to increased shading by periphyton growth (not phytoplankton). In Narragansett Bay, an increase in salinity from 26 to 28.5 ppt corresponds to a reduction in TN from 0.508 to 0.419 mg N/L. If Costa's relationship for a well-flushed estuary holds in Narragansett Bay, this change in TN would yield an estimated increase in maximum depth of seagrass of 1.2 meters in response to reduced periphyton shading.

Model coefficients for Secchi depth capture the response of seagrass to increased light availability. Following improvements in wastewater treatment in the 1970's, the transparency of Narragansett Bay improved as TSS loadings from wastewater treatment plants (WWTPs) decreased (Borkman and Smayda 1988). Turbidity values are now uniformly low along the main axis of the bay (Nu Shuttle data; http://www.narrbay.org/d_projects/nushuttle/shuttletree.htm). It is likely that most of the downstream gradient in transparency in the upper water column is related to changes in chlorophyll *a* concentration.

Assuming that salinity is an indicator of nitrogen loading gradients related to tidal flushing, we can also estimate the effect of N load reductions indirectly by converting the salinity gradient in the study area to the corresponding gradient in total N. Based on an overlay of nutrient sampling stations from Krumholz (2012) with our salinity grid, we calculated the relationship between the estuarine gradients in salinity and total N:

$TNanavg = 1.461 - 0.0417\ Sal$ ($r^2 = 0.86$)

$TNsmravg = 1.431 - 0.0355\ Sal$ ($r^2 = 0.79$)

where TNanavg = annual average total N (mg N/L) for 2006 – 2010 surveys

TNsmravg = summer average total N (mg N/L) for 2006-2010 surveys

Sal = salinity (PSU)

We can estimate the potential effect of N concentration reductions on transparency using a regional model relating chlorophyll a to the diffuse attenuation coefficient, derived from data for 48 estuarine sites in southeastern Massachusetts (Bensen et al. 2013):

$Chl\ a = 5.70 \ln[TN] + 10.53$

$POC = 0.11\ [Chl\ a] + 0.12$

$$K = 1.06\ [POC] + 0.22,$$

where

Chl *a* = chlorophyll *a* (mg/L)

TN = total nitrogen (mg N/L)

POC = particulate organic carbon (mg/L)

K = attenuation coefficient (m^{-1})

and the relationship between the light attenuation coefficient and Secchi depth (Poole and Atkin 1929):

$$K_d = 1.7/\text{Secchi depth}$$

Combining these yields the following equation:

$$\text{Secchi depth} = 1/(0.926 + 0.388 \ln[TN])$$

The apparent light compensation point for seagrass varies with sediment percent TOC (Kenworthy et al. 2014), so estimates of potential seagrass recovery must factor in a lag time for reductions in sediment TOC. We can estimate a short-term response to increased light availability by holding sediment TOC constant, and a potential long-term response assuming reductions in TOC content down to reference levels expected based on percent silt + clay. We downloaded the raw data from McMaster's (1960) sediment collections in Narragansett Bay (Poppe et al. 2003), calculated the average percent (silt + clay) for each of Shepard's sediment classes in the data set (Table 15), and then calculated an expected reference level of sediment total organic carbon based on Pelletier et al (2011):

$$\sqrt{TOC} = 0.11\ \%\ \text{silt-clay} + 0.556$$

To simulate potential short-term effects of N load reductions we estimated a projected change in Secchi depth and a change in salinity "equivalents", the latter to evaluate potential effects of nutrient reductions beyond transparency effects. We kept sediment organic carbon levels constant, assuming that these might take some time to recover. To simulate potential recovery in the long-term, we also applied the model after reducing sediment organic carbon levels back to reference levels based on sediment particle-size class.

Table 15. Average % silt + clay in McMaster surficial sediment samples from Narragansett Bay by Shepard's sediment class and estimated reference level of sediment total organic carbon based on Pelletier et al. (2011).

Shepard_class	Avg %siltclay	Ref %TOC
Gravel	7.06	0.40
Sand	9.78	0.44
Gravelly sediment	31.35	0.81
Silty sand	40.78	1.01
Sandy silt	68.95	1.73
Sand silt clay	72.92	1.84
Clayey silt	90.32	2.40
Silt	96.20	2.61

We estimated potential reductions in total N loading to Narragansett Bay based on projected mandated changes in loadings from sewage treatment plants (Krumholz 2012), urban runoff (as a function of reduced atmospheric loads to impervious surfaces in the watershed), and reductions in direct atmospheric loadings to the water surface related to implementation of the Clean Air Act. Atmospheric load reductions were based on the projected difference in annual loads for the open waters of the estuary between 2006 and 2020, based on results from CMAQ model runs (downloaded from www.epa.gov/edm). In 2005, RIDEM mandated load reductions from sewage treatment plants at the head of Narragansett Bay to reduce wastewater N loading to the bay by 50% by 2014 (Krumholz 2012). After scheduled load reductions from WWTP beyond 2006 and projected decreases in atmospheric deposition, TN loads could be reduced further by ~40%. However, actual reductions might be lower than planned if sediment denitrification rates continue to decrease (Krumholz 2012). Based on the equations above, we would expect a 40% reduction in TN concentrations to result in a 2.2 meter increase of an initial Secchi depth of 2.4 meters to 4.6 meters. (Specific increases will vary by initial Secchi depth because of the inverse relationship.) A reduction in TN concentrations would be equivalent to an increase in salinity of 16.6 to 19 percent in our model; we used a conservative estimate of 16.6%.

Prediction of seagrass presence requires that we choose a threshold probability level. Many modelers choose a default level of 50% (odds ratio of 1) as an indicator of the minimum level at which seagrass is expected to exist. Depending on the relative cost of false positive versus false negative errors, users may choose alternative thresholds (Fielding and Bell 1997). Van derHeide et al. (2007) estimated the minimum density of seagrass required for positive feedback effects promoting sediment stability. For the Wadden Sea this corresponded to a threshold value of 30% maximum biomass. We present estimated changes in seagrass coverage based on a range of alternative thresholds.

4.5 Model predictions of seagrass increase with decreased nitrogen loading

There are physical constraints in the amount of potential seagrass habitat in Narragansett Bay. Based on the current model, the colonized area for all shorelines combined following a 40% reduction in TN loads (and concentration) would increase from 12% of area in the 0 to 5 meter depth zone to about 63% of area in the short term and slightly more in the long term (as sediment organic carbon levels recovered) (Figure 23a). Given a threshold for

presence/absence of 0.5, we predict recovery potential in the short term to differ substantially among shorelines (Figure 23b). Long term recovery (assuming return of sediment organic carbon to reference levels) is estimated to be virtually complete for the majority of the most favorable shorelines, but negligible for Shoreline 17 (yellow lines, Figure 23c). These sensitivity analyses are approximate given that background disturbance levels might change over time and that we haven't factored in the positive feedbacks in sediment stabilization due to seagrass growth. It is possible that specific restoration measures such as co-restoration of shellfish beds (to reduce suspended sediments and effects of tidal currents and wave action) or use of existing or constructed coastal barriers to limit effects of wave action and tides would improve probabilities of initial colonization success and the initiation of positive feedback effects (Bos and van Katwijk 2007). Thus, we can use our model to project ecologically significant increases in seagrass coverage following planned reductions in total N loading based on effects of nitrogen on light availability mediated by phytoplankton and periphyton growth.

4.6 Future improvements

Although our prediction of seagrass absolute or average presence/absence at shoreline locations was very robust based on our current model, our models could be improved in the future with the incorporation of finer resolution data and/or more sophisticated modeling approaches. Accuracy of depth limits could probably be improved with better resolution of digital elevation models (DEMs). NOAA's National Geophysical Data Center (NGDC) is building high-resolution Tsunami Inundation DEMs of select U.S. coastal regions. These DEMs are referenced to a vertical tidal datum of NAVD 88 or Mean High Water (MHW) and horizontal datum of World Geodetic System of 1984 (WGS 84). Cell sizes will range from 1/3 arc-second (~10 meters) to 36 arc-seconds (~1 km) (http://www.ngdc.noaa.gov/mgg/inundation/tsunami/inundation.html). Although not an issue for Narragansett Bay, the improved resolution between the shoreline (mean sea level, MSL) and MLLW (currently set to zero in merged topobathymetry grids) will be particularly important for predicting distributions of intertidal populations. In the near future, we will have access to maps of predicted tidal currents throughout Narragansett Bay as part of a recently developed hydrodynamic model (Abdelrhman 2015). An accompanying water quality model will also include finer scale predictions of nitrogen concentrations throughout the estuary, so that we will not have to rely on salinity as an indicator of total nitrogen gradients (US EPA 2015).

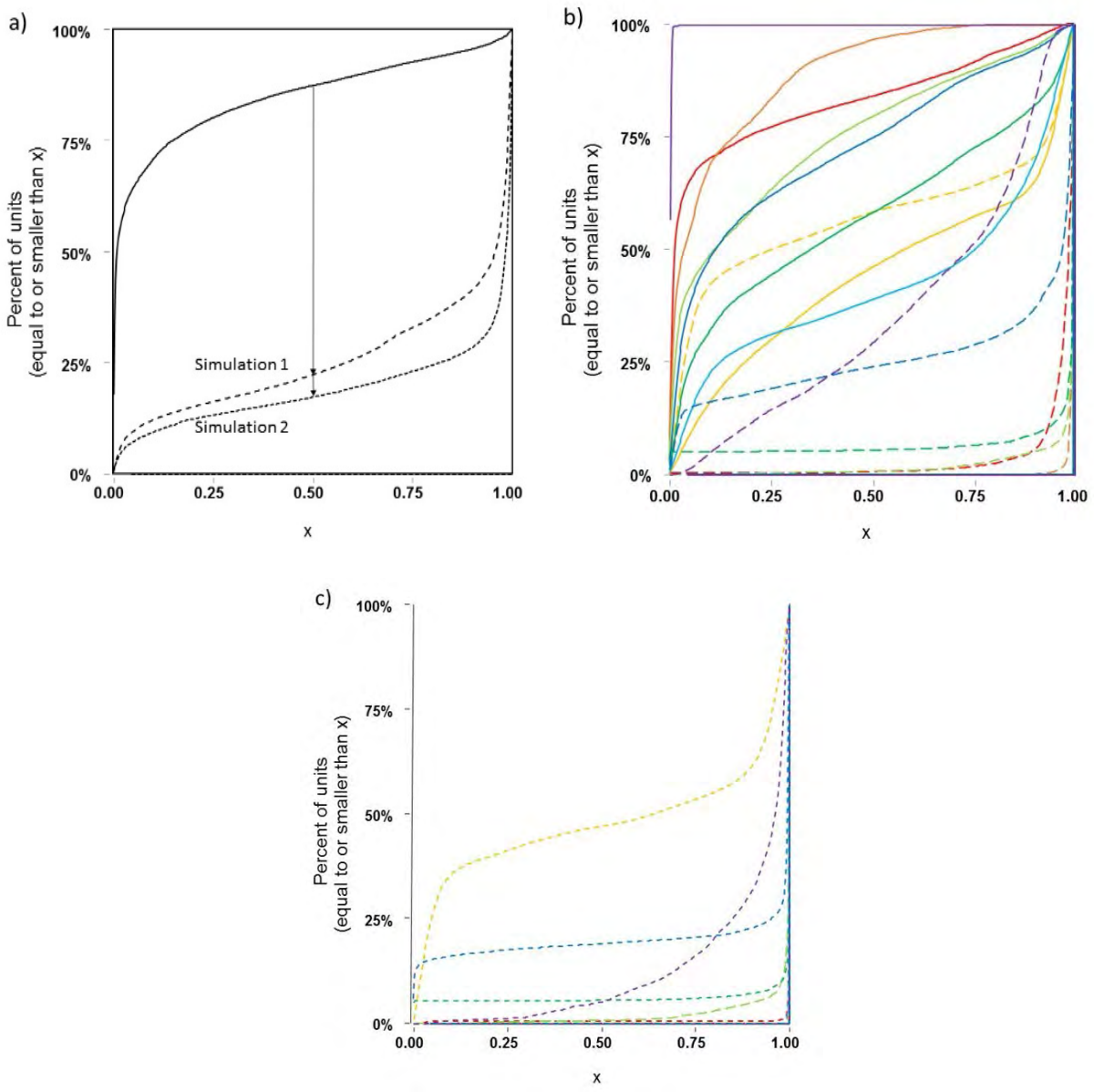

Figure 23. Projected change in cumulative distribution function for probability of occurrence (x) following 40% reduction in total N without (simulation 1) and with (simulation 2) sediment recovery. a) All shorelines combined, b) Projected short-term effect of 40% TN reduction for 8 most favorable shorelines (favorability in order of red – orange – yellow – light green – dark green – light blue – dark blue – indigo) with solid lines indicating condition before TN reductions and dashed lines indicating projected condition after TN reductions, c) Long term recovery following 40% reduction in total N for 8 most favorable shorelines assuming recovery of sediment percent organic carbon to reference levels.

A bio-optical model is also under development for Narragansett Bay (US EPA 2015), which could improve model inputs describing light availability. However, the biggest limitation to model predictions of light availability to seagrass is the lack of information on the resuspension and transport of fine particulates in the nearshore zone. Most measurements of light attenuation (either profiles or Secchi depth measurements) have been made in deeper waters along the main axis of the estuary. Given that the Secchi depth is less than

the depth of highest turbidity associated with resuspended sediments near the seabed[4], S.D. values will represent only background light attenuation and may be of limited use for predicting light availability in seagrass beds or within potential seagrass habitat prior to recovery. Time series of light measurements from underwater HOBOs in areas of potential seagrass habitat both inside and outside of existing seagrass beds are needed, similar to recordings made for Massachusetts coastal bays (Kenworthy et al. 2014).

We would have had access to more methods for incorporating spatial autocorrelation into predictive models if we were not constrained by memory requirements. It is possible that use of parallel computing methods (e.g., simultaneous use of multiple CPUs in an existing quad core processor) through application of R packages such as snow (Tierney et al. 2014) would alleviate this issue. However, we would still have to deal with the limitation of existing packages for methods such as regression kriging in handling anisotropy in spatial autocorrelation.

We could not include all potential interaction terms in our models because of the high variance inflation factors generated. Boosted regression trees (BRTs) are a robust approach for handling models with nonlinear effects and multiple interactions. Of all those approaches tested, Valle et al. (2013) found the best performance for BRT models. Bayesian modeling approaches also could be useful in the future to facilitate adaptive management of habitat (March et al. 2013). Ultimately, managers may need to develop coupled hydrodynamic/water quality/sediment diagenesis models with feedback effects to capture the hysteresis and lags inherent in eelgrass decline and recovery (Eldridge and Morse 2000, van derHeide et al. 2007, delBarrio et al. 2014, Kenworthy et al. 2014).

Our model, like any model, could potentially be improved with additional data or enhanced analyses. Nonetheless, we believe that our predictions are sufficiently robust to support decision-making in an adaptive management framework. Potential applications of our model include identification of ALU zones for setting nutrient criteria for areas of potential seagrass habitat, prioritization of areas and strategies for seagrass restoration, and projection of potential benefits of management actions. Even with the potential for model improvements discussed above, our model represents a significant advancement over existing models which focus on a limited set of factors influencing seagrass growth, fail to incorporate and correct for spatial autocorrelation, and which fail to incorporate interaction of environmental variables or nonlinear effects.

[4] This turbidity maximum is based on depth profiles, and should not be confused with the "turbidity maximum" that occurs in the upper estuary reaches of some estuaries.

Chapter 5. Conclusions

Our pilot project addressed the multiple challenges associated with development of species distribution models for seagrass: the fine-scale patchiness of seagrass distributions with attendant problems of spatial autocorrelation, the large areas of interest for model development and application entailing significant memory demands for modeling, and the potential co-variance of multiple interacting factors affecting seagrass. The fine-scale of data necessary to describe the patchy nature of seagrass distributions coupled with the large model application area for management decisions create significant memory demands which can limit the number and type of R packages that can be applied in practice. However, we found that addition of a residual autocorrelation term in logistic regression models, as suggested by Crase et al. (2013) virtually eliminated the presence of spatial autocorrelation in model residuals. While Crase et al. limited their residual autocorrelation term to zonal averages calculated among adjacent grid cells, we expanded the focal average to cover the range of spatial autocorrelation evident from correlograms. The use of a rectangular zonal average allowed us to account for anisotropy in spatial errors, with a range of up to 1320 meters in the axis parallel to the shoreline but only 200 meters in the offshore direction. As predicted, incorporation of a spatial autocorrelation term in regression models reduced the number of significant variables included, generally reducing the number of distinct sediment type responses and/or interactions detected.

We were able to deal with most but not all of the issues related to cross-correlation of potential explanatory variables. Centering variables prior to incorporating them into models not only kept variance inflation factors values low but also aided in interpretation of model coefficients. Both light availability and wave energy co-vary with depth, and both were significant explanatory variables. We were able to include both types of variables in our models by expressing light availability effects based on the difference between Secchi depth and seagrass bed depth and representing the effect of wave mixing depth as a binary variable (depth > wave mixing depth). Gradients that co-vary with salinity along the main axis of the estuary are more problematical. Although temperature is potentially important in affecting seagrass populations, we excluded it because the populations we were dealing with were subtidal and the range of temperatures measured in our well-flushed system was lower than those at which effects had been observed. We did include salinity which covaries with nutrient concentrations along the axis of the estuary, but given the low range of values, it is probably serving as a proxy for nutrient concentration effects not captured by changes in transparency related to phytoplankton biomass. However, the co-varying gradients of salinity and nutrients would make it difficult to model interactions of nutrients and salinity, although these have been demonstrated experimentally (van Katwijk et al. 1999).

We predicted seagrass distribution at the scale of 10 x 10-meter grid cells, as presence/absence or average presence/absence associated with shoreline locations spaced at 10-meter intervals, and minimum or maximum depth of distributions at those locations. Prediction of seagrass absolute or average presence/absence at shoreline locations was

very robust, with area-under-the-curve (AUC) values associated with Receiver Operating Characteristic (ROC) curves of 0.95 – 0.98 following 10-fold cross-validation of models. Although the model predicting shoreline presence was the most robust of those tested across different scales in Narragansett Bay, other scales of resolution might work better for other types of estuaries. In Narragansett Bay, random shoreline effects varied over several orders of magnitude, probably tied to the distribution of tidal currents which are weak enough to allow persistence of existing seagrass beds, but strong enough to resuspend fine sediments that interfere with successful recolonization. For the model predicting seagrass presence/absence at the grid cell scale, the most influential predictor of fixed effects is Secchi depth, followed by, in order: shoreline isolation, sediment percent total organic carbon, sediment type, and salinity. The least influential variable is water depth greater than average wave mixing depth. For the model predicting presence of seagrass at shoreline locations, the most influential predictor is sediment type, followed by sediment percent total organic carbon (at low Secchi depth), then salinity (as a proxy for water column total nitrogen).

Multiple modes of action for nutrients can be simultaneously incorporated into empirical models for seagrass distribution. We were able to capture the effects of nutrients on light availability, other nutrient impacts potentially related to stimulation of periphyton and macro-algal growth (using salinity as a proxy for total N), and longer term impacts related to the accumulation of organic matter in the sediments. The latter was reflected in a significant interaction between the light compensation point (maximum depth) for seagrass and sediment percent organic carbon.

Application of our model in predictive mode suggests that different shorelines in Narragansett Bay may have very different recovery potentials, and that interventions to reduce tidal energy and/or sediment resuspension may be needed as part of restoration activities. Longer-term recovery potential, related to recovery of sediment organic carbon to levels characteristic of reference condition in New England estuaries, is greater than predicted short-term recovery potential related to reductions in total nitrogen concentrations in the water column and increased transparency related to reductions in phytoplankton biomass.

Many of the georeferenced data layers required to parameterized species distribution models of this type are publically available online through EPA's Estuary Data Mapper application (Detenbeck et al. 2009; www.epa.gov/edm), although spatial resolution of some remotely sensed indicators of water quality for estuaries are limited. In our application, we made substitutions in some cases, based on spatially intensive monitoring that had been carried out in Narragansett Bay. Publically available monitoring data for nutrients are limited to several stations, which tend to be located along the main axis of the bay, are extremely limited in the Sakonnet River arm of the system, and are nonexistent for shallow waters of the bay. Tidal current measurement data are limited for most estuaries and predicted values have only been made publically available online in beta form by NOAA very recently. Finer-scale predictions yielded by future publically available models may be critical in identifying the best protected locations for eelgrass restoration. Our model could be improved by better characterization of optical parameters, particularly in the near-

benthic environment and in the shallow waters in which seagrass is typically found. Long-term records of Secchi depth do not capture trends in the resuspension of fine sediments in the bottom waters which may be limiting seagrass recolonization.

In spite of the methodological challenges and data limitations faced in developing a predictive statistical model for seagrass distribution, our model performed well and was particularly robust for predictions of shoreline occurrence. Our models were also successful in elucidating the effect of multiple interacting stressors in determining the distribution of seagrass patches in Narragansett Bay, as well as factors which might limit recovery. Finally, we were able to demonstrate multiple pathways for nutrient effects on seagrass growth and survival in the bay, including long-term effects of eutrophication on sediment organic carbon which could persist for decades and slow final recovery.

References

Abdelrhman, M.A. *Three-dimensional Modeling of Hydrodynamics and Transport in Narragansett Bay.* EPA/600/R-15/152. Washington, DC: U.S. Environmental Protection Agency, 2015.

Batiuk, R. A., R. J. Orth, K. A. Moore, W. C. Dennison, J. C. Stevenson, L. W. Staver, V. Carter, N. B. Rybicki, R. E. Hickman, S. Kollar, S. Bieber, and P. Heasly. *Chesapeake Bay Submerged Aquatic Vegetation Habitat Requirements and Restoration Targets: A Technical Synthesis*, CBP/TRS 83/92. Annapolis, MD: U.S. Environmental Protection Agency, 1992.

Batiuk, R., P. Bergstrom, M. Kemp, E. Koch, L. Murray, C. Stevenson, R. Bartleson, V. Carter, N. Rybicki, J. Landwehr, C. Gallegos, L. Karrh, M. Naylor, D. Wilcox, K. Moore, S. Ailstock, and M. Teichberg. *Chesapeake Bay Submerged Aquatic Vegetation Water Quality and Habitat-Based Requirements and Restoration Targets: A Second Technical Synthesis*, CBP/TRS 245/00. EPA/903/R-00/014. Annapolis, MD: U.S. Environmental Protection Agency, Chesapeake Bay Program, 2000. http://archive.chesapeakebay.net/pubs/sav/index.html.

Benson, J. L., D. Schlenzinger, and B.L. Howes. Relationship between nitrogen concentration, light, and *Zostera marina* habitat quality and survival in southeastern Massachusetts estuaries. *Journal of Environmental Management* 131: 129-137 (2013).

Borkman, D. G., and T. J. Smayda. Long-term trends in water clarity revealed by Secchi-disk measurements in lower Narragansett Bay. *ICES Journal of Marine Science* 55: 668–679 (1988).

Bos, A. R., and M. M. van Katwijk. Planting density, hydrodynamic exposure and mussel beds affect survival of transplanted intertidal eelgrass. *Marine Ecology Progress Series* 336: 121-129 (2007).

Bradley, M., K. Raposa, and S. Tuxbury. *Report on the Analysis of True-color Aerial Photography to Map and Inventory Zostera marina L. in Narragansett Bay and Block Island, Rhode Island*. Page 1-16 and 9 Mapsheets. Kington, RI: Rhode Island Natural History Survey, 2007.

Burkholder, J. M., D. A. Tomasco, and B. W. Touchette. Seagrasses and eutrophication. *Journal of Experimental Marine Biology and Ecology* 350: 46–72 (2007).

Burnham, K. P., and D. R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd edition. New York, NY: Springer, 2002.

Chambers, P. A. Nearshore occurrence of submersed aquatic macrophytes in relation to wave action. *Canadian Journal of Fisheries and Aquatic Sciences* 44: 1666-1669 (1987).

Collins, B.P. Suspended material transport: Narragansett Bay area, Rhode Island. *Estuarine and Coastal Marine Science* 4: 33-44 (1976).

Compton, J. E., J. Harrison, R. L. Dennis, T. L. Greaver, B. H. Hill, S. J. Jordan, H. Walker, and H. V. Campbell. Ecosystem services altered by human changes in the nitrogen cycle: A new perspective for US decision making. *Ecology Letters* 14: 804-815 (2011).

Costa, J. E. Distribution, production, and historical changes in abundance of eelgrass *(Zostera marina* l.) in southeastern Massachusetts. Ph.D. thesis, Boston University, Boston, MA, 1988. 396 pp.

Costa, J. E., B. L. Howes, and E. Gunn. *Report of the Buzzards Bay Citizens' Water Quality Monitoring Program 1992–1995*. New Bedford, MA: Coalition for Buzzards Bay, 1996.

Crase, B., A. C. Liedloff, and B. A. Wintle. A new method for dealing with residual spatial autocorrelation in species distribution models. *Ecography* 35: 879-888 (2012).

De Boer, W. F. Seagrass–sediment interactions, positive feedbacks and critical thresholds for occurrence: A review. *Hydrobiologia* 591: 5–24 (2007).

DeJonge, V. N., D. J. de Jong, and M. M. van Katwjik. Policy plans and management measures to restore eelgrass (*Zostera marina* L.) in the Dutch Wadden Sea. *Helgoland Marine Research* 54: 151-158 (2000).

Del Barrio, P., N. K. Ganju, A. L. Aretxabaleta, M. Hayn, A. Garcia, and R. W. Howarth. Modeling future scenarios of light attenuation and potential seagrass success in a eutrophic estuary. *Estuarine, Coastal and Shelf Science*, 149: 13-23 (2014).

denHartog, C. *The Seagrasses of the World*. Amsterdam: North-Holland Publication, 1970. 275 pp.

Dennison, W. C. Effects of light on seagrass photosynthesis, growth and depth distribution. *Aquatic Botany* 27: 15-26 (1987).

Dennison, W. C., R. J. Orth, K. A. Moore, J. C. Stevenson, V. Carter, S. Kollar, P. W. Bergstrom, and R. A. Batiuk. Assessing water quality with submersed aquatic vegetation. *Bioscience* 43: 86–94 (1993).

Detenbeck, N., M. Pelletier, M. ten Brink, M. Abdelrhman, and S. Rego. "E-Estuary: Developing a decision-support system for coastal management in the conterminous United States." In *Proceedings, 33rd IAHR 2009 Congress - Water Engineering for a Sustainable Environment*. Vancouver, BC, 2009. Pp. 5284-5292.

Downie, A-L., M. von Numers, and C. Boström. Influence of model selection on the predicted distribution of the seagrass *Zostera marina*. *Estuarine, Coastal and Shelf Science* 121/122: 8-19 (2013).

Duarte, C. M. Seagrass depth limits. *Aquatic Botany* 40: 363-377 (1991).

Eldridge, P. M., and J. W. Morse. A diagenetic model for sediment-seagrass interactions. *Marine Chemistry* 70: 89-103 (2000).

Fielding, A. H. and J. F. Bell. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24: 38-49 (1997).

Fonseca, M. S., and A. Malhotra. *WEMo (Wave Exposure Model) User Manual, Version 4.0*. Beaufort, NC: NOAA National Centers for Coastal Ocean Science, Center for Coastal Fisheries and Habitat Research. March 2010.

Fonseca M. S., W. J. Kenworthy, and G. W. Thayer. *Guidelines for Conservation and Restoration of Seagrass in the United States and Adjacent Waters*. NOAA/NMFS Coastal Ocean Program Decision Analysis Series, No. 12. Silver Spring, MD: NOAA Coastal Ocean Office, 1998.

Fonseca, M. S., P. E. Whitfield, W. J. Kenworthy, D. R. Colby, and B. E. Julius. Use of two spatially explicit models to determine the effect of injury geometry on natural resource recovery. *Aquatic Conservation: Marine and Freshwater Ecosystems* 14: 281-98 (2004).

Garcia, E., T. C. Granata, and C. M. Duarte. An approach to measurement of particle flux and sediment retention within seagrass (Posidonia oceanica) meadows. *Aquatic Botany* 65: 255-268 (1999).

Goodman, J. L., K. A. Moore, and W. C. Dennison. Photosynthetic responses of eelgrass (*Zostera marina* L.) to light and sediment sulfide in a shallow barrier island lagoon. *Aquatic Botany* 50: 37-47 (1995).

Grech, A., and R. G. Coles. An ecosystem-scale predictive model of coastal seagrass distribution. *Aquatic Conservation: Marine and Freshwater Ecosystems* 20: 437-444 (2010).

Hammar-Klose, and E. R. Thieler. *Coastal Vulnerability to Sea-Level Rise: A Preliminary Database for the U.S. Atlantic, Pacific, and Gulf of Mexico Coasts.* U.S. Geological Survey, Digital Data Series DDS-68, 1 CD-ROM, 2001.

Harlin, M. M., and B. Thorne-Miller. Nutrient enrichment of seagrass beds in a Rhode Island coastal lagoon. *Marine Biology* 65: 221-229 (1981).

Holmer, M. and E. J. Bondgaard. Photosynthetic and growth response of eelgrass to low oxygen and high sulfide concentrations during hypoxic events. *Aquatic Botany* 70: 29-38 (2001).

Howarth, R. W. Nutrient limitation of net primary production in marine ecosystems. *Annual Review of Ecology and Systematics* 19: 89-110 (1988).

Howes, D., M. Morris, and M. Zacharias. *British Columbia Estuary Mapping System. Version 1.* Province of British Columbia, Canada: Land Use Coordination Office for the Coastal Task Force, Resource Inventory Committee Resources Inventory Committee, 1999. Downloaded from: http://www.for.gov.bc.ca/ric.

IEC. “Appendices”, In *Narragansett Bay Sustainability Pilot: Phase I Report,* 2012. http://www.epa.gov/research/docs/nbsp-phase-I-report-appendices.pdf.

Jarvis, J. C., and K. A. Moore. The role of seedlings and seed bank viability in the recovery of Chesapeake Bay, USA, *Zostera marina* populations following a large-scale decline. *Hydrobiologia* 649: 55-68 (2010).

Keddy, C. J., and D. G. Patriquin. An annual form of eelgrass in Nova Scotia. *Aquatic Botany* 5: 163-170 (1978).

Keith, D., B. Schaeffer, R. Lunetta, R. Gould, Jr., K. Rocha, and D. Cobb. Remote sensing of selected water-quality indicators with the Hyperspectral Imager for the Coastal Ocean (HICO) Sensor. *International Journal of Remote Sensing* 35: 2927-2962 (2014).

Kelly, N. M., M. Fonseca, and P. Whitfield. Predictive mapping for management and conservation of seagrass beds in North Carolina. *Aquatic Conservation: Marine and Freshwater Ecosystems* 11: 437-451 (2001).

Kendrick, G. A., C. M. Duarte, and N. Marba. Clonality in seagrasses, emergent properties and seagrass landscapes. *Marine Ecology Progress Series* 290: 291-6 (2005).

Kenworthy, W. J., and M. Fonseca. Reciprocal transplant of the seagrass *Zostera marina* L. Effect of substrate on growth. *Aquaculture* 12: 197–213 (1977).

Kenworthy, W. J., C. L. Gallegos, C. Costello, D. Field, and G. di Carlo. Dependence of eelgrass (*Zostera marina*) light requirements on sediment organic matter in Massachusetts coastal bays: Implications for remediation and restoration. *Marine Pollution Bulletin* 83: 446-457 (2014).

Koch, E. W. Beyond light: Physical, geological, and geochemical parameters as possible submersed aquatic vegetation habitat requirements. *Estuaries* 24: 1-17 (2001).

Koch, M. S., S. A. Schopmeyer, O. I. Nielsen, C. Kyhn-Hansen, and C. J. Madden. Conceptual model of seagrass die-off in Florida Bay: Links to biogeochemical processes. *Journal of Experimental Marine Biology and Ecology* 350: 73–88 (2007).

Kopp, B .S. 1848-1994 *Historical Eelgrass Beds of Narragansett Bay,* 1995. Downloaded from www.edc.uri.edu/eelgrass.

Kopp, B. S., A. M. Doherty, and S. W. Nixon. *A Guide to Site Selection for Eelgrass Restoration Projects In Narragansett Bay, Rhode Island. Technical Report to the Narragansett Bay Project.* Narragansett, RI: University of Rhode Island, 1994.

Krause-Jensen, D., M. F. Pedersen, and C. Jensen. Regulation of eelgrass (*Zostera marina*) cover along depth gradients in Danish coastal waters. *Estuaries* 26: 866-977 (2003).

Krumholz, J.S. Spatial and temporal patterns in nutrient standing stock and mass-balance in response to load reductions in a temperate estuary. Ph.D. Thesis, University of Rhode Island, 2012.

Latimer, J. S. and S. A. Rego. Empirical relationship between eelgrass extent and predicted watershed-derived nitrogen loading for shallow New England estuaries. *Estuarine, Coastal and Shelf Science* 90: 231-240 (2010).

Lee, K., S. R. Park, and Y. K. Kim. Effects of irradiance, temperature, and nutrients on growth dynamics of seagrasses: A review. *Journal of Experimental Marine Biology and Ecology* 350: 144–175 (2007).

Maggini, R., A. Lehmann, N. E. Zimmermann, and A. Guisan. Improving the generalized regression analysis for the spatial prediction of forest communities. *Journal of Biogeography* 33: 1729-1749 (2010).

Marba, N. and C. M. Duarte. Rhizome elongation and seagrass clonal growth. *Marine Ecology Progress Series* 174: 269-280 (1998).

March, D., J. Alos, M. Cabanellas-Reboredo, E. Infantes, A. Jordi, and M. Palmer. A Bayesian spatial approach for predicting seagrass occurrence. *Estuarine, Coastal and Shelf Science* 131: 206-12 (2013).

McMaster, R. L. *Sediment Types, Narragansett Bay Sediments; A map of Narragansett Bay sediments* from McMaster, 1960, Published in 1993, 1:24000 (1in=2000ft) scale, State of Rhode Island and Providence Plantations, 1960. Downloaded from http://www.narrbay.org/d_downloads/D_Biological/D_habitat/sediment.htm#6.

Murphey P. L., and M. S. Fonseca. Role of high and low energy seagrass beds as nursery areas for *Penaeus duorarum* in North Carolina. *Marine Ecology Progress Series* 121: 91–98 (2005).

Neckles, H. A., F. T. Short, S. Barker, and B. S. Kopp. Disturbance of eelgrass *Zostera marina* by commercial mussel *Mytilus edulis* harvesting in Maine: Dragging impacts and habitat recovery. *Marine Ecology Progress Series* 285: 57-73 (2005).

Nelson, W.G. (ed.) *Seagrasses and Protective Criteria: A Review and Assessment of Research Status,* EPA/600/R-09/050. Washington, DC: U.S. Environmental Protection Agency, 2009.

Nielsen, S. L., J. Borum, O. Geertz-Hansen, and J. Sand-Jensen. Marine bundplanters dybdegraense. *Vand Miljo* 5: 217-220 (1989).

Nixon, S. W., and M. E. Pilson. "Nitrogen in estuarine and coastal marine systems." In *Nitrogen in the Marine Environment*, eds. E. J. Carpenter, and D. G. Capone, 565-648. New York: Academic Press, 1983.

Oakley, B. A., J. D. Alvarez., and J. C. Boothroyd. Benthic geologic habitats of shallow estuarine environments: Greenwich Bay and Wickford Harbor, Narragansett Bay, Rhode Island, U.S.A. *Journal of Coastal Research* 28: 760-773 (2012).

Oczkowski, A., S. Nixon, K. Henry, P. DiMilla, M. Pilson, S. Granger, B. Buckley, C. Thornber, R. McKinney, and J. Chaves. Distribution and trophic importance of anthropogenic nitrogen in Narragansett Bay: An assessment using stable isotopes. *Estuaries and Coasts* 31: 53-69 (2008).

Orth, R. J., M. Luckenbach, and K. A. Moore. Seed dispersal in a marine macrophyte: Implications for colonization and restoration. *Ecology* 75: 1297–1939 (1994).

Orth, R. J., T. J. B Carruthers, W. C. Dennison, C. M. Duarte, J. W. Fourqurean, K. L. Heck, Jr., A. R. Hughes, G. A. Kendrick, W. J. Kenworthy, S. Olyarnik, F. T. Short, M. Waycott, and S. L. Williams. A global crisis for seagrass ecosystems. *Bioscience* 56: 987 – 996 (2006).

Oviatt, C. A., and S. W. Nixon. Sediment resuspension and deposition in Narragansett Bay. *Estuarine and Coastal Marine Science* 3: 201-217 (1975).

Oviatt, C., P. Doering, B. Nowicki, L. Reed, J. Cole, and J. Frithsen. An ecosystem level experiment on nutrient limitation in temperate coastal marine environments. *Marine Ecology Progress Series* 116: 171-179 (1995).

Pelletier, M. C., D. E. Campbell, K. T. Ho, R. M. Burgess, C. T. Audette, and N. E. Detenbeck. Can sediment total organic carbon and grain size be used to diagnose organic enrichment in estuaries? *Environmental Toxicology and Chemistry* 30: 538-547 (2011).

Pilson, M. E. Q. On the residence time of water in Narragansett Bay. *Estuaries* 8: 2–14 (1985).

Poole, H.H. and W.R.G. Atkins. Photo-electric measurements of the penetration of light into sea water. *Journal of the Marine Biological Association U.K.* 16: 297 (1929).

Poppe, L. J., V. F. Paskevich, S. J. Williams, M. E. Hastings, J. T. Kelley, D. F. Belknap, L. G. Ward, D. M. FitzGerald, and P. F. Larsen. *Surficial Sediment Data from the Gulf of Maine, Georges Bank, and Vicinity: A GIS Compilation,* U.S. Geological Survey Open-File Report 03-001. Woods Hole, MA: U.S. Geological Survey Coastal and Marine Geology Program, Woods Hole Field Center, 2003.

R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2014. http://www.R-project.org/.

Ralph, P. J., M. J. Durako, S. Enríquez, C. J. Collier, and M. A. Doblin. Impact of light limitation on seagrasses. *Journal of Experimental Marine Biology and Ecology* 350: 176–193 (2007).

RIGIS. Marinas of Rhode Island: s44fsm96. Rhode Island Geographic Information System Database, 1996 (s44fsm96 accessed from http://www.edc.uri.edu/rigis/data/data.aspx?ISO=structure August 2014).

RIGIS. Eelgrass Beds in Rhode Island (Polygon): s44nep99. Rhode Island Geographic Information System Database, 2000a (s44nep99 accessed from http://www.edc.uri.edu/rigis/data/data.aspx?ISO=biota August 2014).

RIGIS. Eelgrass Beds in Rhode Island (Line): s44nel99. Rhode Island Geographic Information System Database, 2000b (s44nel99 accessed from http://www.edc.uri.edu/rigis/data/data.aspx?ISO=biota August 2014).

RIGIS. Hardened Shorelines in Narragansett Bay: hrdshore. RI Dept. of Environmental Management, Narragansett Bay Estuary Program, 2003. (hrdshore accessed from http://www.edc.uri.edu/rigis/data/data.aspx?ISO=structure August 2014).

RIGIS. Submerged Aquatic Vegetation (SAV) in Rhode Island Coastal Waters (2006); SAV06. Kingston, RI: Rhode Island Geographic Information System (RIGIS) Data Distribution System, 2013, Kingston, Rhode Island: Environmental Data Center, University of Rhode Island (http://www.edc.uri.edu/rigis/data/data.aspx?ISO=biota, last date accessed: 1 May 2013).

Rivers, D. O., and F. T. Short. Effect of grazing by Canada geese Branta canadensis on an intertidal eelgrass *Zostera marina* meadow. *Marine Ecology Progress Series* 333: 271-279 (2007).

Roberts J. J., B. D. Best, D. C. Dunn, E. A. Treml, and P. N. Halpin. Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. *Environmental Modelling & Software* 25: 1197-1207 (2010).

Rohweder, J., J. T. Rogala, B. L. Johnson, D. Anderson, S. Clark, F. Chamberlin, D. Potter, and K. Runyon. *Application of Wind Fetch and Wave Models for Habitat Rehabilitation and Enhancement Projects–2012 Update*. Contract report prepared for U.S. Army Corps Engineers' Upper Mississippi River Restoration – Environmental Management Program, 2012. 52p.

Schonert, T., and P. Milbradt. Hydro- and morphodynamic simulation considering ecological model components on asynchronous parallel cellular automata. *Computing in Civil Engineering* 2005: 1-11 (2005).

Shipman, H. "The geomorphic setting of Puget Sound: Implications for shoreline erosion and the impacts of erosion control structures" In *Puget Sound Shorelines and the Impacts of Armoring—Proceedings of a State of the Science Workshop,* U.S. Geological Survey Scientific Investigations Report 2010-5254, by Shipman, H., M. N. Dethier, G. Gelfenbaum, K. L. Fresh, and R. S. Dinicola (eds.), May 2009, 2010. p. 19-34.

Short, F. T. Effects of sediment nutrients on seagrasses, literature review and mesocosm experiment. *Aquatic Botany* 27: 41–57 (1987).

Short, F. T., L. K. Muehlstein, and D. Porter. Eelgrass wasting disease: Cause and recurrence of a marine epidemic. *Biological Bulletin* 173: 557-62 (1987).

Short, F. T., B. W. Ibelings, and C. Den Hartog. Comparison of a current eelgrass disease to the wasting disease of the 1930's. *Aquatic Botany* 30: 295-304 (1988).

Short, F. T., D. Burdick, J. Wolf, and G. E. Jones. *Eelgrass in Estuarine Research Reserves Along the East Coast, USA. Part I: Declines from Pollution and Disease and Part II: Management of Eelgrass Meadows.* Durham, NH: NOAA, Coastal Ocean Program Publ., 1993.

Short, F. T. *The Port of New Hampshire Interim Mitigation Success Assessment Report*. Report to the New Hampshire Department of Transportation. Durham, NH: Jackson Estuarine Laboratory, 1993.

Short, F. T., and D. M. Burdick. *Eelgrass Restoration Site Selection Model V.1.2*. CDROM, 2005.

Smayda, T. J., and D. G. Borkman. "Nutrient and plankton dynamics in Narragansett Bay", Ch. 15 In *Science for Ecosystem-based Management,* By Desbonnet, A. and B. A. Costa-Pierce (eds.),. New York, NY: Springer, 2008.

Teichberg, M., S. E. Fox, Y. S. Olsenz, I. Valielaw, P. Martinetto, O. Iribarnes, E. Y. Muto, M. A. V. Petti, T. N. Corbisier, M. Soto-Jimenez, F. Paez-Osuna, P. Castro, H. Freitas, A. Zitelli, M. Cardinaletti, and D. Tagliapietra. Eutrophication and macroalgal blooms in temperate and tropical coastal waters: nutrient enrichment experiments with Ulva spp. *Global Change Biology* 16: 2624–2637 (2010).

Terrados, J., C. M. Duarte, L. Kamp-Nielsen, N. S. R. Agawin, E. Gacia, D. Lacap, M. D. Fortes, J. Borum, M. Lubanski, and T. Greve. Are seagrass growth and survival constrained by the reducing conditions of the sediment? *Aquatic Botany* 65: 175-197 (1999).

Thorne-Miller, B., M. M. Harlin, G. B. Thursby, M. M. Brady-Campbell, and B. A. Dworetzky. Variations in the distribution and biomass of submerged macrophytes in five coastal lagoons in Rhode Island, USA. *Botanica Marina* XXVI:231-242 (1983).

Touchette, B. W., and J. M. Burkholder. Seasonal variations in carbon and nitrogen constituents in eelgrass (*Zostera marina* L.) as influenced by increased temperature and water column nitrate. *Botanica Marina* 45: 23-34 (2002).

Touchette, B.W. Seagrass-salinity interactions: Physiological mechanisms used by submersed marine angiosperms for a life at sea. *Journal of Experimental Marine Biology and Ecology* 350: 194–215 (2007).

U.S. Environmental Protection Agency. *Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries*. EPA/903/R/03/002. Annapolis, MD: U.S. Environmental Protection Agency, Region 3, Chesapeake Bay Program Office, 2003.

U.S. Environmental Protection Agency. *Quantitative Models Describing Past and Current Nutrient Fluxes and Associated Ecosystem Level Responses in the Narragansett Bay Ecosystem*. EPA/600/R-15/174. Washington, DC: U.S. Environmental Protection Agency, 2015.

Valiela, I., S. Mazzilli, J. Bowen, K. Kroeger, M. Cole, G. Tomasky, and T. Isaji. ELM, an estuarine nitrogen loading model: Formulation and verification of predicted concentrations of dissolved inorganic nitrogen. *Water Air and Soil Pollution* 157: 365-391 (2004).

Valle, M., M. M. van Katwijk, D. J. de Jong, T. J. Bouma, A. M. Schipper, G. Chust, B. M. Benito, J. M. Garmendia, and Á. Borja. Comparing the performance of species distribution models of *Zostera marina*: Implications for conservation. *Journal of Sea Research* 83: 56-64 (2013).

van derHeide, T., E. H. van Nes, G. W. Geerling, A. J. P. Smolders, T. J. Bouma, and M. M. van Katwijk. Positive feedbacks in seagrass ecosystems: Implications for success in conservation and restoration. *Ecosystems* 10: 1311–1322 (2007)

van der Heide,T., E. T. H. M Peeters, D. C. R Hermus., M. M van Katwijk., J. G. M Roelofs, and A. J. P. Smolders. Predicting habitat suitability in temperate seagrass ecosystems. *Limnology and Oceanography* 54: 2018–2024 (2009).

van Katwijk, M.M., G. H. W. Schmitz, A. P. Gasseling, and P.H. van Avesaath. Effects of salinity and nutrient load and their interaction on *Zostera marina*. *Marine Ecology Progress Series* 190: 155-165 (1999).

van Katwijk, M.M., A.R Bos, V. N. de Jonge, L. S. A. M. Hanssen, D. C. R. Hermus, and D.J. de Jong. Guidelines for seagrass restoration: Importance of habitat selection and donor population, spreading of risks, and ecosystem engineering effects. *Marine Pollution Bulletin* 58: 179-188 (2009).

Venables, W. N., and B.D. Ripley, B. D. *Modern Applied Statistics with S*. Fourth Edition. New York: Springer, 2002.

Wazniak, C. E., and M. R. Hall (Eds). *Maryland's Coastal Bays: Ecosystem Health Assessment 2004*, DNR-12-1202-0009. Annapolis, MD: Maryland Department of Natural Resources, Tidewater Ecosystem Assessment, 2005.

Wetzel, R. L., and P. A. Penhale. Production ecology of seagrass communities in the lower Chesapeake Bay. *Marine Technology Society Journal* 17: 22-31 (1983).

Wicks, E. C., E. W. Koch., J. M. O'Neil, and K. Elliston. Effects of sediment organic content and hydrodynamic conditions on the growth and distribution of *Zostera marina. Marine Ecology Progress Series* 378: 71-80 (2009).

Wortmann, J., J. W. Hearne, and J. B. Adams. A mathematical model of an estuarine seagrass. *Ecological Modeling* 98: 137-149 (1997).

Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. M. Smith. *Mixed Effects Models and Extensions in Ecology with R*. New York, NY: Springer, 2009.

R-package references

Andrus, J. M., T. Kuehlhorn, L. F. Rodriguez, A. D. Kent, and J. L. Zilles. CommunityCorrelogram: Ecological Community Correlogram. R package version 1.0, 2014. http://CRAN.R-project.org/package=CommunityCorrelogram.

Bjornstad, Ottar N. ncf: spatial nonparametric covariance functions. R package version 1.1-5, 2013. http://CRAN.R-project.org/package=ncf.

Broström, G. glmmML: Generalized linear models with clustering. R package version 1.0, 2013 http://CRAN.R-project.org/package=glmmML.

Fournier, D. A., H. J. Skaug, J. Ancheta, J. Ianelli, A. Magnusson, M. Maunder, A. Nielsen, and J. Sibert. AD Model Builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods Software* 27: 233-249 (2012).

Fox, J. and J. Hong. Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package. *Journal of Statistical Software* 32: 1-24 (2009).

Signorell, A. Includes R source code and/or documentation previously published by: Anderegg, N., T. Aragon, A. Arppe, B. Bolker, S. Champely, D. Chessel, L. M. D. Chhay, H. C. Doran, S. Dray, C. Dupont, J. Enos, C. Ekstrom, M. Elff, J. Fox, T. Galili, M. Gamer, J. Gross, G/ Grothendieck, F. E. Harrell, Jr., M. Hoehle, C. W. Hoffmann, M. Huerzeler, R. J. Hyndman, V. A. Iglesias, D. Kahle, M. Kohl, M. Korpela, J. Lemon, M. Maechler, A. Magnusson, D. Malter, A. Matei, D. Meyer, Y. Min, M. Naepflin, D. Ogle, S. Pavoine, R. Rapold, W. Revelle, V. E. Seshan, G. Snow, M. Smithson, W. A. Stahel, Y. Tille, A. Trapletti, K. Ushey, J. VanDerWal, B. Venables, J. Verzani, G. R. Warnes, S. Wellek, P. Wolf, and A. Zeileis. DescTools: Tools for descriptive statistics. R package version 0.99.7, 2014. http://CRAN.R-project.org/package=DescTools.

Sing, T., O. Sander, N. Beerenwinkel, and T. Lengauer. 2005. ROCR: Visualizing classifier performance in R. *Bioinformatics* 21: 7881 (2005).

Tierney, L., A. J. Rossini, N. Li, and H. Sevcikova. snow: Simple Network of Workstations. R package version 0.3-13, 2013. http://CRAN.R-project.org/package=snow

Venables, W. N., and B. D. Ripley. *Modern Applied Statistics with S*. Fourth Edition. New York, NY: Springer, 2002.

Wang, Xiao-Feng, and B. Wang. 2011. Deconvolution estimation in measurement error models: The R Package decon. *Journal of Statistical Software* 39: 1-24 (2011).

Wickham, H. ggplot2: Elegant graphics for data analysis. New York, NY: Springer, 2009.

Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* 73: 3-36 (2011).

Wood, S. and F. Scheipl. gamm4: Generalized additive mixed models using mgcv and lme4. R package version 0.2-3, 2014. http://CRAN.R-project.org/package=gamm.

Appendix A. National and Regional Data Sources for Seagrass Habitat Model Development

Table A-1. National and regional data sources for seagrass habitat model development

Parameter	Data Source	Original Source	Source Date	Description	Link
Current Seagrass	Estuary Data Mapper	Rhode Island Geographic Information System (RIGIS) / Bradley, M., K. Raposa, and S. Tuxbury. 2007. Report on the Analysis of True Color Aerial Photography to Map and Inventory Zostera marina L. in Narragansett Bay and Block Island, Rhode Island. Page 1-16 and 9 Mapsheets. Rhode Island Natural History Survey, Kingston, RI.	2006	Data layer of Submerged Aquatic Vegetation (SAV) in Rhode Island Coastal Waters, 2006 from Rhode Island Geographic Information System (RIGIS). SAV presence / absence coded as: Present (1), Absent (0). Grid cell size 10m.	http://www.edc.uri.edu/rigis/data/
Sediments	RI DEM Narragansett Bay Estuary Program	McMaster (1960)	1960	Between 1988 and 1992 the Narragansett Bay Project developed an extensive listing of GIS data layers for applications involving Narragansett Bay. All data were compiled at the University of Rhode Island's Environmental Data Center in 1993 Prior to documenting, all data were reviewed for spatial and topological correctness.	http://www.narrbay.org/biological_data.htm
Temperature	Estuary Data Mapper	MUR SST - Multi scale Ultra high Resolution Sea Surface Temperature (NASA - Jet Propulsion Laboratory - California Institute of Technology)	varies	Used Estuary Data Mapper (EDM) to download daily sea surface temperature (SST) from ftp://podaac-ftp.jpl.nasa.gov/allData/ghrsst/data/L4/GLOB/JPL/MUR/, and developed weekly mean SST data for the years 2003 to 2012. In areas of the estuaries which were not covered by the SST data, a Euclidean distance method was used to fill in areas without data.	http://mur.jpl.nasa.gov/index.php
Salinity	Estuary Data Mapper	R.I. Department of Environmental Management. Bay Assessment Response Team.	varies	Downloaded State of Rhode Island Department of Environmental Management Bay Assessment & Response Team (BART) daily salinity data and developed weekly mean salinity data for stations in Narragansett Bay for the years 2003 to 2012. Thiessen polygons were created from the point data to fill in all areas of Narragansett Bay.	http://www.narrbay.org/d_projects/buoy/buoydata.htm
Bathymetry	Estuary Data Mapper	NOAA Coastal Data Model		Only one source of merged topographic and bathymetric data is currently available in EDM, NOAA's Coastal Data Model (Topography/Bathymetry (NOAA)	www.epa.gov/edm

Parameter	Data Source	Original Source	Source Date	Description	Link
Wave Exposure	WeMO (Wave Exposure Model)	The wave exposure model is a free software modeling tool developed by NOAA to forecast wave energy/exposure along coastal areas. It provides important habitat and erosional information for habitats. This software application was developed by NOAA but is no longer supported.	2003-2006 meteorological data.	Data input for WeMO requires the following: a local meteorological file (containing wind speed and direction), bathymetry covering the area of interest and a shoreline file which covers the boundary of the area of interest. A point file (shape file format) must also be provided to allow the model to perform calculations at specific points of interest. Instructions on processing these data can be found in the WeMO 4.0 manual available at the link provided. WeMo model runs for this project were performed for each estuary using default model values with a modified interrogation distance of (5000 m).	http://www.csc.noaa.gov/ digitalcoast/ tools/wemo
Transparency	Narragansett Bay Commission; US EPA's National Coastal Assessment program; MODIS satellite imagery	Narragansett Bay Commission; US EPA's National Coastal Assessment program; MODIS satellite imagery		Transparency across the bay was described using three data sources. Secchi depths measured by the Narragansett Bay Commission between 2008 and 2012 (http://snapshot.narrabay.com/app/MonitoringInitiatives/WaterClarity) and by the US EPA's National Coastal Assessment program between 2000 and 2006 (http://oaspub.epa.gov/coastal/coast.search) were combined and averaged by Water Body ID (WBID) estuarine segments used for assessment and listing by the Rhode Island DEM. Secchi depths greater than the maximum depth at the site of measurement were removed from the records before averaging. Gaps in transparency data along the southern shore of Conamicutt Island (Sakonnett and Newport Bays) were filled in using Kd estimates from offshore MODIS satellite imagery. The latter were downloaded using the EDM tool, averaged over the growing season (May – October) for nonmissing cells for the years 2008-2012. Average grid cell values were extracted for cells greater than 30 meters in depth, and averaged over a swath of offshore cells parallel to the south shore of Conamicutt Island.	http://snapshot.narrabay.com/app/ MonitoringInitiatives/WaterClarity; http://oaspub.epa.gov/coastal/coast.search; EDM (MODIS)
Organic Carbon	USEPA National Coastal Assessment	USEPA National Coastal Assessment	varies	Total organic carbon was estimated across the Bay based on surficial sediment grabs collected by the US EPA National Coastal Assessment (downloaded from http://oaspub.epa.gov/coastal/coast.search under Sediment Grain Composition Data category). Values were interpolated to create a complete grid within the shallow-water zone using inverse distance weighting in ArcMap 10.1.	http://oaspub.epa.gov/coastal/coast.search

Parameter	Data Source	Original Source	Source Date	Description	Link
Secchi	Secchi (3) data sources for Narragansett Bay; Narragansett Bay fixed monitoring network (RIDEM); Narragansett Bay Commission and MODIS data Kd downloaded from EDM	Secchi (3) data sources for Narragansett Bay; Narragansett Bay fixed monitoring network (RIDEM); Narragansett Bay Commission and MODIS data Kd downloaded from EDM	varies	Secchi depth was estimated from Kd values using an empirical relationship developed by Batiuk et al. (2000).	http://www.dem.ri.gov/bart/netdata.htm ; http://snapshot.narrabay.com/app/ MonitoringInitiatives/WaterClarity; EDM

Appendix B. Tutorial on Finding and Downloading Data for Seagrass Habitat Prediction Models Using EPA's Estuary Data Mapper

B.1. Overview of Estuary Data Mapper

The objective of EPA's Estuary Data Mapper (EDM) project is to produce an easily accessible, standalone software product to automate the retrieval and pre-processing of GIS coverages (including remote sensing data) and associated environmental data (e.g., tidal, hydrologic, and weather time series; water quality and sediment quality data) to populate 1) a GIS data model for estuaries and their watersheds, and 2) tools and models to assess, visualize, diagnose, predict, prioritize, and manage condition of estuaries and coastal watersheds (Detenbeck et al. 2009). The EDM has been designed as a stand-alone application requiring no other specialized software for implementation. EDM is written in C++, OpenGL, and FLTK for extremely fast graphical visualization, user interaction and minimal memory consumption. The interface has been designed with three tabs, the first enabling selection of an area of interest, with background layers such as political boundaries, watersheds, estuaries, and hydrography provided for reference and drop-down boxes to zoom in on states then estuary or watershed. The second tab allows selection of geospatial layers and environmental data time series of interest to be selected for download. This page also allows the user to visualize time series or other data before download. The third tab allows the user to choose among a series of download formats (comma-delimited time series, ASCII grid, shapefile, kmz for Google maps, png or mpg for visualization of images or time series) as well as a location to save the files. Outputs also include a text file (edm_output.txt) containing all WCS calls to rsigserver. This file demonstrates by example how to obtain data from the web service rather than using the EDM GUI - so scripts and other applications can leverage it.

B.2. Installation of Estuary Data Mapper Tool

The latest version of EDM can be downloaded from www.epa.gov/edm. Right-click on the appropriate line in the menu on the right-hand portion of the screen to download the zip file for the version of EDM associated with your operating system and follow the instructions on this page to save as EDM.zip and extract to a folder. Once EDM is unzipped, you can click on the EDM.bat file to start an interactive EDM session.

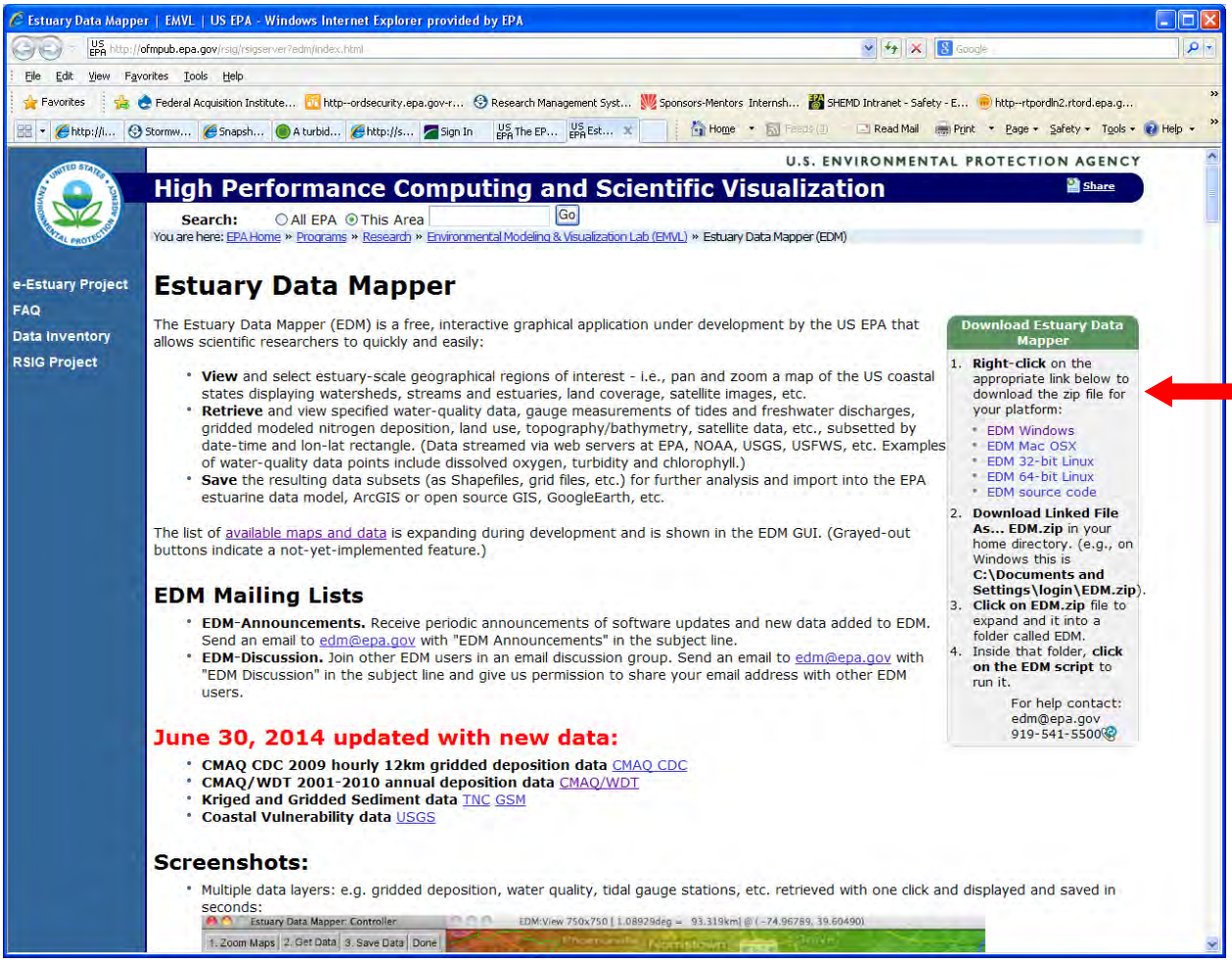

Figure B-1. Estuary Data Mapper home page.

B.3. Zooming to Area of Interest

The first time you start the EDM application, you will view the first tab and a default display extent:

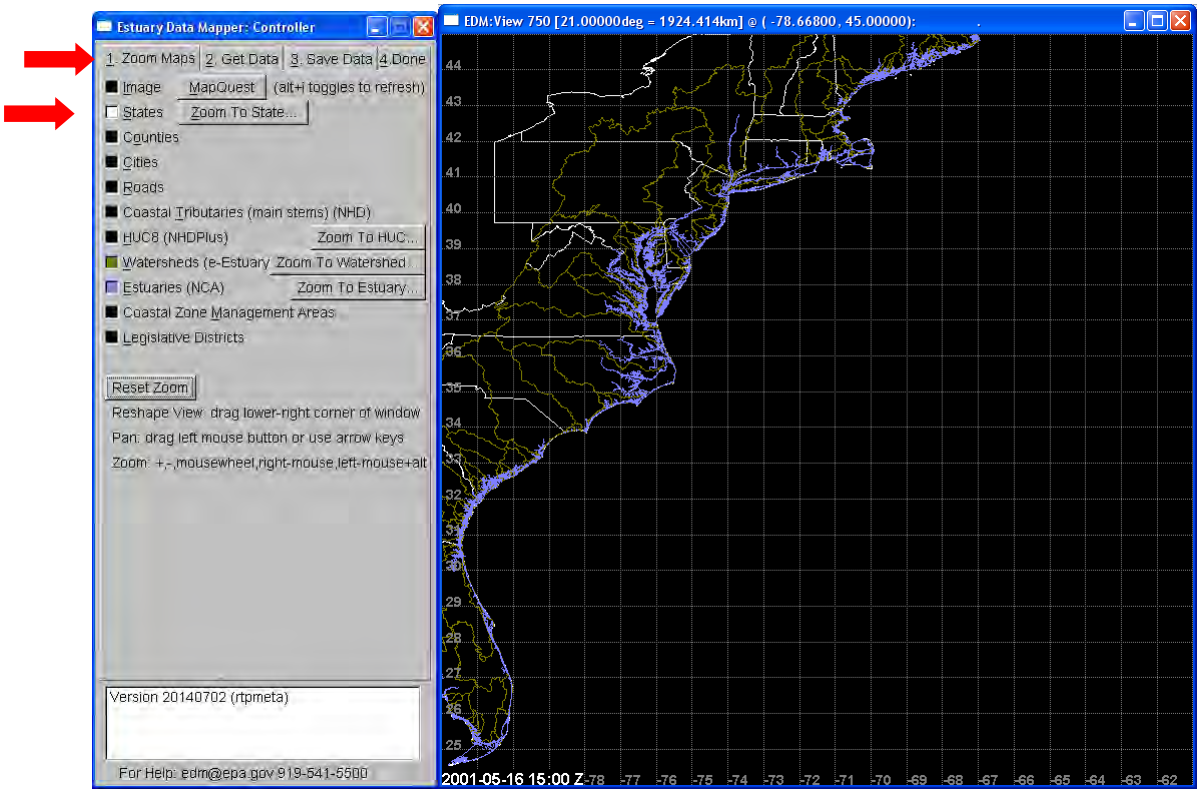

Figure B-2. First tab of Estuary Data Mapper with default display extent.

Although the default display extent consists of the Atlantic seaboard, EMD provides access to estuaries and coastal watersheds across the entire conterminous United States. To facilitate finding an area of interest, make sure the State, Watersheds (e-Estuary), and Estuaries (NCA) boundaries are turned on, e.g., click the box next to States until it turns white. When selections are toggled "on", the color of the boxes next to the selections turns from black (not selected) to the color of the boundary line of interest. To narrow down your search area, click on the Zoom to State button and select a state of interest.

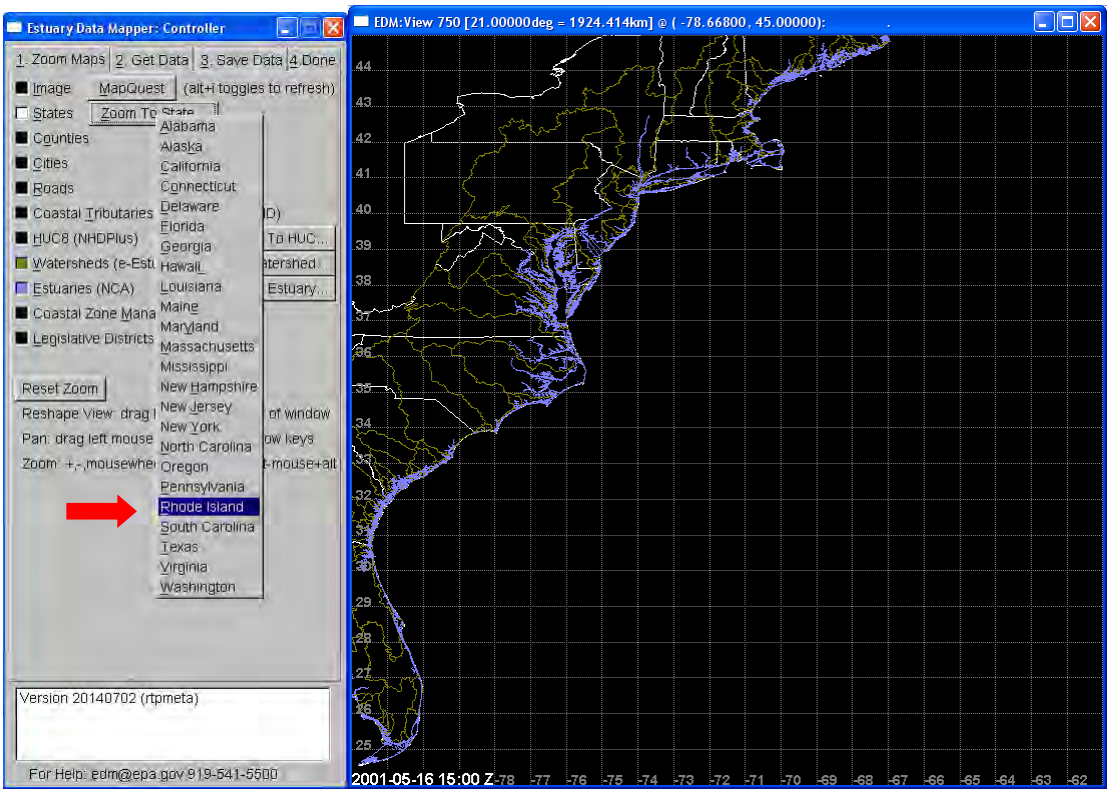

Figure B-3. Selection for zooming in to a state on first tab of Estuary Data Mapper.

Once zoomed into a state, when you click on Zoom to Estuary, you will be provided with a list of only those estuarine boundaries appropriate for the selected state:

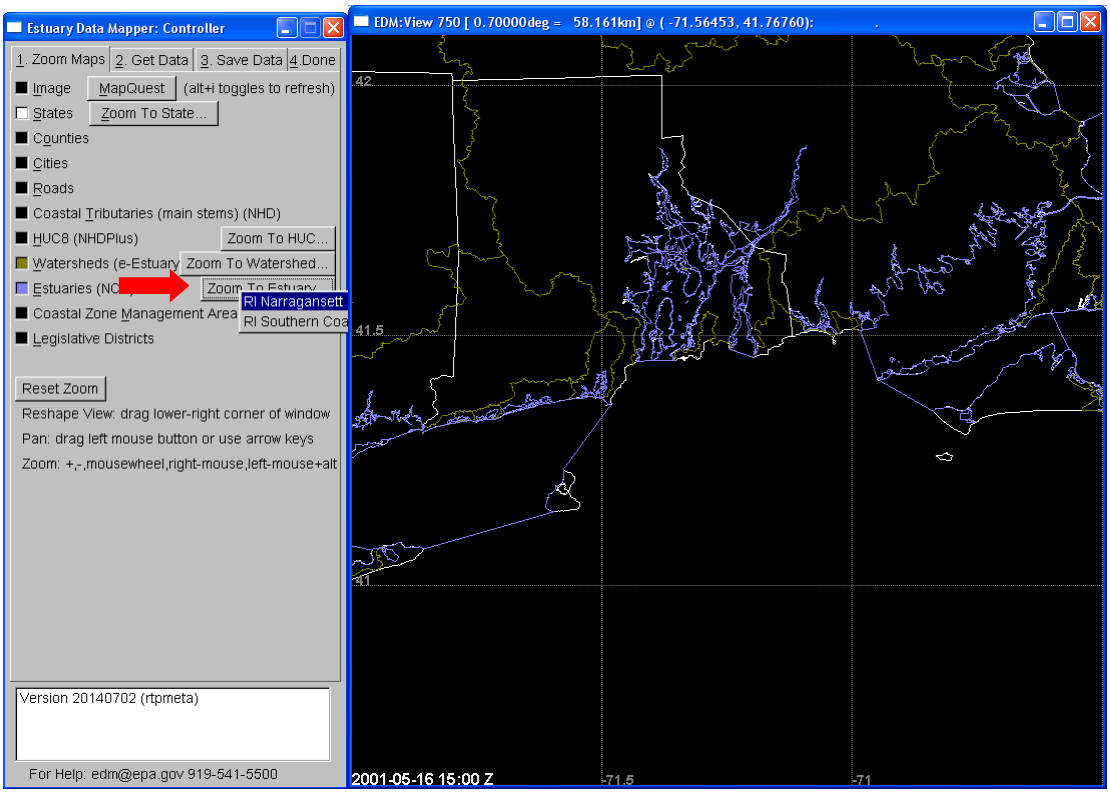

Figure B-4. Process of zooming into an individual estuary in Estuary Data Mapper.

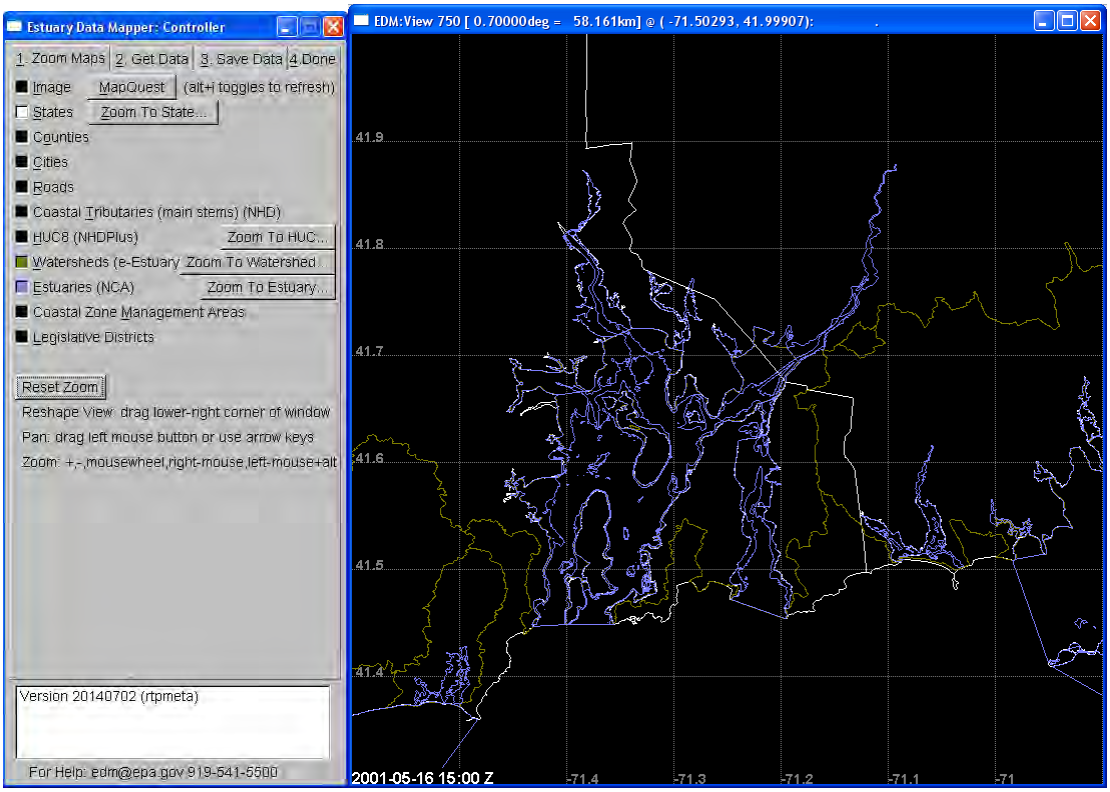

Figure B-5. Estuary Data Mapper zoomed in to Narragansett Bay.

The display extent will shift to the boundaries of the selected estuary. The view can be reshaped using your mouse following the instructions on the left-hand panel.

Alternatively, if you wish to identify and download data associated with the full watershed for an estuary, you would click on Zoom to Watersheds (e-Estuary) to zoom to the associated watershed boundary.

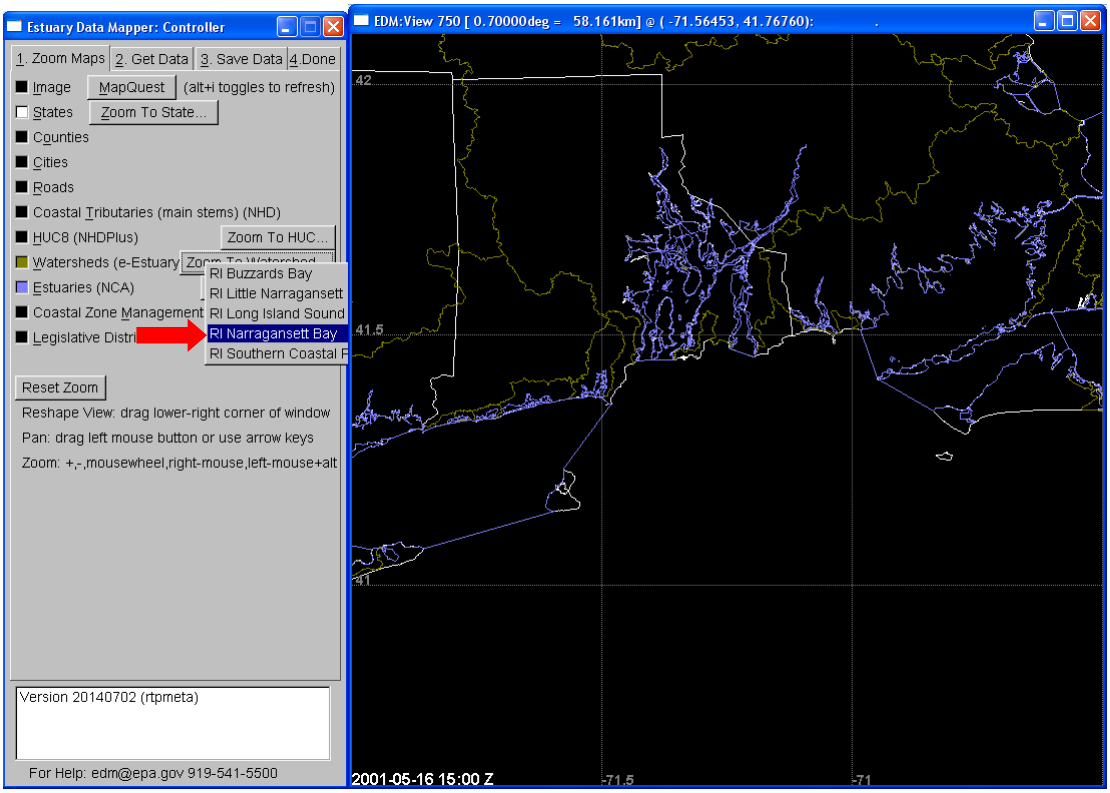

Figure B-6. Zooming in to an estuarine watershed in Estuary Data Mapper.

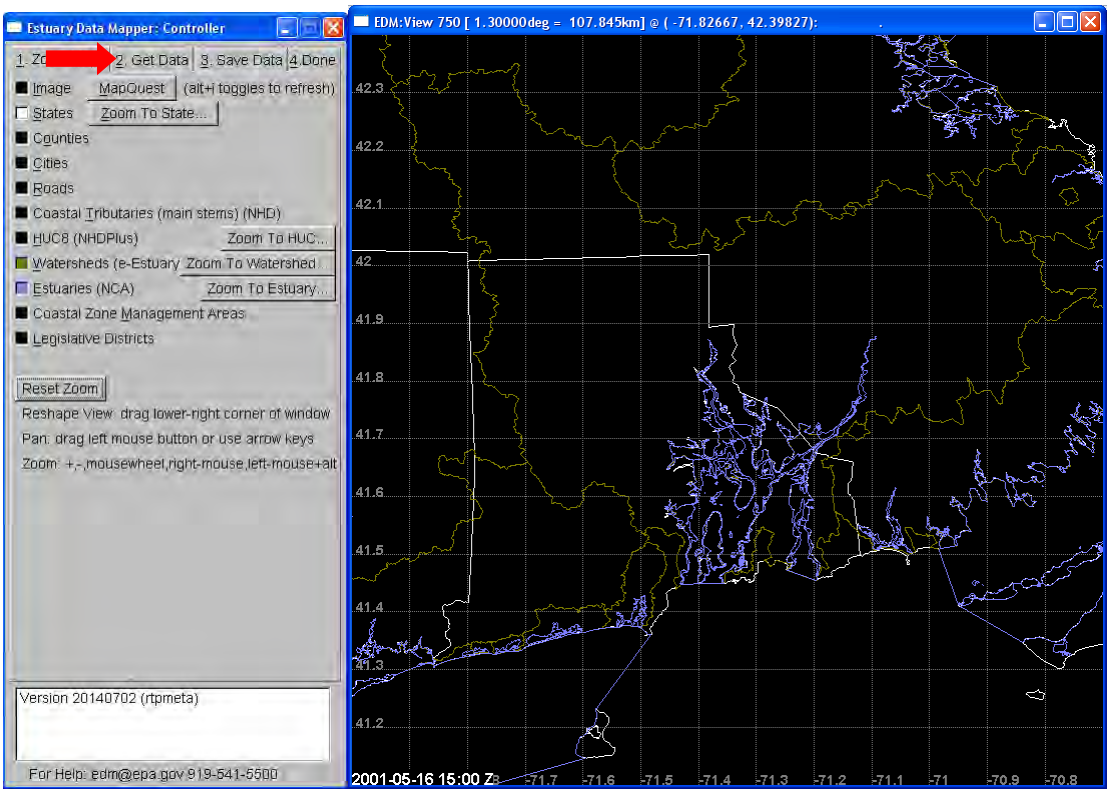

Figure B-7.Get Data tab of Estuary Data Mapper.

Now you will see the full extent of the watershed of interest. To begin exploring data of interest, you now click on the 2 Get Data tab on the top of the control panel.

B.4. Identifying Data Sources

B.4.1. System Boundaries

System boundaries that are checked on Tab1 are automatically downloaded when you save other data from that system. However, normally a smoothed version of the watershed boundaries is saved to speed up the process of saving data. If the original higher-resolution watershed boundary is desired, click on the Save Map Polygons/DBF box in Tab 3 Save Data before selecting Save Data (Shapefile) and type in an appropriate directory location to save the shapefiles to (Choose Directory Folder for Saved Files):

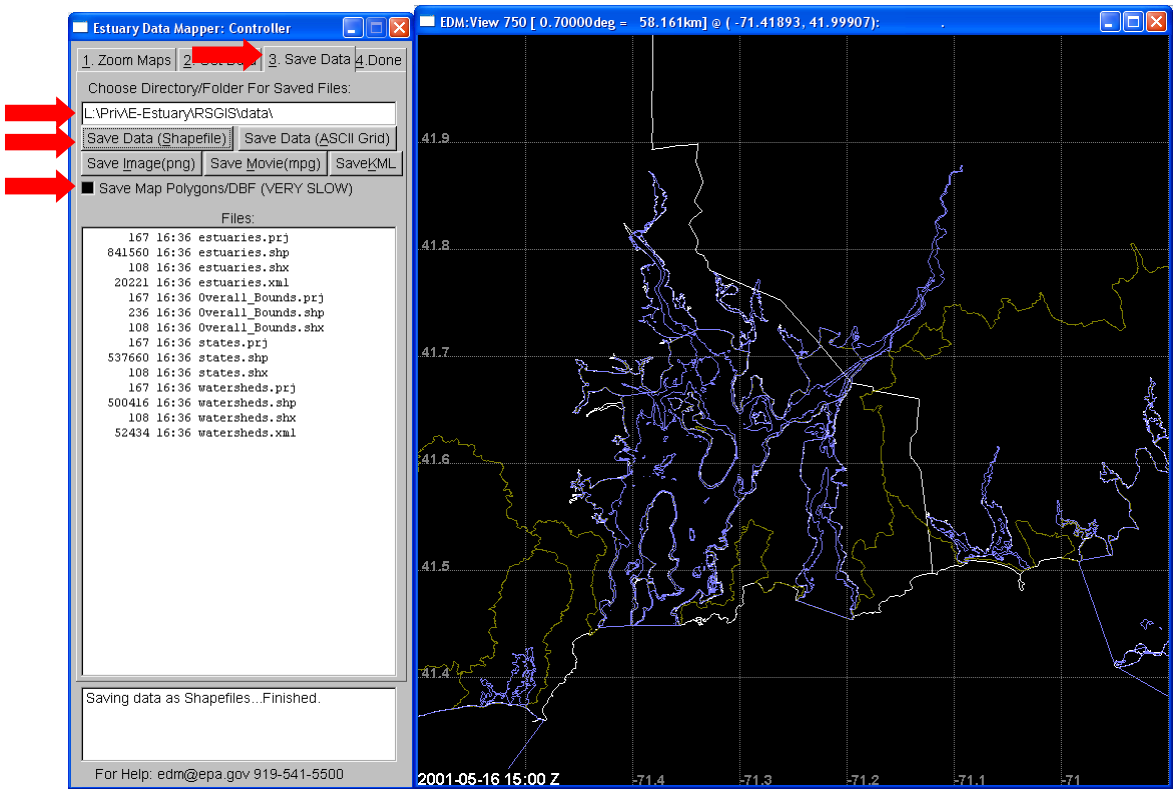

Figure B-8. Save Data tab on Estuary Data Mapper.

B.4.2. Seagrass

Click on the second tab to start the data discovery process. There is currently only one option for retrieving seagrass coverages, which returns a composite coverage with data combined across multiple sources representing the most up-to-date sources publically available as of 2013. Coding of seagrass abundance categories was standardized as explained in the metadata and original sources are also described in the metadata. Click on the Seagrass (State/NOAA) button and then on the Retrieve and Show Selected Data button. In the text box at the base of the menu panel you will messages indicating which data are being retrieved, when the data retrieval has finished, and how many records were returned. In this case, because there is only one composite coverage, any dates you enter will be ignored. In this example, we have manually zoomed into an area in the southern portion of Narragansett Bay in order to see the seagrass pixels. (If you do this, you will need to remember to zoom back out to the full extent before saving or you will only retrieve this subset.)

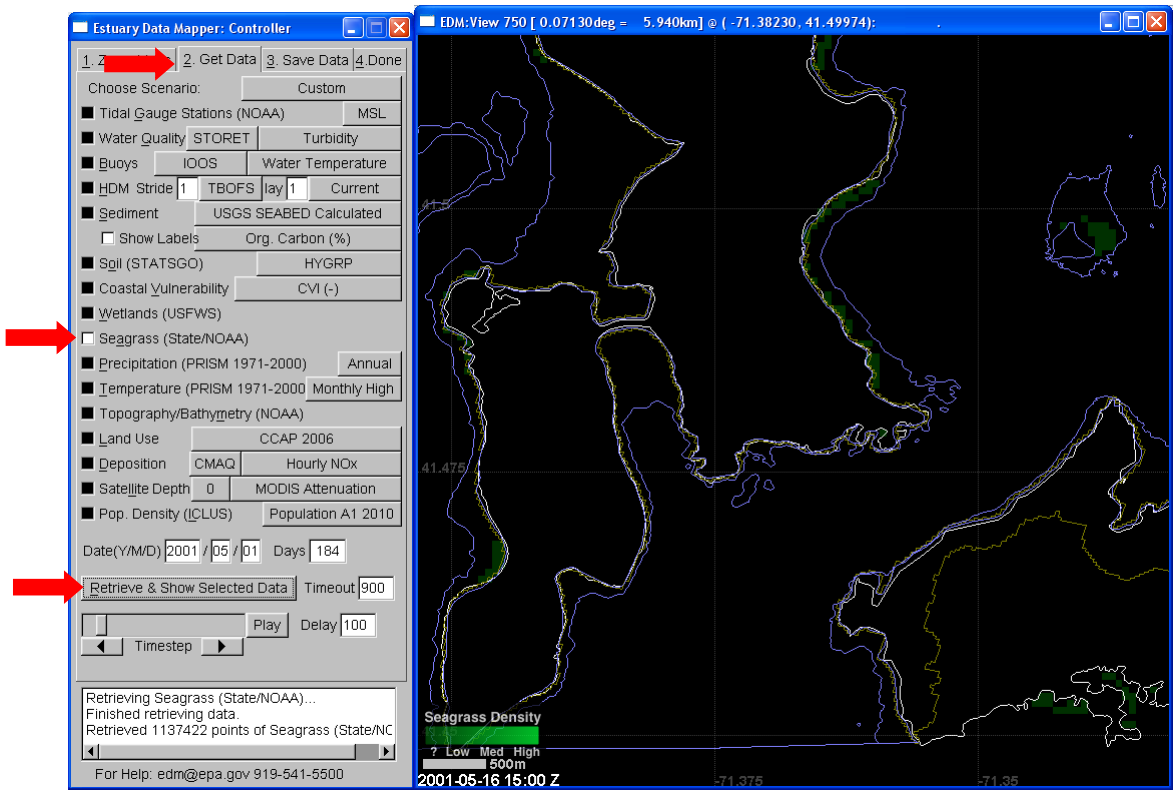

Figure B-9. Retrieving seagrass data in Estuary Data Mapper.

B.4.3. Depth

Only one source of merged topographic and bathymetric data is currently available in EDM, NOAA's Coastal Data Model (Topography/Bathymetry (NOAA)). Select this button and hit Retrieve and Show Selected Data again. Note that if you hover your mouse over the item in the menu, you will see a popup screen displaying the url for information on the original data source. Sources for data are also available on the EPA EDM web page at http://ofmpub.epa.gov/rsig/rsigserver?edm/data_inventory.html.

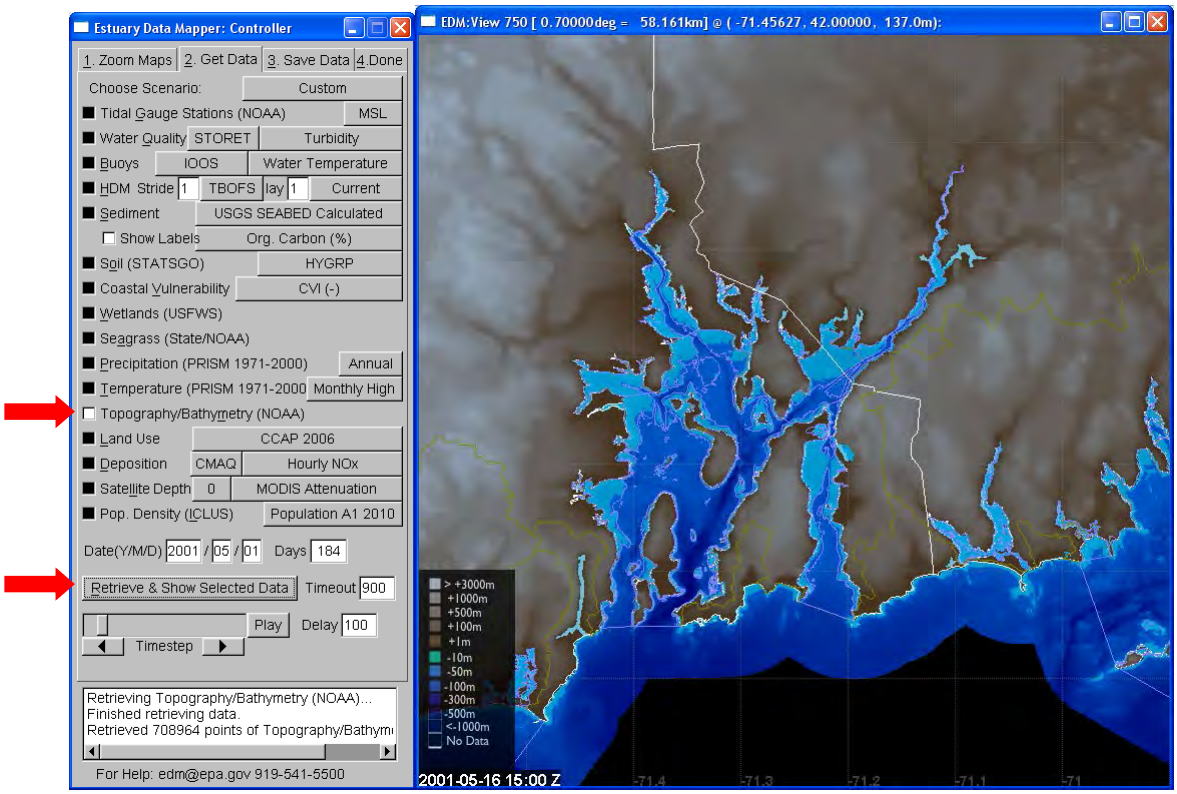

Figure B-10. Retrieving topobathymetry data in Estuary Data Mapper.

Users requiring finer resolution bathymetry data may wish to check local sources or the NOAA Tsunami Inundation Digital Elevation Models (DEMs; http://www.ngdc.noaa.gov/mgg/inundation/tsunami/), available for select regions only.

B.4.3. Transparency

Multiple data sources are available to describe transparency, including light attenuation coefficients, Secchi depth, turbidity, chlorophyll, and total suspended solids. Data associated with grab samples or instantaneous sensor readings for water quality parameters can be retrieved through web services provided by the National Estuarine Research Reserve System (NERRS) program or by the joint USGS/EPA (STORET button) web services. Click on the STORET button next to Water Quality to see the drop down list which allows you to select between the NERRS and STORET web services. Clicking on the button to the right of STORET activates the dropdown list to allow the user to choose a water quality parameter. At this time, only one parameter can be selected at a time, but users can choose to download results then select another parameter. Retrieval of water quality values requires that the user select a starting date and number of days of record (up to 365) for retrieval. Depending on the time of day, available bandwidth, demand for web services, and amount of data retrieved, these requests can take up to a few minutes. The user may need to increase the timeout parameter to the right of the Retrieve & Show Selected Data button to up to 900 seconds to prevent the tool from timing out and returning a message indicating no data points are available.

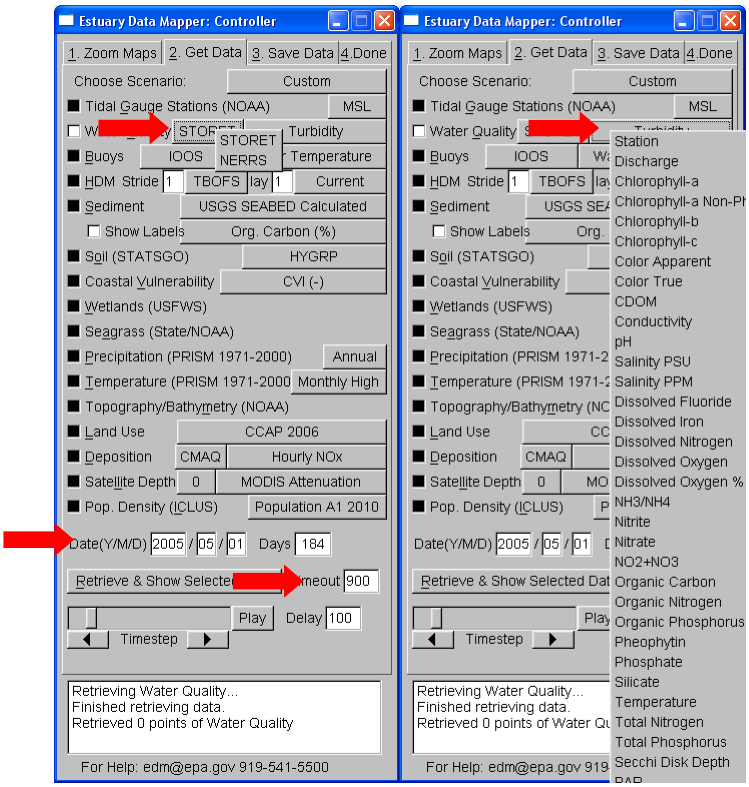

Figure B-11. Selecting STORET variables for data retrieval in Estuary Data Mapper.

Queries through STORET web services often fail to locate data collected through EPA's National Coastal Assessment surveys (e.g., Secchi depth, light attenuation coefficients, TSS, chlorophyll) so users may wish to retrieve data directly from the EPA's Environmental Monitoring and Assessment web site (http://oaspub.epa.gov/coastal/coast.search) for data up to 2006 or the National Coastal Assessment site (http://water.epa.gov/type/oceb/assessmonitor/ncca.cfm) for data from 2008 or later (due to be

added by the end of 2014). Selected parameters from EPA's NCA surveys will be added to EDM in the near future to fill the gap in WQ web services.

Remotely sensed light attenuation coefficients based on the MODIS satellite are also available through NASA web services. Caution should be exercised in using these data for shallow systems (< 30m depth) as algorithms were developed for the open ocean and do not include corrections for bottom reflectance. Therefore, data should be checked against in situ measurements to determine if a relationship exists. Algorithms for ocean color parameters are currently under development and are being tested for shallow and more turbid coastal systems (Keith et al. 2014) but these data are not yet readily available through web services.

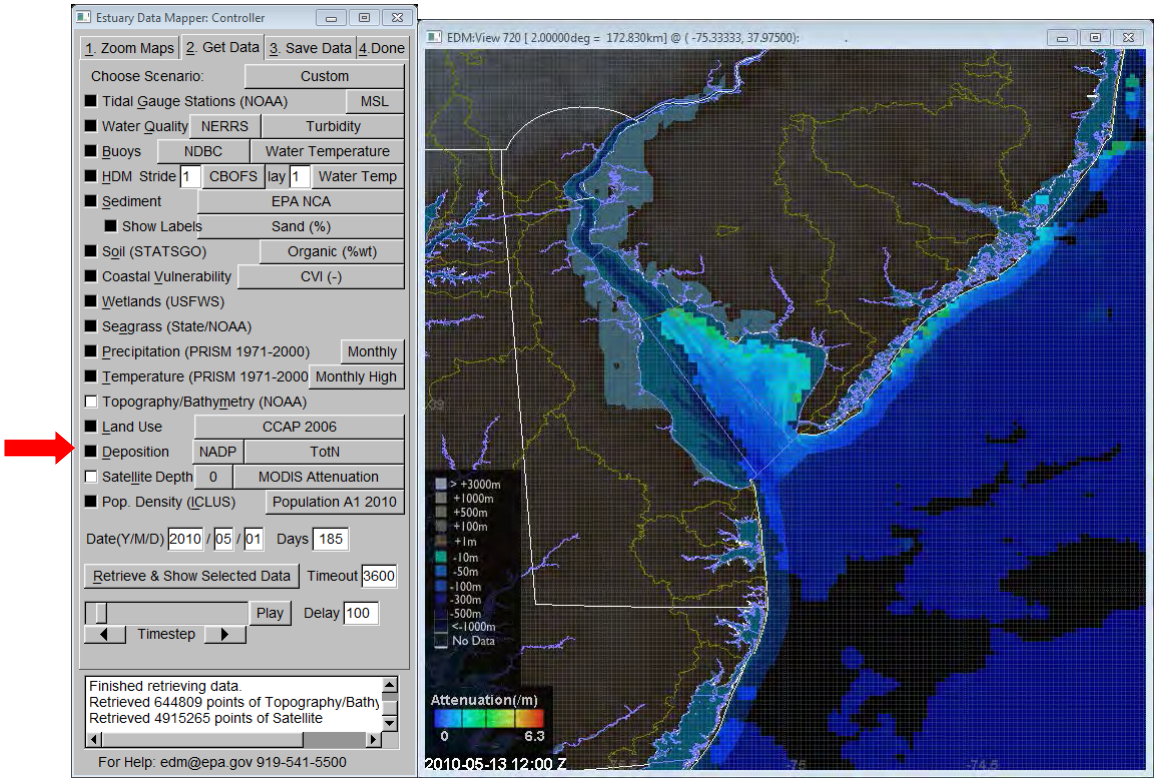

Figure B-12. Retrieving remotely sensed light attentuation data for Chesapeake Bay in Estuary Data Mapper.

B.4.4. Energy Environment

B.4.4.1. Wave Energy Model Inputs

There are a few options available for the user to calculate Relative Wave Energy. Previously, users could use the NOAA Wave Energy Model (WEMO; http://products.coastalscience.noaa.gov/wemo/) to calculate fetch and relative wave energy based on input data including system boundaries, merged topo-bathymetry, and wind data. Unfortunately WEMO is no longer being supported and is not compatible with ArcMap versions later than 9.3. Recently, the USGS WAVE extension for ArcMap was upgraded for use with ArcMap 10 and could be used in a similar fashion. See Sections B.4.1 and B.4.3 for information on downloading data on system boundaries and merged topo-bathymetry. Wind data can be downloaded from the NERRS web service (wind speed and direction or from various NOAA web services for buoy data (Buoys button, using IOOS, NDBC or NERACOOS dropdown options).

B.4.4.2. Relative Exposure

USGS has recently calculated a Coastal Vulnerability Index, including a component related to wave energy (http://woodshole.er.usgs.gov/project-pages/cvi/). Unfortunately values have been collapsed onto an ordinal scale (1-5), leading to a loss of information. For some systems such as the Narragansett Bay (below), the range of values may be restricted, making this index less useful as a predictive tool. However, for other systems with a wider range of energy environments, the Wave Rank score might prove useful.

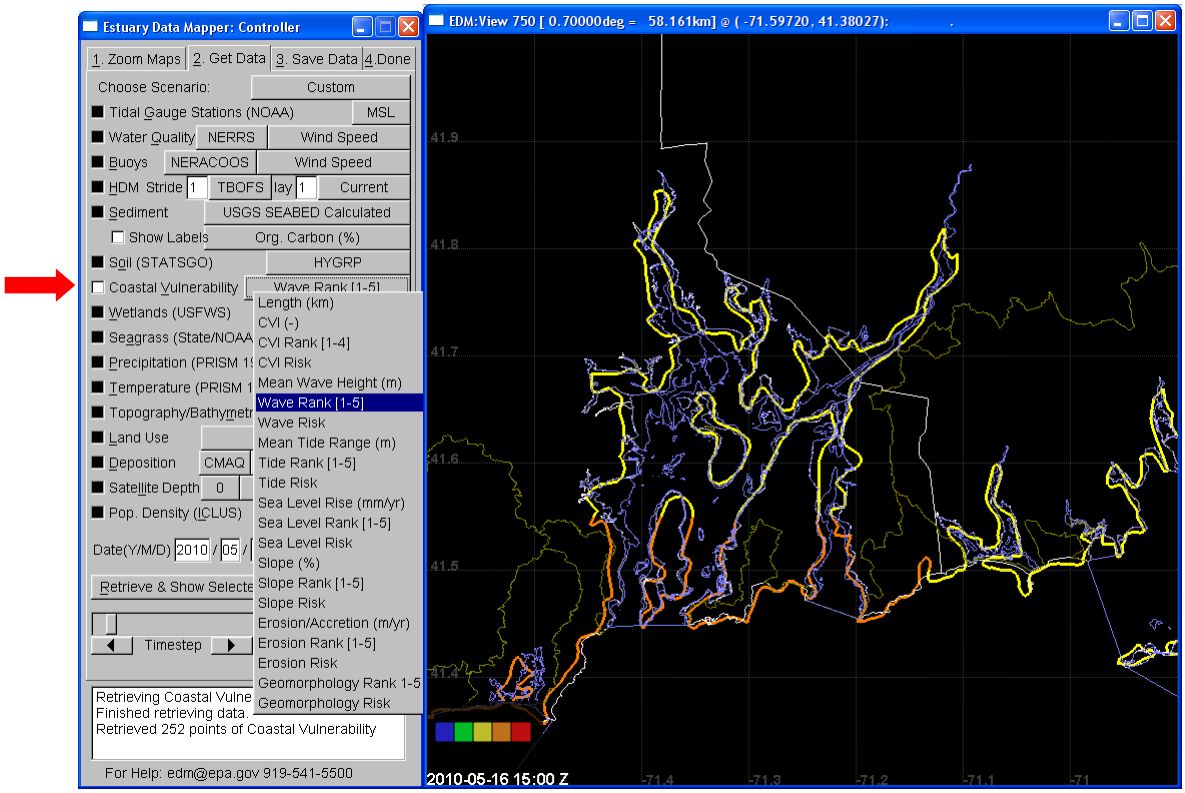

Figure B-13. Retrieving Coastal Vulnerability Index data in Estuary Data Mapper.

B.4.4.3 Current velocity

Current velocity data are available at some Buoy locations; however, these stations tend to be sparse. For a limited number of systems, a more complete coverage of the current velocity environment can be obtained from model output. Hydrometeorological model outputs for NOAA's Operational Forecast Systems are available for seven systems. Choose system using dropdown menu by HDM button for the layer of interest.

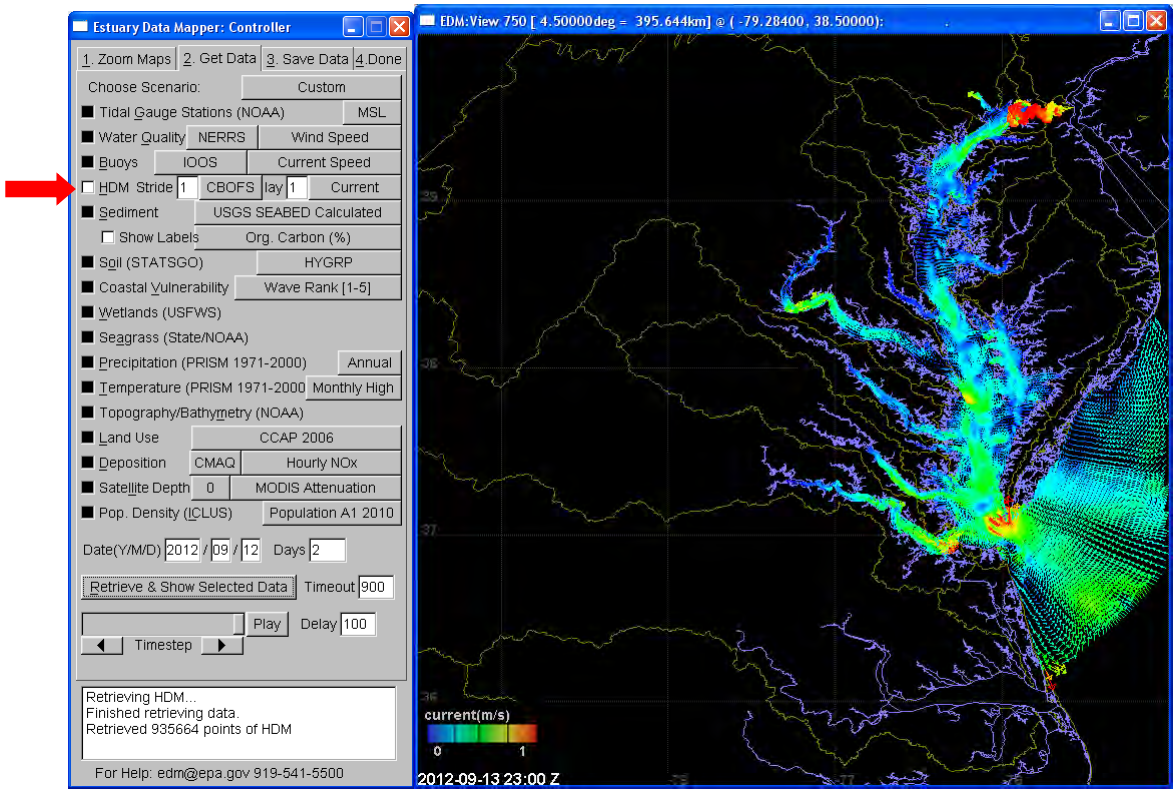

Figure B-14. Retrieving current velocity data in Estuary Data Mapper.

B.4.5. Sediment Characteristics

Sediment characteristics are available both for grab samples from US EPA National Coastal Assessments (URL) or the USGS Seabed database (URL) and as continuous grids developed using kriging methods for the North Atlantic coast (https://www.conservationgateway.org/ConservationByGeography/NorthAmerica/UnitedStates/edc/reportsdata/marine/namera/Pages/default.aspx) or Gulf coast (http://instaar.colorado.edu/~jenkinsc/dbseabed/resources/gsmseabed/). You can choose between these options on the Sediment selection (a), which leads you to the drop down menu for sediment characteristics for the NCA (b), or USGS Seabed datasets (c), or for the krigged datasets (d) and e):

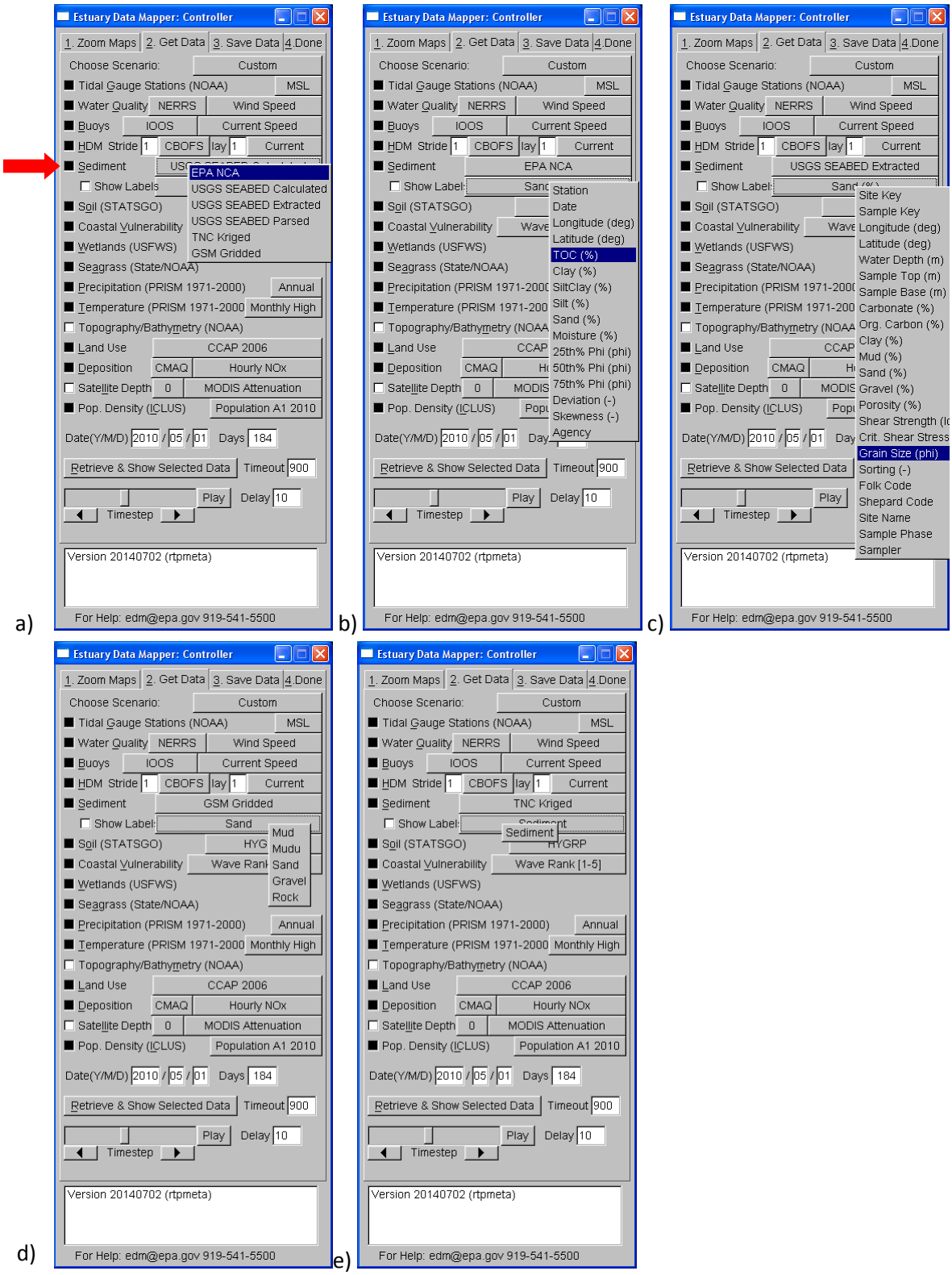

Figure B-15. Drop-down selections for selection of sediment parameters from EPA National Coastal Assessment dataset.

B.4.6. Water Quality

B.4.6.1. Temperature

EDM provides access to instantaneous temperature readings through the EPA/USGS web services (Section 4.3) time series at fixed stations through the NERRS or various buoy web services (Section 4.4.1), and remotely sensed time series with continuous gridded coverages. Although several sources are available in the Satellite drop-down menu, the MUR option provides grids with the finest resolution and best coverage for many estuarine systems. Note that the color scale has been optimized to represent the range in temperature over time, so will not do a good job of illustrating gradients in temperature for a particular day within EDM. Users investigating the effects of temperature for systems in which seagrasses are exposed to the air can also access local air temperatures through the web services for NERRS or NOAA's buoy systems.

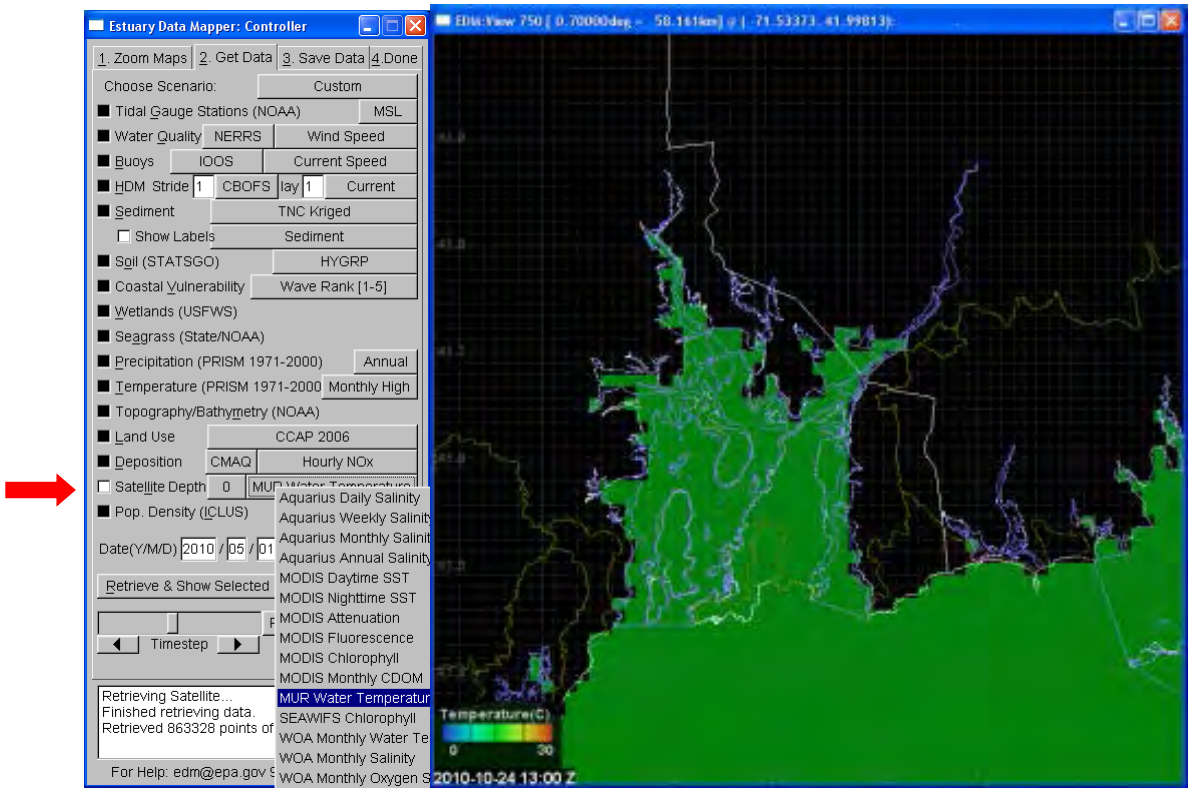

Figure B-16. Retrieval of remotely sensed temperature data with Estuary Data Mapper.

B.4.6.2. Salinity

EDM provides access to instantaneous salinity readings through the EPA/USGS web services (Section 4.3), time series at fixed stations through the NERRS, or various buoy web services (Section 4.4.1), and remotely sensed time series with continuous gridded coverages. The latter are available through NASA web services with data aggregated from daily to annual time steps. Different satellite coverages can be selected from the dropdown box selection under Satellites (see below).

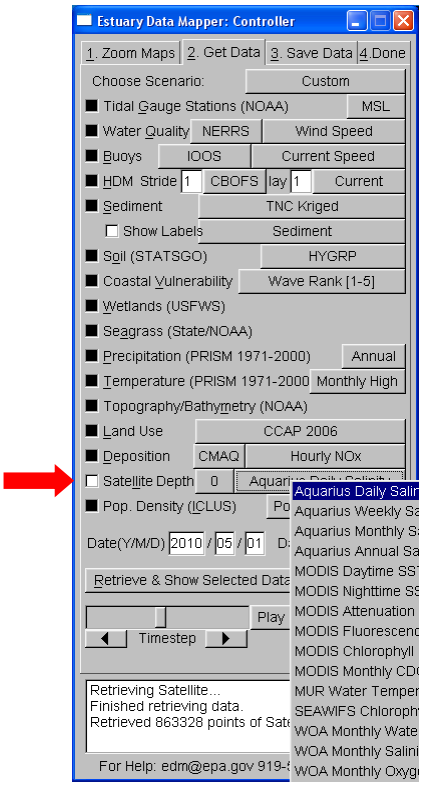

Figure B-17. Selection of remotely sensed salinity data for download in Estuary Data Mapper.

B.4.7. Nitrogen Concentration and Loading

B.4.7.1. Nutrient Concentrations

Nutrient concentrations for estuaries and their tributaries can be retrieved via the EPA/USGS and NERRS web services (see Section B.4.3). As mentioned previously, neither the EPA/USGS web service nor the online STORET database query interface reliably allow the retrieval of data collected during NCA surveys. See Section B.4.3 for directions on current web access to NCA data sets.

B.4.7.2. Atmospheric Loads

Data regarding nitrogen sources or loads to estuaries and their watersheds are available through the Nitrogen menu in EDM. Atmospheric loading data for nitrogen and phosphorus can be retrieved through EDM based on either 1) interpolation of monitoring data collected by the National Atmospheric Deposition Network (NADP; http://nadp.sws.uiuc.edu/) or 2) modeled deposition based on results of modeling runs of EPA's Community Multiscale Air Quality model (http://www.epa.gov/AMD/Data/wdtData.html; http://www.epa.gov/heasd/research/cdc.html). CMAQ deposition data are available to estimate both deposition over the estuarine watershed as well as directly to the estuary surface area for hourly, monthly or annual time steps. NADP grids only cover watershed deposition for an annual time step. Summaries of loading by estuary are covered below in Section B.4.7.3.

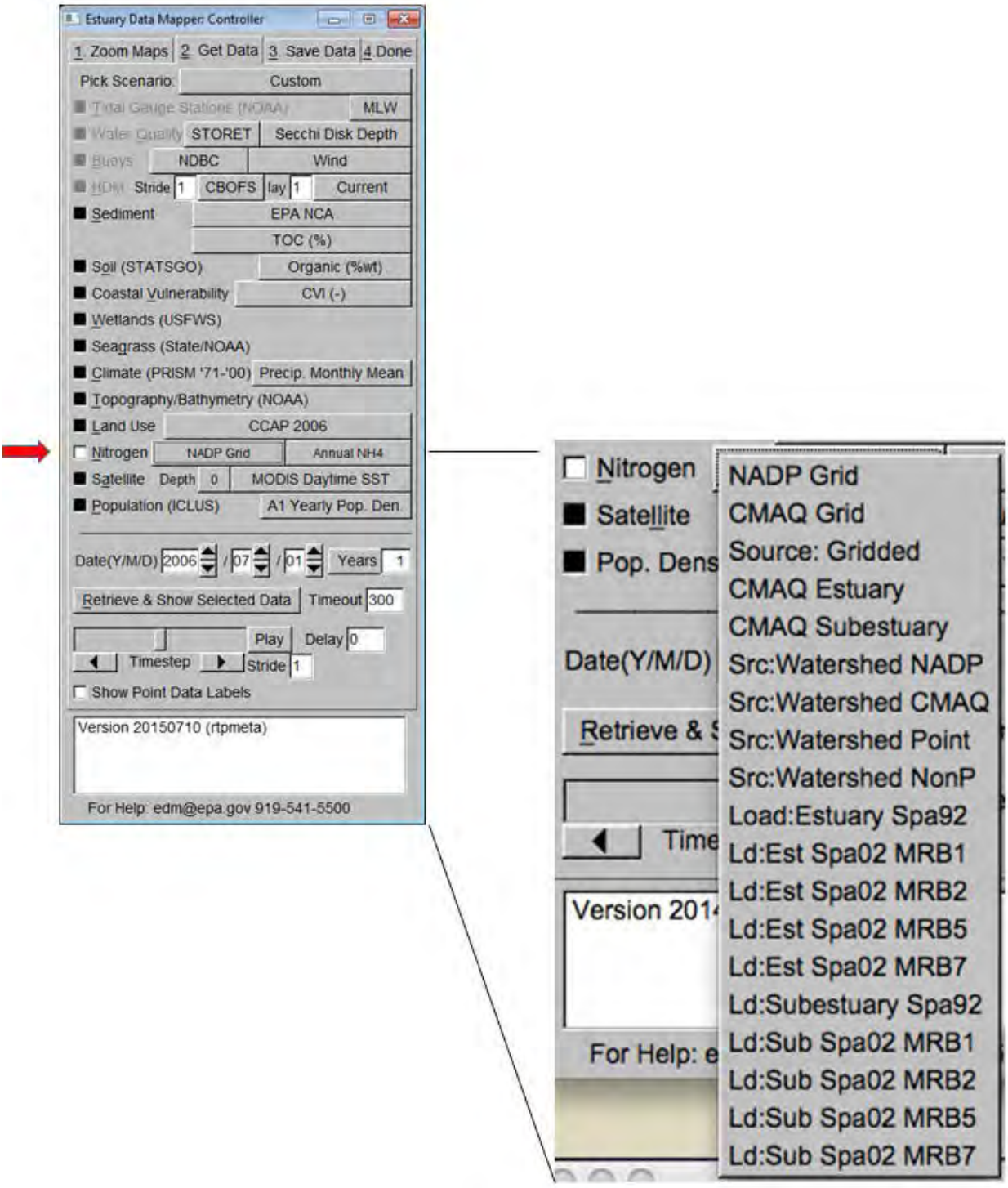

Figure B-18. Selection of nitrogen loading data sets in Estuary Data Mapper.

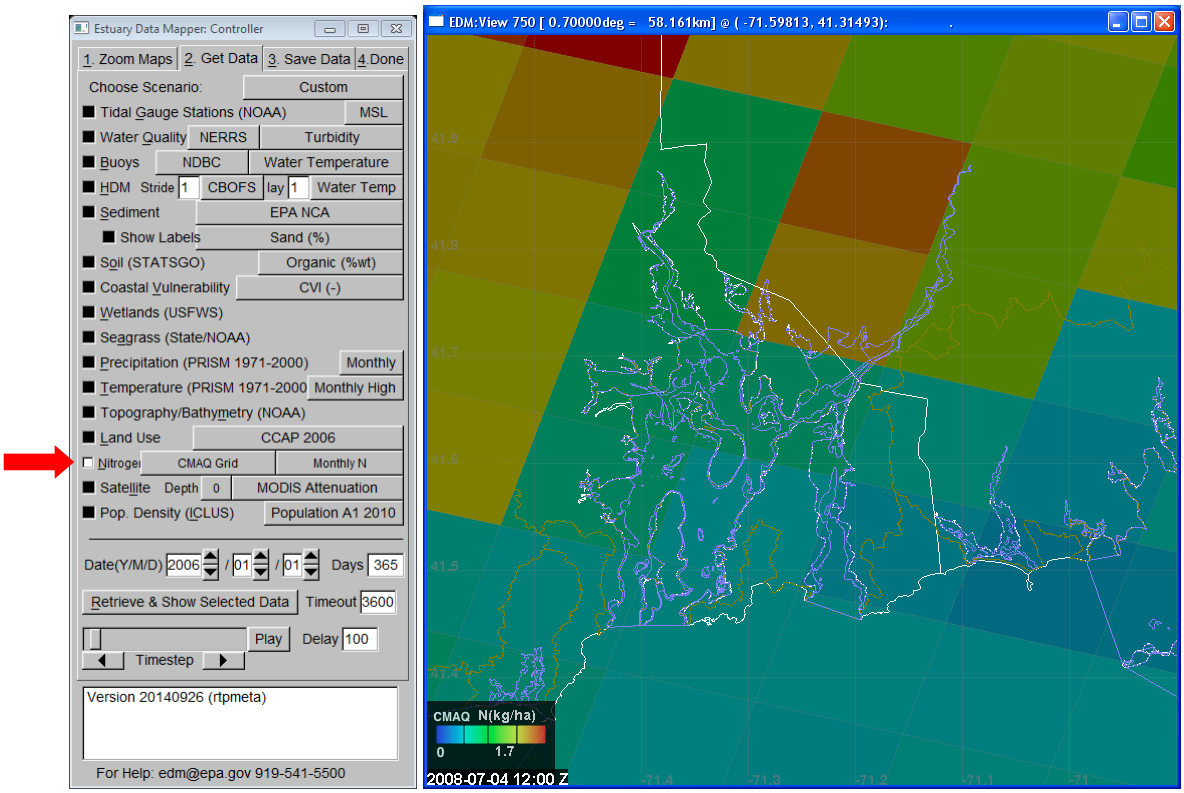

Figure B-19. Retrieval of CMAQ monthly nitrogen deposition data with Estuary Data Mapper.

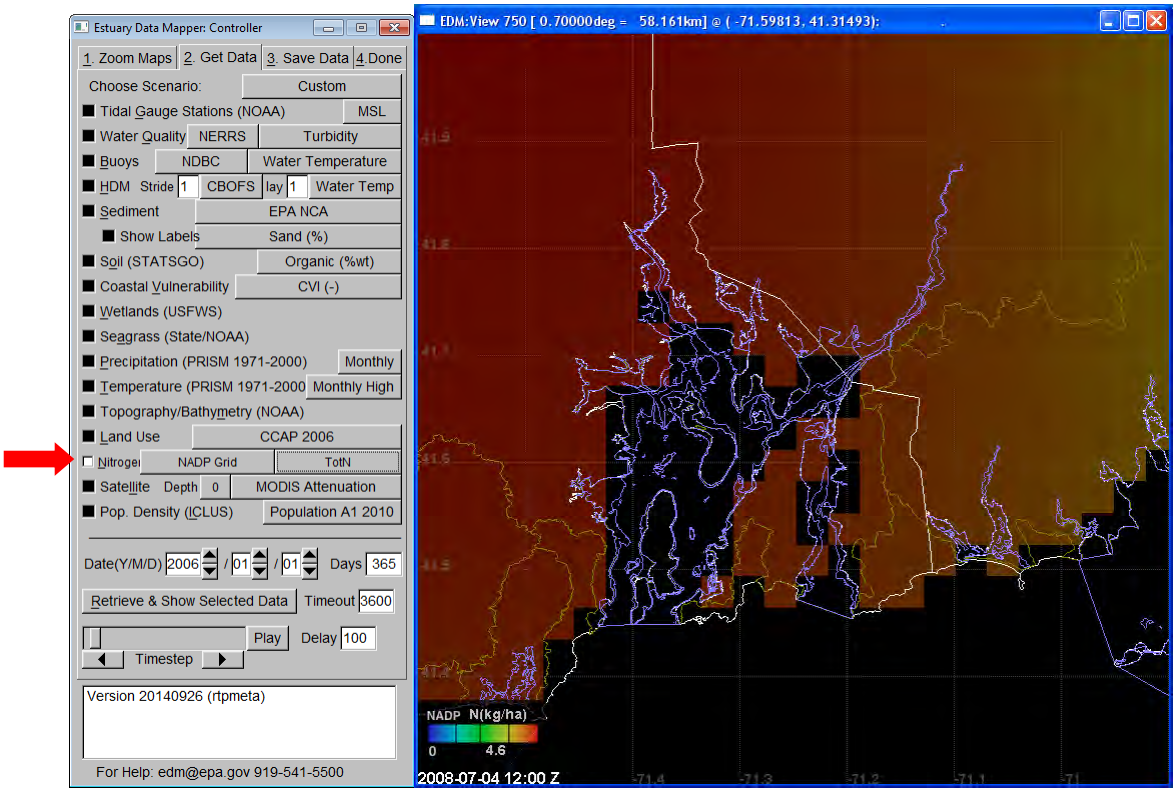

Figure B-20. Retrieval of NADP nitrogen deposition data with Estuary Data Mapper.

B.4.7.3. Watershed Sources

Nitrogen source and loading data are available through EDM in gridded form for estuarine watersheds. In addition, summaries of annual watershed-based loads and sources and direct atmospheric loading to estuaries by estuary are provided in dbf files associated with estuary and watershed shapefiles. Atmospheric deposition data can be accessed from the Nitrogen menu (a), which gives access to both gridded data at hourly (CMAQ), monthly (CMAQ), or annual time steps (CMAQ, NADP). In addition, summaries of annual CMAQ or NADP atmospheric N deposition and monthly CMAQ atmospheric N deposition are available for estuarine watersheds and estuaries/subestuaries (CMAQ only). NADP deposition grids have been interpolated for terrestrial areas only.

Selection of the CMAQ grid menu (b) provides access to a wide array of time steps and nitrogen forms.

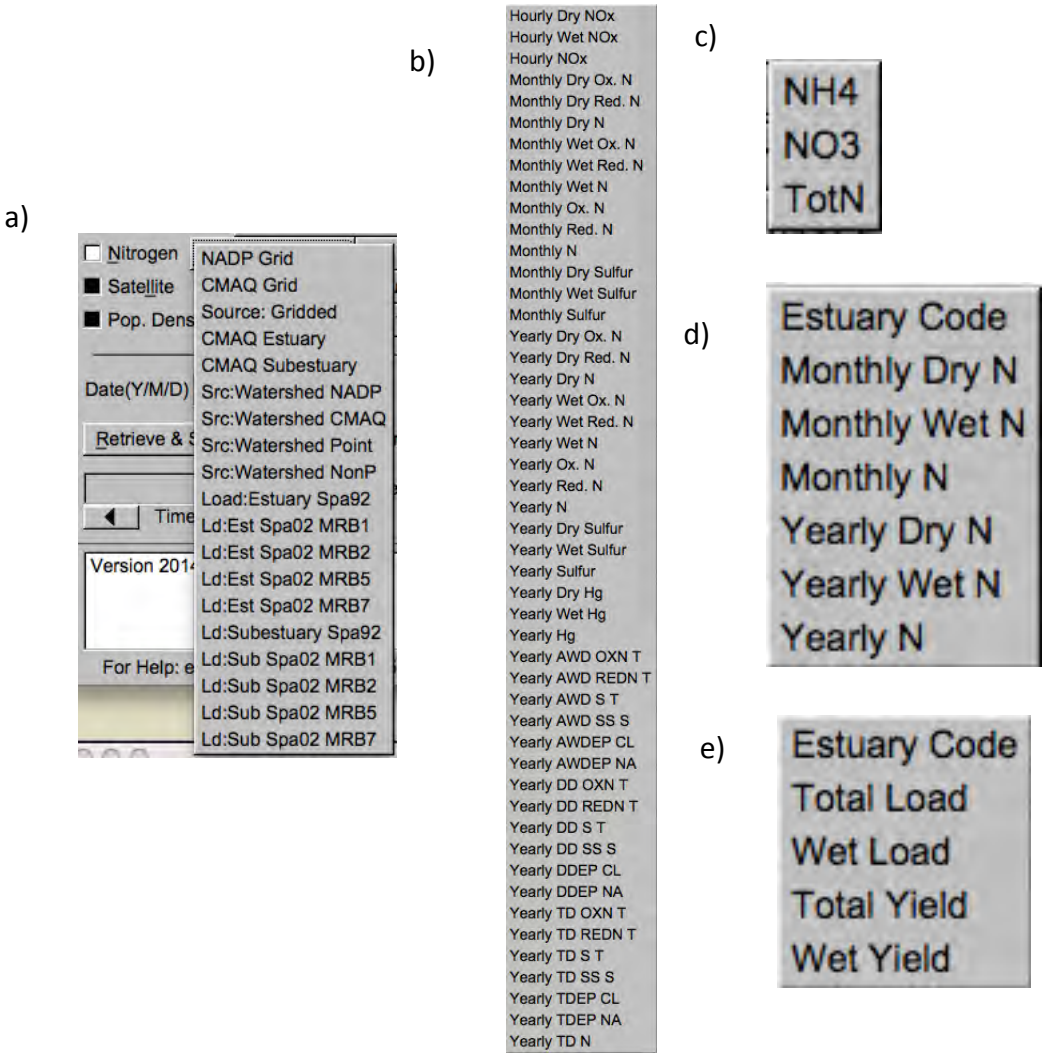

Figure B-21. Nitrogen submenus for CMAQ and NADP nitrogen deposition.

Selection of NADP grid sources in the main menu then provides access to a submenu (c) allowing access to different N fractions: total ammonia, nitrate and total N. Selection of CMAQ summaries at the estuary (d) or watershed (e) scales provides access to choices between wet, dry or total N fractions aggregated at monthly or annual time steps.

Additional land-based nitrogen sources (a) can be assessed with b) gridded maps or c) at the estuarine watershed scale.

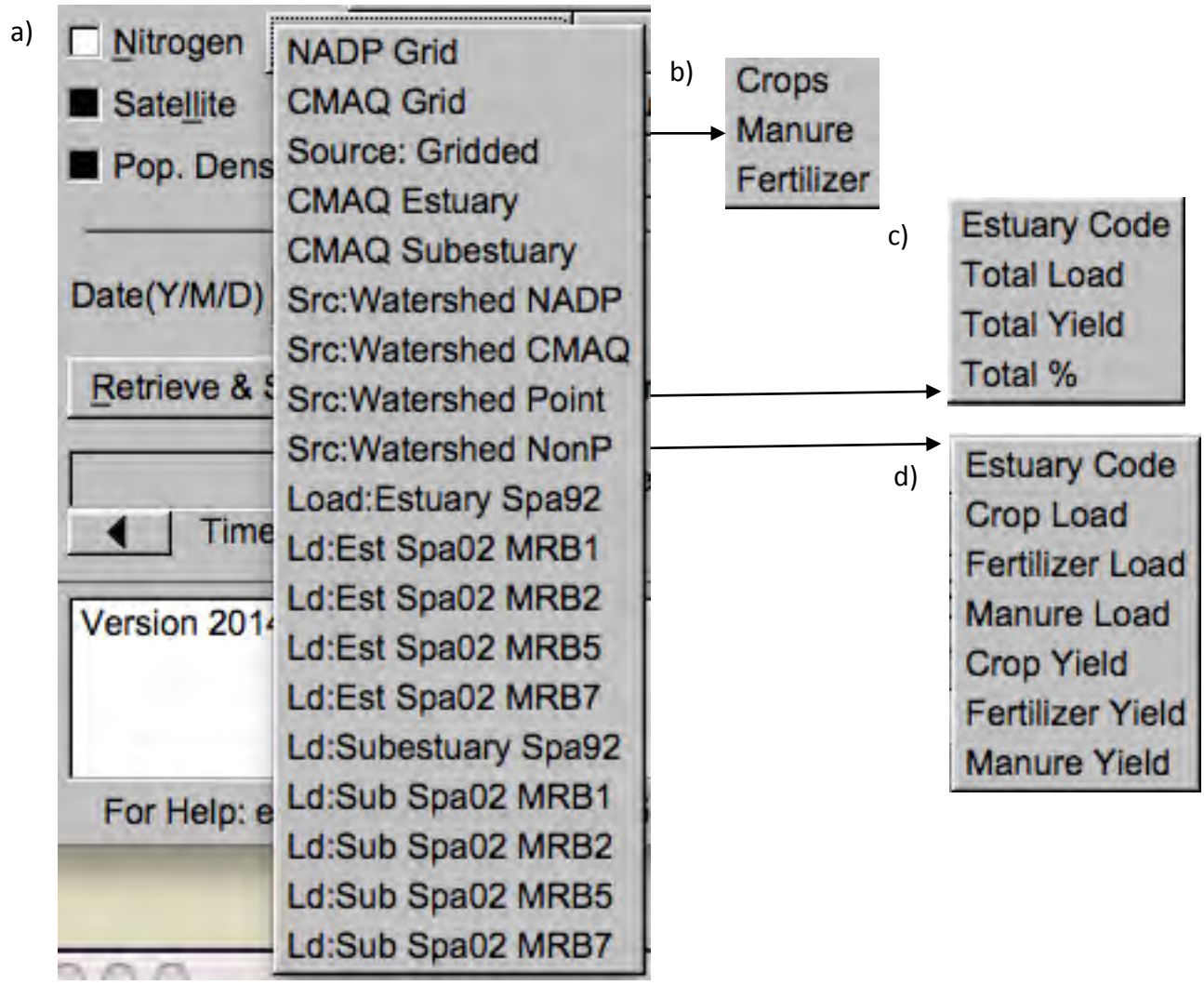

Figure B-22. Submenus for nitrogen source data at gridded or estuarine scales.

Finally, estuarine watershed loads and yields of nitrogen can be explored based on the output of 1992 National and 2002 regional SPARROW models (http://water.usgs.gov/nawqa/sparrow/mrb/) that have been aggregated to the estuarine watershed scale.

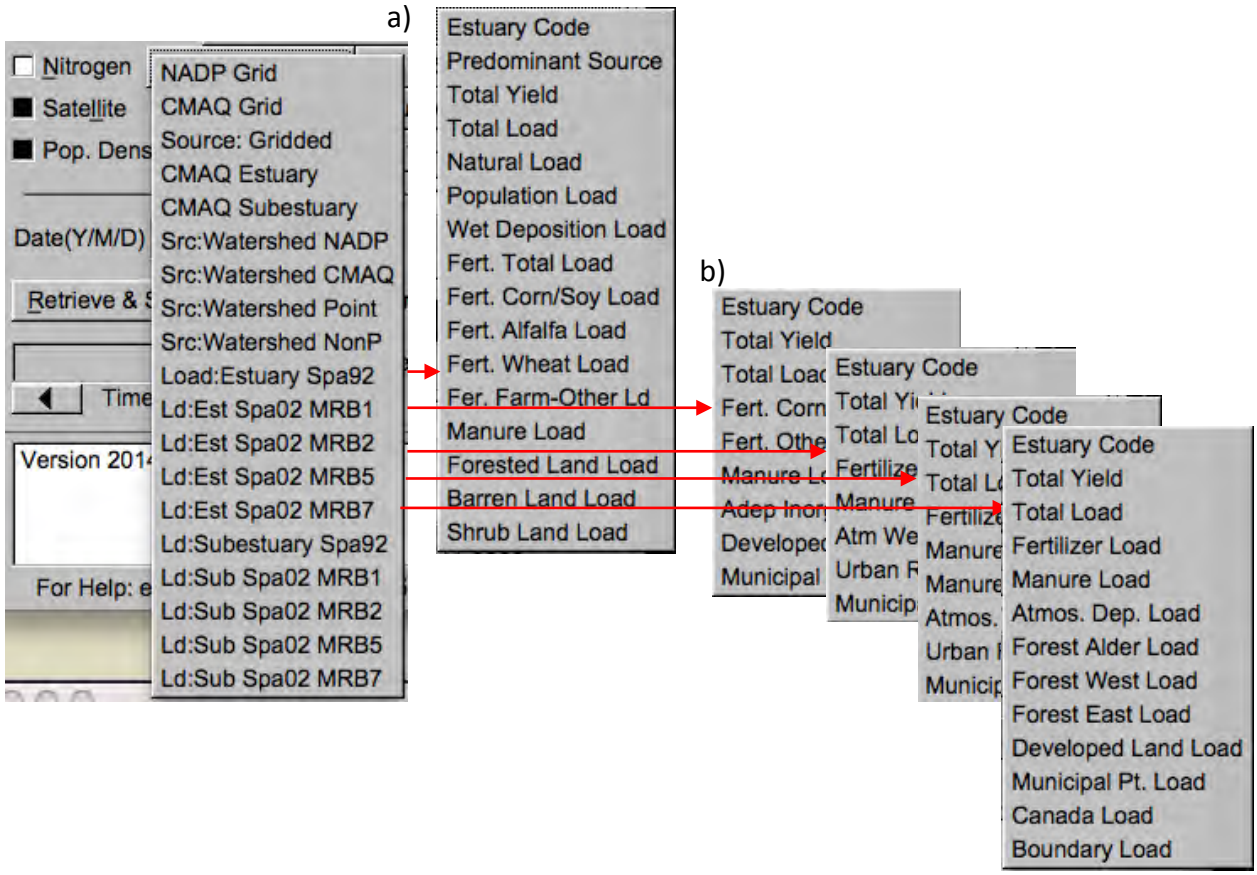

Figure B-23. Selection of N loading data from SPARROW models.

B.5. Downloading Data

The data displayed in the view window as the result of reference boundary layers selected under Tab 1. Zoom Maps, data selected under Tab 2 (Get Data) can be downloaded to the directory of the user's choice in Tab 3. Save Data. Data can be downloaded in a variety of formats for later viewing (.png, .mpg, .kml) or import into decision support modeling applications (shapefile, ASCII grid). Metadata are automatically provided in associated .txt or .xml files. See the message box at the bottom for an indication of when data downloads have been completed and Files Listing box for a list of files that have been downloaded. Time series of remote sensing data can be saved either as a series of ASCII grids, one for each date, or as a single shapefile. The latter output provides a more compact format for downloading and subsequent data calculations; each date of data of the time series is provided in a separate column of the associated dbf file. Users will need to pay attention to missing value indicators (e.g., -9999) in using these data for later calculations.

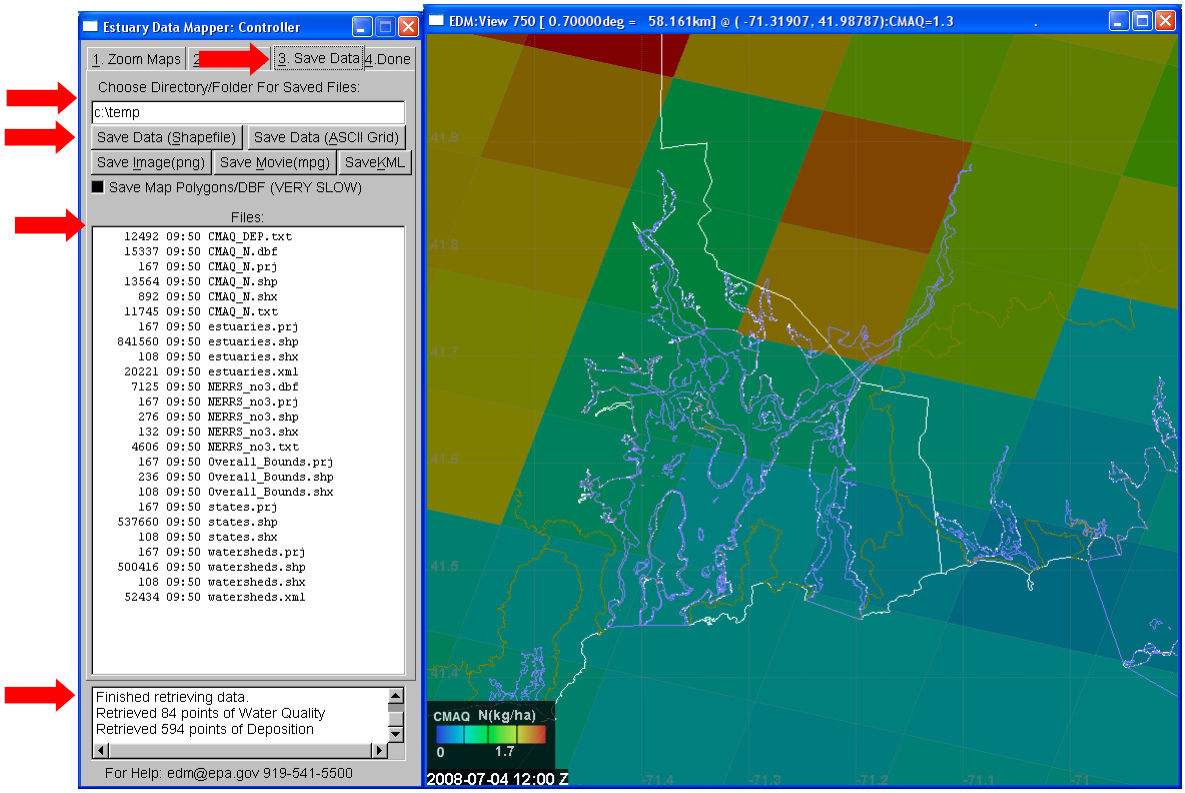

Figure B-24. Save Data tab in Estuary Data Mapper, illustrating saving shapefiles.

B.6. Terminating an EDM Session and EDM Updates

Users should end an EDM session using Tab 4. Done. Exiting in this manner will save user settings so that the next time EDM is started the user will automatically be zoomed in to the area of interest selected in the last session run. The user can also set preferences for updating EDM versions on this tab by selecting among the options given:

Figure B-25. Final tab in Estuary Data Mapper to close out program and check for updates.

Users are encouraged to use the contact email (edm@epa.gov) to provide feedback, to report bugs in the system, or to be added to one of two mailing lists:

- **EDM-Announcements.** Receive periodic announcements of software updates and new data added to EDM. Send an email to edm@epa.gov with "EDM Announcements" in the subject line.
- **EDM-Discussion.** Join other EDM users in an email discussion group. Send an email to edm@epa.gov with "EDM Discussion" in the subject line and give us permission to share your email address with other EDM users.

Appendix C. R Packages and Commands Used in Development of Predictive Seagrass Habitat Models

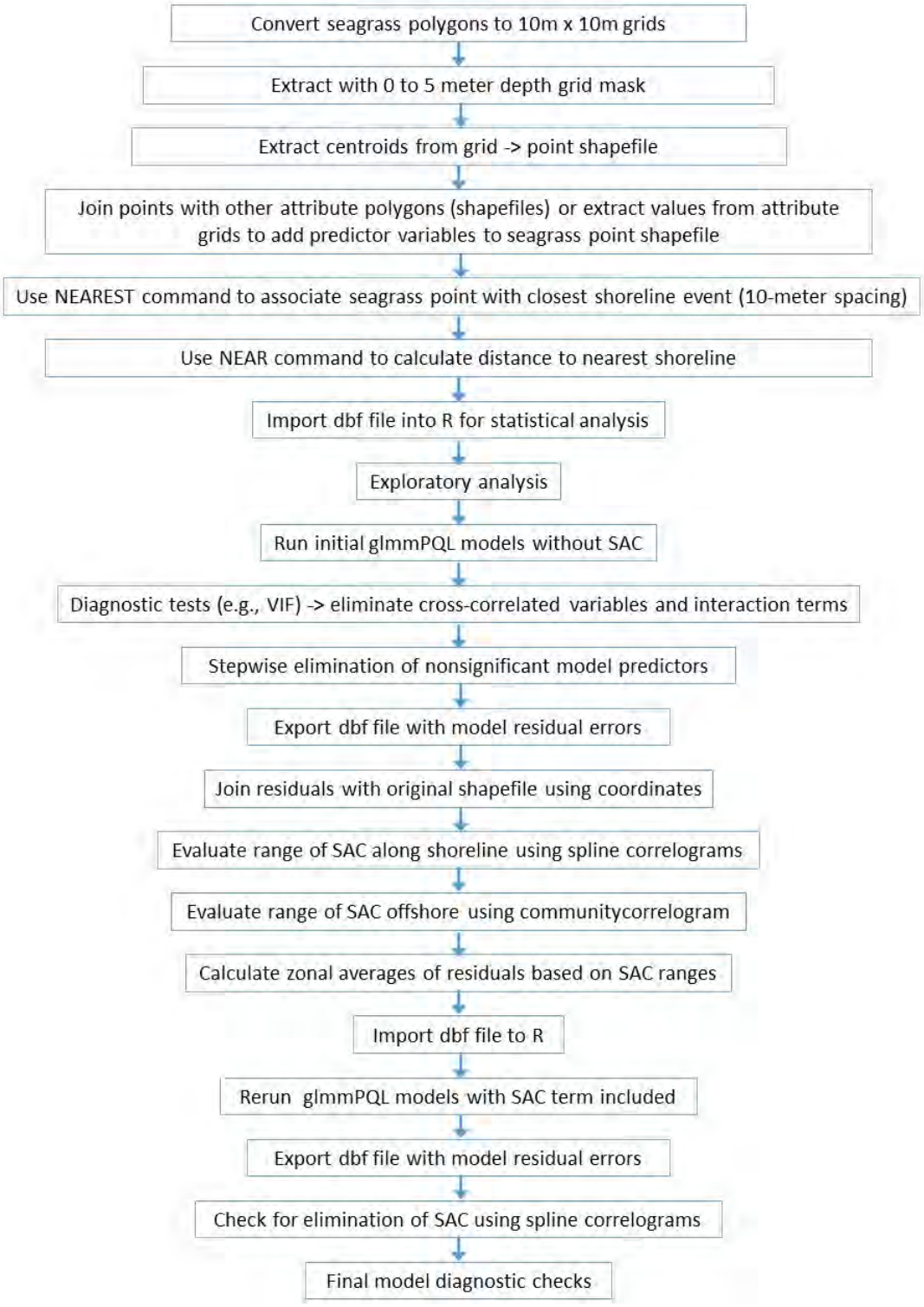

Figure C- 1 General sequence of GIS and R analyses to create generalized linear mixed models to predict seagrass presence/absence.

In the following examples, R command lines are preceded by a ">" and are typed in boldface. Responses from the console are shown in regular typeface.

C.1. Exploratory analysis (graphics package)

Spine plots

> # Explore linearity of binary responses

> spineplot(fsavcode~cSAL, data=nb9cc)

…

> spineplot(fsavcode~cUSRMARIkm2, data=nb9cc)

C.2. General linear mixed models with diagnostics plots (MASS, rms packages)

> library (MASS)

># Define formula

> fo.glmm15 <- formula(fsavcode ~ cSAL + cSAL2 + cSDavgrtrZ + cSDavgrt_1 + cSDavgrt_2 + cptTOC + cptTOC2 + cptTOC3 + fZgtMXZavT + cCG046avkm + cDstoMarin + fISOLATED + fSED4 + PResid14fa)

> # Run generalized linear mixed model

> GLMM15w.nb9cc <- glmmPQL(fo.glmm15, data = nb9cc14fa1300x200, weights = weight, random =~1|fSHORLIN, family = "binomial")

iteration 1

iteration 2

iteration 3

iteration 4

iteration 5

iteration 6

iteration 7

iteration 8

> # summarize results

> summary(GLMM15w.nb9cc)

Linear mixed-effects model fit by maximum likelihood

Data: nb9cc14fa1300x200

AIC BIC logLik

NA NA NA

Random effects:

Formula: ~1 | fSHORLIN

(Intercept) Residual

StdDev: 6.373279 0.9092619

Variance function:

Structure: fixed weights

Formula: ~invwt

Fixed effects: fsavcode ~ cSAL + cSAL2 + cSDavgrtrZ + cSDavgrt_1 + cSDavgrt_2 + cptTOC + cptTOC2 + cptTOC3 + fZgtMXZavT + cCG046avkm + cDstoMarin + fISOLATED + fSED4 + PResid14fa

	Value	Std.Error	DF	t-value	p-value
(Intercept)	0.420257	1.6594588	518856	0.25325	0.8001
cSAL	1.434985	0.1887963	518856	7.60071	0.0000
cSAL2	-0.567210	0.1552607	518856	-3.65327	0.0003
cSDavgrtrZ	2.612055	0.0536428	518856	48.69350	0.0000
cSDavgrt_1	-0.233461	0.0328527	518856	-7.10632	0.0000
cSDavgrt_2	-0.056192	0.0095881	518856	-5.86057	0.0000
cptTOC	-2.443826	0.0858941	518856	-28.45162	0.0000
cptTOC2	1.315678	0.0707202	518856	18.60401	0.0000
cptTOC3	-0.316580	0.0547752	518856	-5.77962	0.0000

fZgtMXZavT1 0.533763 0.1427521 518856 3.73909 0.0002

cCG046avkm 0.011350 0.0038698 518856 2.93287 0.0034

cDstoMarin -0.001454 0.0000522 518856 -27.86994 0.0000

fISOLATED1 -10.666443 1.0603103 518856 -10.05974 0.0000

fSED46 -1.119823 0.2133335 518856 -5.24917 0.0000

fSED47 0.632220 0.0722058 518856 8.75582 0.0000

fSED412 -0.833132 0.1516006 518856 -5.49558 0.0000

fSED4124 3.657733 0.3345729 518856 10.93254 0.0000

fSED4810 -0.377593 0.1502137 518856 -2.51370 0.0119

PResid14fa 3.172977 0.0560561 518856 56.60364 0.0000

Correlation:

(Intr) cSAL cSAL2 cSDvgZ cSDv_1 cSDv_2 cptTOC cpTOC2 cpTOC3

cSAL -0.063

cSAL2 -0.069 0.931

cSDavgrtrZ -0.007 0.093 0.048

cSDavgrt_1 -0.036 0.043 0.038 -0.324

cSDavgrt_2 0.032 -0.080 -0.065 -0.129 -0.844

cptTOC -0.007 0.170 0.141 -0.268 0.049 0.024

cptTOC2 -0.030 0.084 0.022 0.104 0.043 -0.074 0.069

cptTOC3 0.003 -0.024 0.022 0.010 -0.048 0.045 -0.636 -0.427

fZgtMXZavT1 -0.091 0.052 0.054 0.020 0.066 0.003 -0.008 0.013 -0.007

cCG046avkm -0.051 0.558 0.578 0.034 0.084 -0.071 0.056 0.091 0.016

cDstoMarin -0.016 0.318 0.331 -0.095 -0.078 0.089 0.261 -0.060 0.000

fISOLATED1 -0.194 0.040 0.048 -0.092 0.063 -0.037 0.064 -0.034 0.006

fSED46 -0.005 -0.011 0.025 0.000 -0.088 0.106 -0.049 0.066 -0.003

fSED47 -0.002 -0.239 -0.171 0.149 -0.030 -0.009 -0.097 0.103 -0.126

fSED412 -0.013 -0.062 -0.023 -0.039 0.024 -0.011 -0.109 -0.020 -0.074

fSED4124 0.003 -0.095 -0.085 0.063 -0.017 0.005 -0.097 0.031 -0.022

fSED4810 0.006 -0.029 -0.011 0.136 -0.175 0.114 0.149 -0.012 -0.128

PResid14fa 0.020 0.051 -0.031 0.485 -0.109 -0.013 -0.372 0.242 -0.060

fZMXZT cCG046 cDstMr fISOLA fSED46 fSED47 fSED412 fSED4124 fSED48

cSAL

cSAL2

cSDavgrtrZ

cSDavgrt_1

cSDavgrt_2

cptTOC

cptTOC2

cptTOC3

fZgtMXZavT1

cCG046avkm 0.114

cDstoMarin -0.006 0.296

fISOLATED1 -0.001 0.016 0.047

fSED46 -0.002 0.104 0.241 -0.026

fSED47 -0.022 -0.005 -0.089 -0.079 0.284

fSED412 -0.022 -0.012 -0.095 0.020 0.069 0.239

fSED4124 0.033 -0.050 -0.136 -0.093 0.094 0.201 0.104

fSED4810 -0.031 0.004 0.245 -0.048 0.277 0.370 0.038 0.110

PResid14fa 0.033 0.037 -0.243 -0.145 0.019 0.152 -0.063 0.107 0.118

Standardized Within-Group Residuals:

Min Q1 Med Q3 Max

-82.898446484 -0.049605288 -0.007560305 -0.001533338 361.324250323

Number of Observations: 518890

Number of Groups: 16

> fixed.effects(GLMM15w.nb9cc, family = binomial)

(Intercept) cSAL cSAL2 cSDavgrtrZ cSDavgrt_1 cSDavgrt_2 cptTOC cptTOC2 cptTOC3

0.420257164 1.434985095 -0.567209800 2.612055480 -0.233461322 -0.056191589 -2.443825859 1.315678476 -0.316580172

fZgtMXZavT1 cCG046avkm cDstoMarin fISOLATED1 fSED46 fSED47 fSED412 fSED4124 fSED4810

0.533762818 0.011349608 -0.001454164 -10.666443279 -1.119823166 0.632220198 -0.833132486 3.657733496 -0.377592950

PResid14fa

3.172976995

> # Output random effects

> random.effects(GLMM15w.nb9cc, family = binomial)

(Intercept)

-99 2.7822797

2 -9.7044680

3 -5.6673015

6 -9.6165676

7 -6.2225803

8 -4.1356439

9 0.4226314

10 3.0034085

11 -2.6426397

12 1.4561820

13 -4.3301970

14 0.2790325

15 6.6477236

16 9.7049135

17 11.4413276

18 6.5818993

Diagnostics examples

> # Check assumption of collinearity using variance inflation factor

> library(rms)

> vif(GLMM15w.nb9cc)

> # Check for heterogeneity of variance and patterns in residuals

> nb9cc7fa1300x200$PResid14wfa <- residuals(GLMM14w.nb9cc, type = "pearson")

> nb9cc7fa1300x200$Predict14wfa <- predict(GLMM14w.nb9cc, data = nb9ccfa, type = "response")

> plot(x = nb9cc7fa1300x200$Predict14wfa, y = nb9cc7fa1300x200$PResid14wfa, main = "Pearson Residuals vs Predicted")

> plot(x = nb9cc7fa1300x200$cSDavgrtrZ, y = nb9cc7fa1300x200$PResid14wfa)

> plot(x = nb9cc7fa1300x200$cSAL, y = nb9cc7fa1300x200$PResid14wfa)

> plot(x = nb9cc7fa1300x200$cptTOC, y = nb9cc7fa1300x200$PResid14wfa)

> plot(x = nb9cc7fa1300x200$cSAL, y = nb9cc7fa1300x200$PResid14wfa)

> plot(x = nb9cc7fa1300x200$cSDavgrtrZ, y = nb9cc7fa1300x200$PResid14wfa)

> spineplot(fsavcode~Predict14wfa, data=nb9cc7fa1300x200)

> # Export for calculating smoothed residual

> # Fit spline correlogram

> # Create separate dataset for each shoreline to plot spline correlograms

> nb9cc7fa1300x200.PResid14w <-data.frame(nb9cc7fa1300x200[,c(7,15,16,106)])

> str(nb9cc7fa1300x200.PResid14w)

'data.frame': 518890 obs. of 4 variables:

$ fSHORLIN : int 11 11 11 11 11 11 11 11 11 11 ...

$ ShLnDist : int 2815150 2815150 2815150 2815150 2815150 2815150 2815150 2815150 2815150 2815150 ...

$ Distance : int 0 0 0 0 0 0 0 0 0 0 ...

$ PResid14wfa: atomic -0.0186 -0.0117 -0.0114 -0.0114 -0.0114 ...

..- attr(*, "label")= chr "Standardized residuals"

C.3. General linear mixed model with interaction plots (effects package)

> fo.glm3y <- formula(savcodeAv ~ csalAv + csalAv2 + csecchiminAv + csecchiminAv2 + cpttocAv + cpttocAv2 + cZgtMXZavAv + cZgtMXZavAv2 + cDistToHdShAv + cDstoMarinaAv + cUSRMARIkm2Av + fISOLATED + cwindAv * fSED5 + PR3xFA460)

> GLM3y.avgPAbyShLnDist <- glm(fo.glm3y, data = avgPAbyShLnDistwPResid3xFA460, family = "binomial", weights = Nweight)

> # effects interaction plots for fSED5 * cwindAv

> library(effects)

> plot(effect("cwindAv:fSED5",GLM3y.avgPAbyShLnDist))

C.4. Generalized additive mixed models (mcgv package)

> fo.gamm4w4w <- formula(fsavcodeMa ~ s(csalAv) + s(csecchimin) + s(cpttocMin) + csecchimin:cpttocMin + fZgtMXZavM + cCG046avkm + cDistToHdS + cDstoMarin + cwindMin * fSED4 + PR4wFA1370)

> GAMM4wfa4w.maxPAbyShLnDist <- gamm(fo.gamm4w4w, data = maxPAbyShLnDistwPR4wFA1370m, random=list(fSHORLIN=~1),weights = weight, family = "binomial")

C.5. Spline correlogram evaluations (ncf package)

```
> # Spline correlogram

> load("D:\\savhabitat\\workingmodels\\NB9\\nb9cc14fa1300x200_ShLn14.rda")

> library(ncf)

> # Randomly select 9000 to evaluate SA at coarser scales

> nb9cc14fa1300x200.PResid15w.ShLn14.sub9000 <-
nb9cc14fa1300x200.PResid15w.ShLn14.subset[sample(1:nrow(nb9cc14fa1300x200.PResid15w.ShLn14
.subset),9000),]

># Generate spline correlogram using shoreline distance as x coordinate and setting y (Distance) to 0

> fit15wShLn14 <- spline.correlog(nb9cc14fa1300x200.PResid15w.ShLn14.sub9000$ShLnDist,
nb9cc14fa1300x200.PResid15w.ShLn14.sub9000$Distance,

+ nb9cc14fa1300x200.PResid15w.ShLn14.sub9000$PResid15wfa,

+ w = NULL, df = NULL, type = "boot",resamp = 100, npoints = 300, save = FALSE, filter = FALSE, fw = 0,
max.it = 25, xmax = 15000, latlon = FALSE, na.rm = FALSE,

+ quiet = FALSE)

1  of  100

2  of  100

…

99  of  100

100  of  100

> plot.spline.correlog(fit15wShLn14)

> # export elements of spline correlogram fit for plotting because ncf doesn't allow adjustment of y-
axis

> fit15wShLn14.predictedy <- fit15wShLn14$real$predicted$y

> write.csv(fit15wShLn14.predictedy, file =
"D:\\savhabitat\\workingmodels\\NB9\\fit15wShLn14predictedy.csv")

> fit15wShLn14.predictedx <- fit15wShLn14$real$predicted$x
```

```
> write.csv(fit15wShLn14.predictedx, file =
"D:\\savhabitat\\workingmodels\\NB9\\fit15wShLn14predictedx.csv")

> fit15wShLn14.xint <- fit15wShLn14$real$x.intercept

> write.csv(fit15wShLn14.xint, file = "D:\\savhabitat\\workingmodels\\NB9\\fit15wShLn14xint.csv")

> fit15wShLn14.bootxint <- fit15wShLn14$boot$boot.summary$x.intercept

> write.csv(fit15wShLn14.bootxint, file =
"D:\\savhabitat\\workingmodels\\NB9\\fit15wShLn14bootxint.csv")

> # bootstrap summaries - plot rows 3, 6, and 9 for 5%ile, median and 95%ile

> fit15wShLn14.bootpredy <- fit15wShLn14$boot$boot.summary$predicted$y

> write.csv(fit15wShLn14.bootpredy, file =
"D:\\savhabitat\\workingmodels\\NB9\\fit15wShLn14bootpredy.csv")

> plot(fit15wShLn14.predictedx,fit15wShLn14.predictedy, type = "l", main = "Spline Correlogram", sub
= "Model GLM15w shoreline 14 random subset of 9000", xlab = "Distance (m)",

+ ylab = "Correlation")

> lines(fit15wShLn14.predictedx,fit15wShLn14.bootpredy[3,],col="blue")

> lines(fit15wShLn14.predictedx,fit15wShLn14.bootpredy[9,],col="red")

> abline(h = 0)

> plot(fit15wShLn14.predictedx,fit15wShLn14.predictedy, type = "l", main = "Spline Correlogram", sub
= "Model GLM15.nb9cc Shoreline 14 random subset of 9000", xlab = "Distance (m)",

+ ylab = "Correlation")

> lines(fit15wShLn14.predictedx,fit15wShLn14.bootpredy[3,],col="blue")

> lines(fit15wShLn14.predictedx,fit15wShLn14.bootpredy[9,],col="red")

> abline(h = 0)
```

C.6. Community correlograms with anisotropy (CommunityCorrelogram package)

```
# Extract dataframe with x-y coordinates;

PResid4w.xy <- nb9cc.PResid4w.ShLn14.sub1200[,c(2,4)]
```

```
head(PResid4w.xy)

# Find maximum range of distances;

max(PResid4w.xy$DistToShor)

# Figure out optimum lag size and number;

# lagmin should be greater than smallest distance (10m)

# lagmax should be less than 2/3 maximum interpoint distance

lagSelect(sampleData=PResid4w,sampleLocation=cbind(PResid4w.xy,z=0),sampleTime=NULL,Location
Names=NULL,lagmin=11,lagmax=250,by=30,option=1,plot=T,anisotropic=T,azimuth=90,azimuthTol=0,
bandwidth=0,dipAngle=0,dipTol=0,dipBandwidth=0,distmeth='euclidean')

commcorrelogram(sampleData=PResid4w,sampleTime=NULL,sampleLocation=cbind(PResid4w.xy,z=0)
,LocationNames=NULL,option=1,metric='mantel',lagNumber=25,lagSize=11,lagTol =
5.5,numTests=99,anisotropic=TRUE,azimuth=90,azimuthTol=0,bandwidth=0,dipAngle=0,dipTol=0,dipB
andwidth=0,distmeth='euclidean',mantmeth='spearman',adj='holm',prog=TRUE,alternative='one.side
d')
```

C.7. Cross-validation and ROC construction (pROC package)

Subsetting data and generating training and test data sets

```
> library(MASS)

> # XVAL_100514.R Cross-validation for final models

> load(file = "D:\\savhabitat\\workingmodels\\NB9\\nb9cc14fa1300x200.RDA")

> load(file = "D:\\savhabitat\\workingmodels\\NB9\\GLMM15w_nb9cc.RDA")

> # First sample each random effects group (SHORLIN) separately and combine rows, then repeat 10
times

> df.0 <- nb9cc14fa1300x200 # original data frame

> df.2 <- df.0[df.0$fSHORLIN == 2,]

> df.7 <- df.0[df.0$fSHORLIN == 7,]

> df.6 <- df.0[df.0$fSHORLIN == 6,]

> df.9 <- df.0[df.0$fSHORLIN == 9,]

> df.14 <- df.0[df.0$fSHORLIN == 14,]
```

```
> df.8 <- df.0[df.0$fSHORLIN == 8,]
> df.10 <- df.0[df.0$fSHORLIN == 10,]
> df.12 <- df.0[df.0$fSHORLIN == 12,]
> df.13 <- df.0[df.0$fSHORLIN == 13,]
> df.3 <- df.0[df.0$fSHORLIN == 3,]
> df.17 <- df.0[df.0$fSHORLIN == 17,]
> df.11 <- df.0[df.0$fSHORLIN == 11,]
> df.15 <- df.0[df.0$fSHORLIN == 15,]
> df.18 <- df.0[df.0$fSHORLIN == 18,]
> df.16 <- df.0[df.0$fSHORLIN == 16,]
> df.99 <- df.0[df.0$fSHORLIN == -99,]
>
> # create subsample 1 of 1/10 each SHORLIN n set with replacement
> df.2.1 <- df.2[sample(1:163486,size = 16347, replace = TRUE),]
> df.7.1 <- df.7[sample(1:84233,size = 8423, replace = TRUE),]
> df.6.1 <- df.6[sample(1:63230,size = 6323, replace = TRUE),]
> df.9.1 <- df.9[sample(1:58307,size = 5831, replace = TRUE),]
> df.14.1 <- df.14[sample(1:48586,size = 4859, replace = TRUE),]
> df.8.1 <- df.8[sample(1:44745,size = 4475, replace = TRUE),]
> df.10.1 <- df.10[sample(1:22029,size = 2203, replace = TRUE),]
> df.12.1 <- df.12[sample(1:11176,size = 1118, replace = TRUE),]
> df.13.1 <- df.13[sample(1:5389,size = 539, replace = TRUE),]
> df.3.1 <- df.3[sample(1:4768,size = 477, replace = TRUE),]
> df.17.1 <- df.17[sample(1:3189,size = 319, replace = TRUE),]
> df.11.1 <- df.11[sample(1:3025,size = 303, replace = TRUE),]
```

> df.15.1 <- df.15[sample(1:2354,size = 235, replace = TRUE),]

> df.18.1 <- df.18[sample(1:2135,size = 214, replace = TRUE),]

> df.16.1 <- df.16[sample(1:1172,size = 117, replace = TRUE),]

> df.99.1 <- df.99[sample(1:1084,size = 108, replace = TRUE),]

> dftest.1 <-
rbind(df.2.1,df.7.1,df.6.1,df.9.1,df.14.1,df.8.1,df.10.1,df.12.1,df.13.1,df.3.1,df.17.1,df.11.1,df.15.1,df.18.1,df.16.1,df.99.1)

> dftest.1$TEST1 <- 1

> TEST1.JOINID <- dftest.1[,c("JOINID","TEST1")]

> dfall.1 <- merge(df.0,TEST1.JOINID,all.x = TRUE)

> dftrain.1 <- dfall.1[-which(dfall.1$TEST1 == 1),]

…

> # subsample 10

> df.2.10 <- df.2[sample(1:163486,size = 16347, replace = TRUE),] # create subsample 1 of 1/10 SHORLIN 2 set with replacement

> df.7.10 <- df.7[sample(1:84233,size = 8423, replace = TRUE),]

> df.6.10 <- df.6[sample(1:63230,size = 6323, replace = TRUE),]

> df.9.10 <- df.9[sample(1:58307,size = 5831, replace = TRUE),]

> df.14.10 <- df.14[sample(1:48586,size = 4859, replace = TRUE),]

> df.8.10 <- df.8[sample(1:44745,size = 4475, replace = TRUE),]

> df.10.10 <- df.10[sample(1:22029,size = 2203, replace = TRUE),]

> df.12.10 <- df.12[sample(1:11176,size = 1118, replace = TRUE),]

> df.13.10 <- df.13[sample(1:5389,size = 539, replace = TRUE),]

> df.3.10 <- df.3[sample(1:4768,size = 477, replace = TRUE),]

> df.17.10 <- df.17[sample(1:3189,size = 319, replace = TRUE),]

> df.11.10 <- df.11[sample(1:3025,size = 303, replace = TRUE),]

> df.15.10 <- df.15[sample(1:2354,size = 235, replace = TRUE),]

```
> df.18.10 <- df.18[sample(1:2135,size = 214, replace = TRUE),]

> df.16.10 <- df.16[sample(1:1172,size = 117, replace = TRUE),]

> df.99.10 <- df.99[sample(1:1084,size = 108, replace = TRUE),]

> dftest.10 <-
rbind(df.2.10,df.7.10,df.6.10,df.9.10,df.14.10,df.8.10,df.10.10,df.12.10,df.13.10,df.3.10,df.17.10,df.11.
10,df.15.10,df.18.10,df.16.10,df.99.10)

> dftest.10$TEST10 <- 1

> TEST10.JOINID <- dftest.10[,c("JOINID","TEST10")]

> dfall.10 <- merge(df.0,TEST10.JOINID,all.x = TRUE)

> dftrain.10 <- dfall.10[-which(dfall.10$TEST10 == 1),]

> # Save training and test data sets so that they can be selectively reloaded to conserve memory for
model runs

> save(dftrain.1,file = "D:\\savhabitat\\workingmodels\\NB9\\dftrain1.rda")

> save(dftest.1,file = "D:\\savhabitat\\workingmodels\\NB9\\dftest1.rda")

> save(dftrain.2,file = "D:\\savhabitat\\workingmodels\\NB9\\dftrain2.rda")

> save(dftest.2,file = "D:\\savhabitat\\workingmodels\\NB9\\dftest2.rda")

> save(dftrain.3,file = "D:\\savhabitat\\workingmodels\\NB9\\dftrain3.rda")

> save(dftest.3,file = "D:\\savhabitat\\workingmodels\\NB9\\dftest3.rda")

> save(dftrain.4,file = "D:\\savhabitat\\workingmodels\\NB9\\dftrain4.rda")

> save(dftest.4,file = "D:\\savhabitat\\workingmodels\\NB9\\dftest4.rda")

> save(dftrain.5,file = "D:\\savhabitat\\workingmodels\\NB9\\dftrain5.rda")

> save(dftest.5,file = "D:\\savhabitat\\workingmodels\\NB9\\dftest5.rda")

> save(dftrain.6,file = "D:\\savhabitat\\workingmodels\\NB9\\dftrain6.rda")

> save(dftest.6,file = "D:\\savhabitat\\workingmodels\\NB9\\dftest6.rda")

> save(dftrain.7,file = "D:\\savhabitat\\workingmodels\\NB9\\dftrain7.rda")

> save(dftest.7,file = "D:\\savhabitat\\workingmodels\\NB9\\dftest7.rda")

> save(dftrain.8,file = "D:\\savhabitat\\workingmodels\\NB9\\dftrain8.rda")
```

> save(dftest.8,file = "D:\\savhabitat\\workingmodels\\NB9\\dftest8.rda")

> save(dftrain.9,file = "D:\\savhabitat\\workingmodels\\NB9\\dftrain9.rda")

> save(dftest.9,file = "D:\\savhabitat\\workingmodels\\NB9\\dftest9.rda")

> save(dftrain.10,file = "D:\\savhabitat\\workingmodels\\NB9\\dftrain10.rda")

> save(dftest.10,file = "D:\\savhabitat\\workingmodels\\NB9\\dftest10.rda")

> save.image("D:\\savhabitat\\workingmodels\\NB9\\101214a_xval.RData")

> # For each set run model with training set and predict results for test set

…

> # Repeat for sample 10;

> load("D:\\savhabitat\\workingmodels\\NB9\\dftrain10.rda")

> load("D:\\savhabitat\\workingmodels\\NB9\\dftest10.rda")

> # Run models first using original PResid zonal averages;

> # Run model foglmm15 with training sample 10 dftrain.10

> GLMM15w.dftrain.10 <- glmmPQL(fo.glmm15, data = dftrain.10, weights = weight, random =~1|fSHORLIN, family = "binomial")

iteration 1

iteration 2

iteration 3

iteration 4

iteration 5

iteration 6

iteration 7

iteration 8

> # Generate predictions with test sample 10, type = response dftest.10

> dftest.10$Predict15 <- predict(GLMM15w.dftrain.10, newdata = dftest.10, type = "response")

> # Generate raw residuals with test sample 10 after converting factor back to original numeric value

> dftest.10$Resid15 <- as.numeric(levels(dftest.10$fsavcode))[dftest.10$fsavcode] - dftest.10$Predict15

> save(dftest.10,file = "D:\\savhabitat\\workingmodels\\NB9\\dftest10xval.rda")

> # Combine predicted values, residuals and calculate mean residual, MSE sum

> load("D:\\savhabitat\\workingmodels\\NB9\\dftest1xval.rda")

> load("D:\\savhabitat\\workingmodels\\NB9\\dftest2xval.rda")

> load("D:\\savhabitat\\workingmodels\\NB9\\dftest3xval.rda")

> load("D:\\savhabitat\\workingmodels\\NB9\\dftest4xval.rda")

> load("D:\\savhabitat\\workingmodels\\NB9\\dftest5xval.rda")

> load("D:\\savhabitat\\workingmodels\\NB9\\dftest6xval.rda")

> load("D:\\savhabitat\\workingmodels\\NB9\\dftest7xval.rda")

> load("D:\\savhabitat\\workingmodels\\NB9\\dftest8xval.rda")

> load("D:\\savhabitat\\workingmodels\\NB9\\dftest9xval.rda")

> load("D:\\savhabitat\\workingmodels\\NB9\\dftest10xval.rda")

> #Append test files

> dftest.1to10 <- rbind(dftest.1[,-113],dftest.2[,-113],dftest.3[,-113],dftest.4[,-113],dftest.5[,-113],dftest.6[,-113],dftest.7[,-113],dftest.8[,-113],dftest.9[,-113],

+ dftest.10[,-113])

>#Calculate summary statistics for residuals and squared residuals

> dftest.1to10$Resid15_2 <- dftest.1to10$Resid15 * dftest.1to10$Resid15

> summary(dftest.1to10$Resid15)

```
    Min.   1st Qu.    Median      Mean   3rd Qu.      Max.      NA's
-1.000000 -0.073900 -0.001545 -0.128500 -0.000043  1.000000        18
```

> summary(dftest.1to10$Resid15_2)

```
   Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
0.000000 0.000000 0.000004 0.117600 0.009783 1.000000       18
```

ROC statistics

> **library(pROC)**

Type 'citation("pROC")' for a citation.

Attaching package: ‘pROC’

The following objects are masked from ‘package:stats’:

cov, smooth, var

> **roc(dftest.1to10$fsavcode,dftest.1to10$Predict15, plot = TRUE)**

Call:

roc.default(response = dftest.1to10$fsavcode, predictor = dftest.1to10$Predict15, plot = TRUE)

Data: dftest.1to10$Predict15 in 506133 controls (dftest.1to10$fsavcode 0) < 12759 cases (dftest.1to10$fsavcode 1).

Area under the curve: 0.7144

Appendix D. Exploratory Analyses and Diagnostic Tests

D.1. Models for Seagrass Grid Cell Presence/Absence

D.1.1. Exploratory Analyses

Exploratory analyses were conducted to evaluate the distribution of predictor variables, e.g., to determine if there are potential problems with outliers. Figures D1a-r illustrate the distribution of continuous variables used in model development. With the exception of wave energy, variables exhibit few influential outliers. If necessary, a square root transformation could be applied to the wave energy variable to even out the distribution; however that would change the form of the response to wave energy.

Conditional plots for the binary response variable as a function of predictors (spine plots) provide information on potential nonlinearities in response (before the effect of other variables are factored out). Spine plots show conditional probabilities of seagrass presence as a function of predictor variables, segmented into groups (Figures D2a-l). The width of each band is inversely proportional to the number of observations in that band. The inverted light color bands at the top of plots represent the relative frequency of occurrence of current (2006) eelgrass and the darker gray band segments at the bottom represent the relative frequency of current eelgrass absence[1]. Most of the continuous potential predictors appear to be related linearly to the relative frequency of seagrass occurrence, with two exceptions. The response to wind-generated wave mixing depth appears to be unimodal, with maximum response at intermediate values. The effect of distance from historic seagrass patch edge appears to drop off exponentially rather than linearly. Seagrass probability of occurrence appears to vary among shoreline segments. Counter to our initial hypothesis, probability of seagrass appears to be greater along isolated shoreline segments. Seagrass shows evidence of persistence, with greater probability of occurrence in areas of historic occurrence, and lessening likelihood of occurrence as distance from historic patch edges increases.

[1] R also includes the cdplot function which will produce a smoothed version of conditional probability plots, but these are less useful because they can include artifacts where there is sparse representation of points along a gradient.

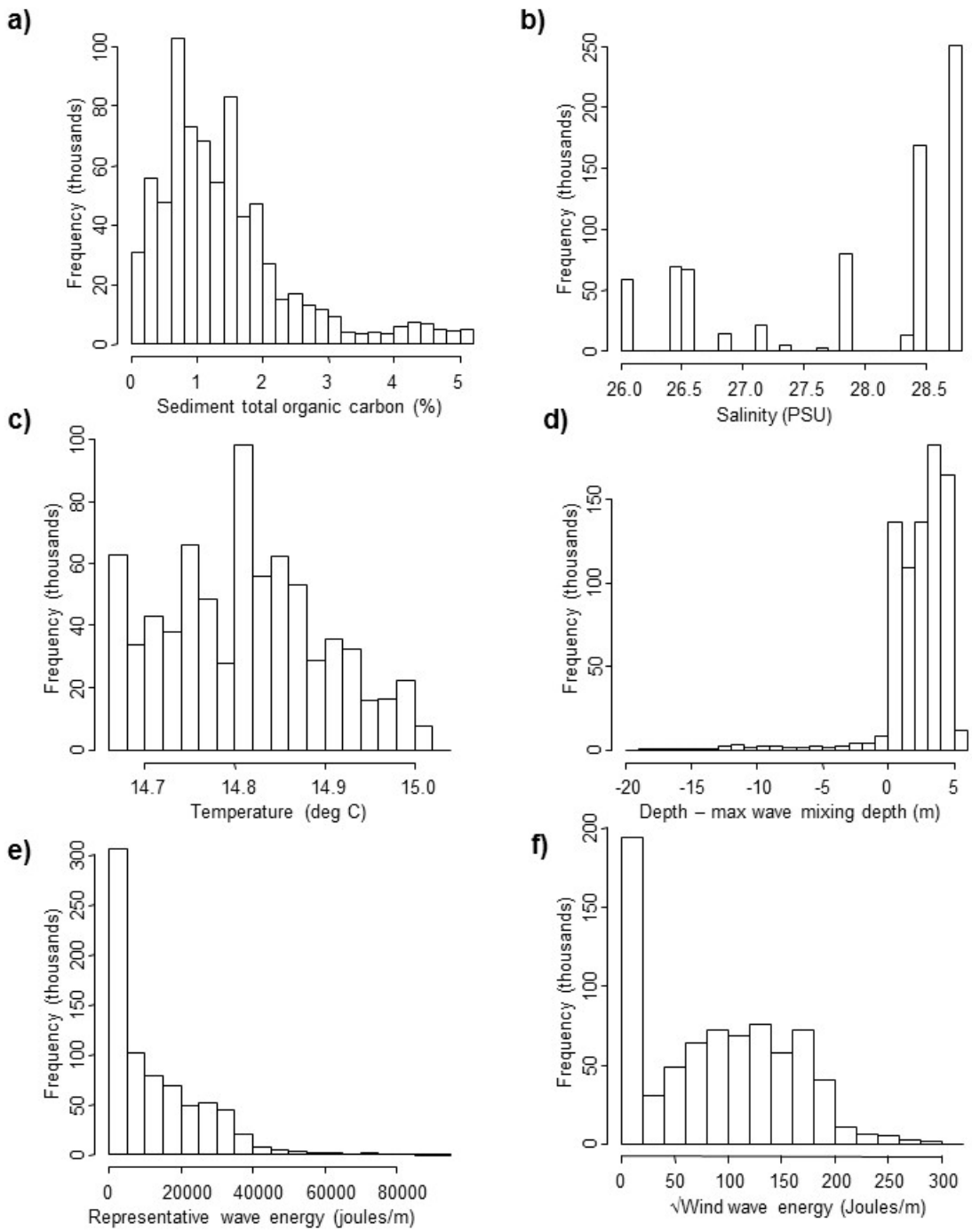

Figure D-1. Distribution of independent continuous variables entered into general linear models and general additive models: a) sediment percent total organic carbon; b) salinity (PSU); c) temperature (deg C); d) Depth minus max wave mixing depth (m); e) representative wave energy (joules/m); f) square root wind wave energy (Joules/m); g) Avg Secchi depth minus water depth; h) Minimum Secchi depth minus water depth; i) Canada goose (numbers/km²); j) $\log_{10}$ (Canada goose density/km² + 1); k) distance to hardened shoreline (m); l) √distance to hardened shoreline (m); m) density of unsewered residences on high infiltration soils (no/km²); n) distance to nearest marina (m); o) √distance to nearest marina; p) area of 1999 seagrass patch (m2); q) distance to edge of nearest 1999 seagrass patch (m); r) log10(max wave mixing depth/water depth).

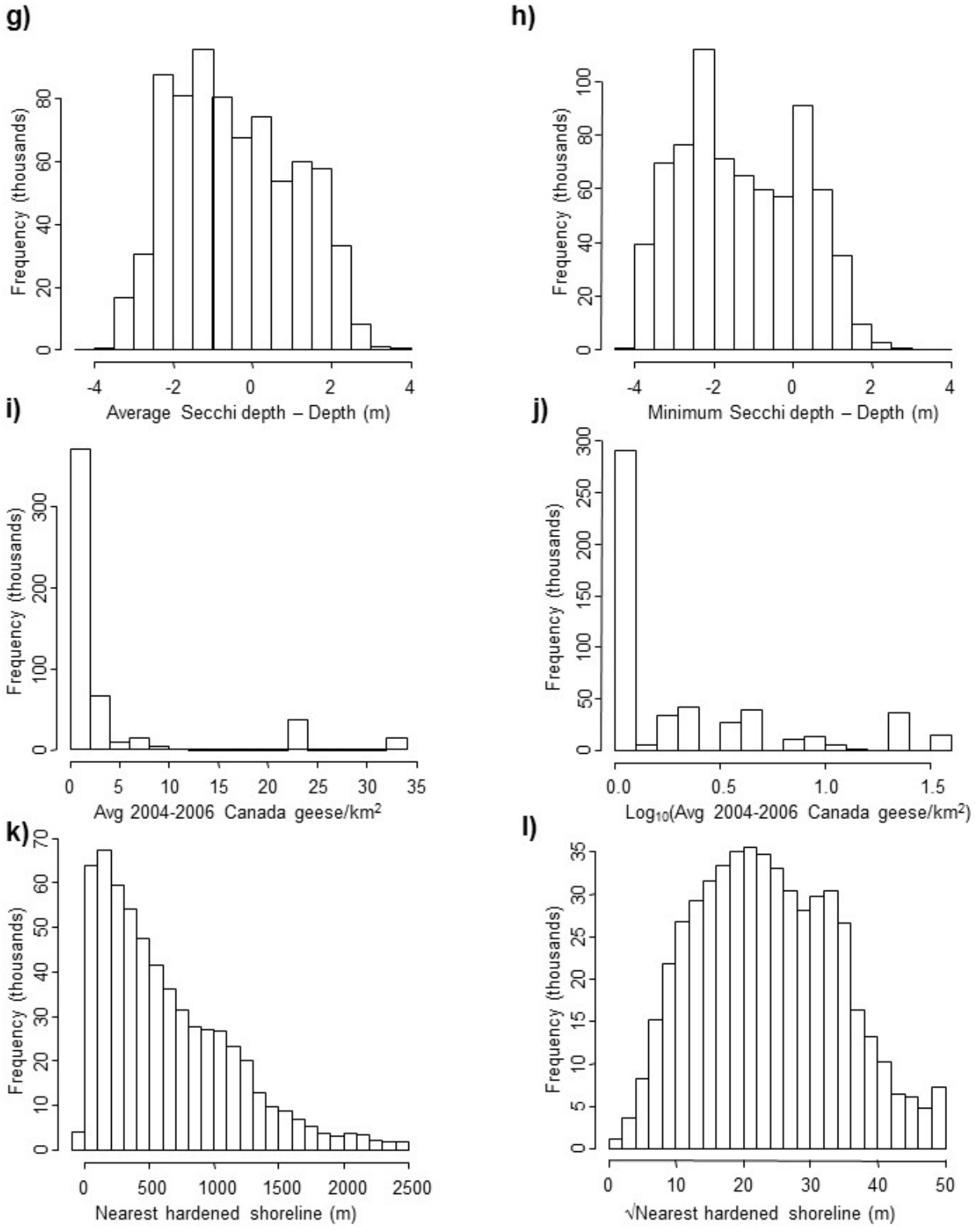

Figure D-1 (cont'd)

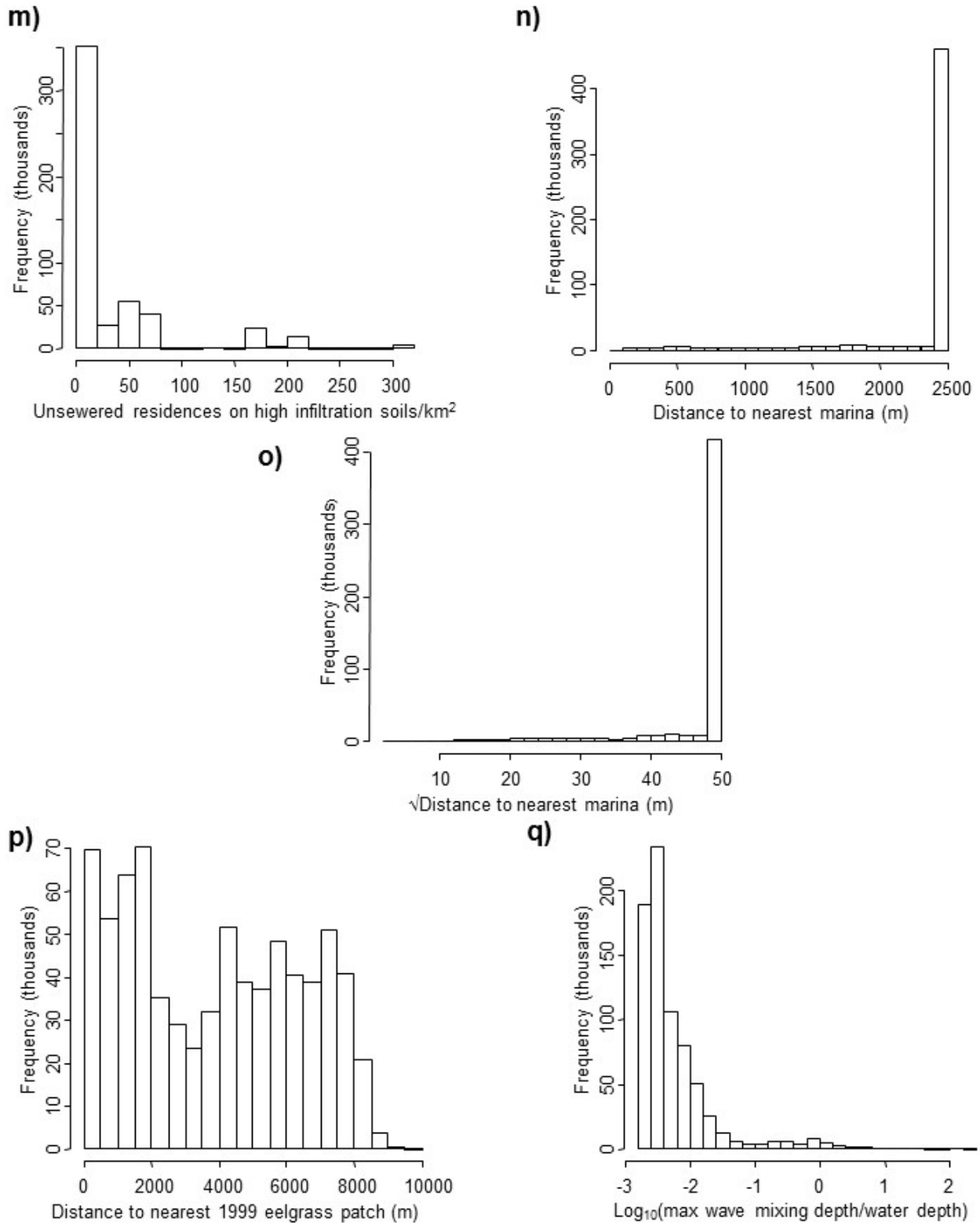

Figure D-1 (cont'd)

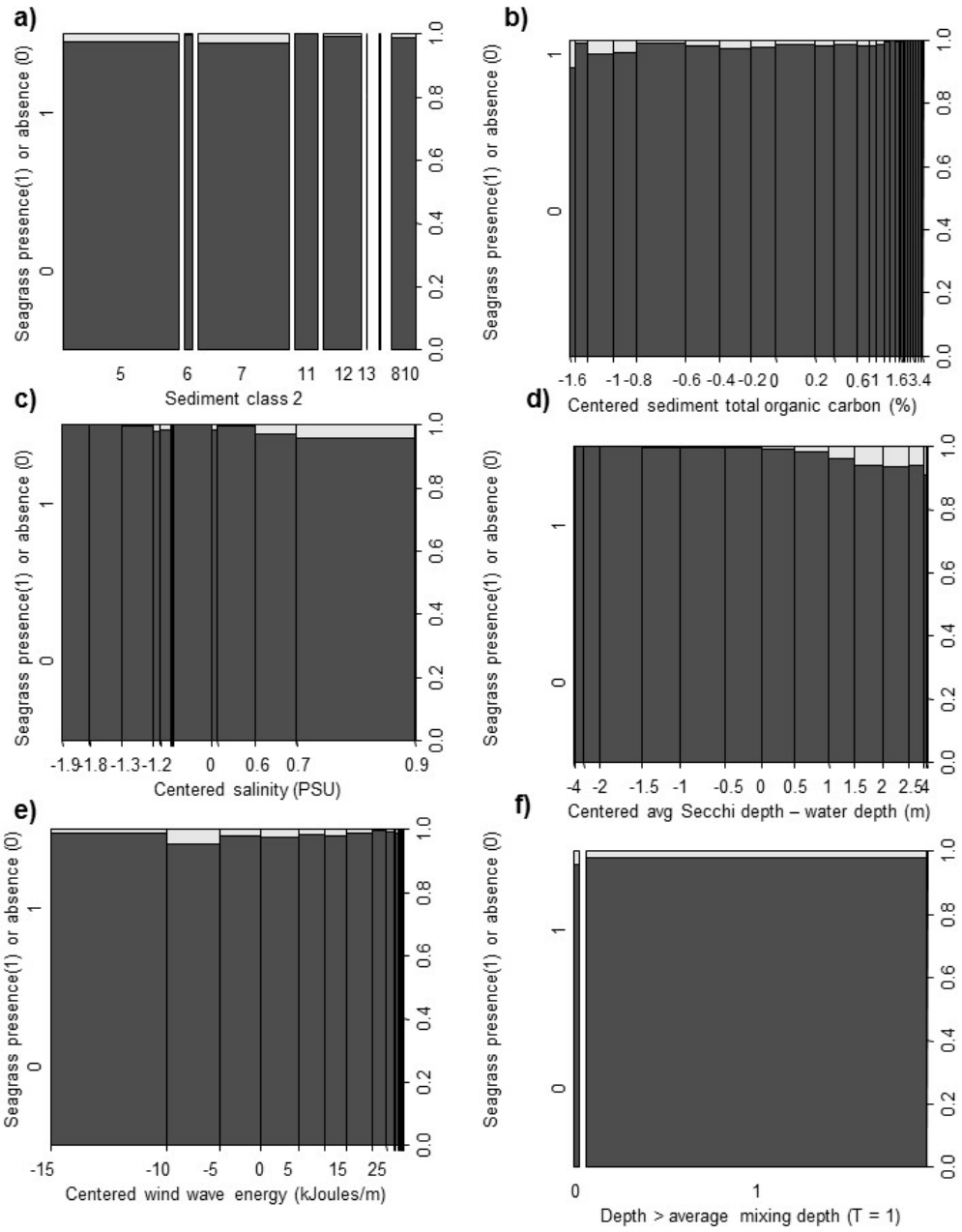

Figure D-2. Spine plots showing conditional probabilities of seagrass presence as a function of potential independent predictor variables. Width of band is inversely proportional to number of observations in band. Inverted light color bands at top of plots represent relative frequency of occurrence of current (2006) eelgrass and darker gray band segments at bottom represent relative frequency of current eelgrass absence : a) Sediment particle-size class (5 = sand, 6 = gravelly-sand, 7= silty-sand, 11 = clay-silt, 12 = sand-silt-clay, 13 = gravel-silt-clay, 1+2+4 = gravel + sandy gravel + gravel-sand-silt, 8+10= silty+sandy-silt); b) Sediment percent total carbon; c) centered salinity (PSU); d) centered average Secchi depth – water depth (m); e) centered wave energy(kJoules/m); f) dummy variable indicating water depth in excess of wave mixing depth; g) dummy variable indicating isolated shoreline segment in middle of channel; h) shoreline segment in Narragansett Bay (see Figure 3); i) dummy variable indicating historic (1996-97) eelgrass presence/absence; j) centered distance to historic (1996-97) eelgrass edge; k) distance to hardened shoreline (km); l) area of 1999 seagrass patch.

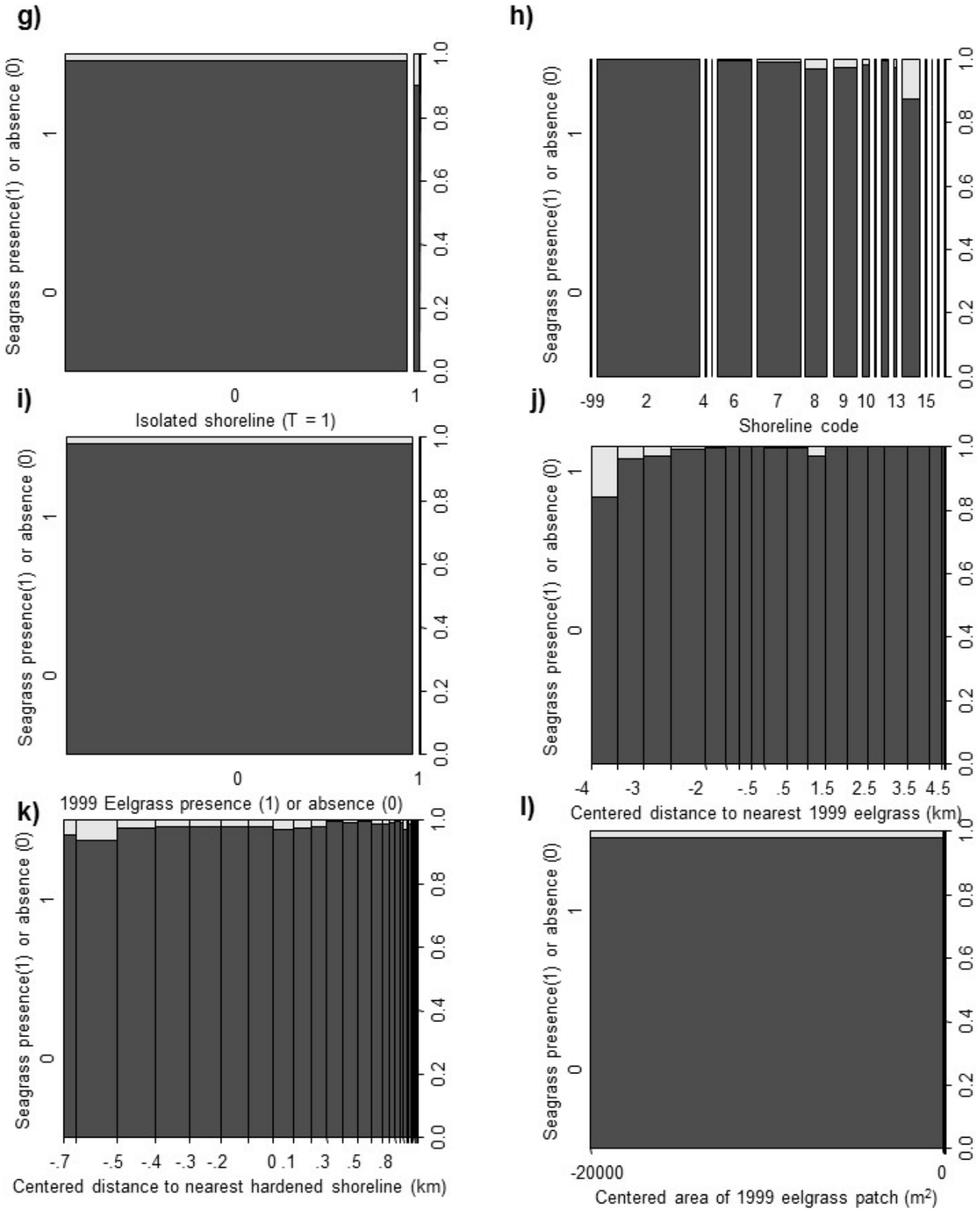

Figure D-2. (Cont'd)

D.1.2. Evaluation of Model Assumptions in Preliminary Seagrass Grid P/A Models

We performed diagnostic tests to check on validity of model assumptions for the preliminary seagrass grid presence/absence models, e.g., multicollinearity, heterogeneity of variance, random distribution of model residuals, and spatial independence of model residuals. We evaluated an initial generalized linear mixed model (GLMM) (Table 3) to predict seagrass presence at the scale of 10-meter grid cells, with shoreline included as a random effect and all potential predictors (Table D-1) and interaction terms included. To avoid problems with multi-collinearity of independent variables we modified the initial model to remove all interaction terms except for Wave Energy x Sediment Type. A review of the working model residuals showed multiple problems with heterogeneity of variance and potential nonlinearities (Figure D-3a-c). In general, values of predictor variables associated with high probability of seagrass occurrence also showed a greater range of Pearson residuals. (Pearson residuals are adjusted to account for higher expected variance with the mean, so should show no pattern when graphed against predicted values or against each predictor.)

Diagnostic tests showed evidence of spatial autocorrelation for working model residuals. A run of the model using the quasibinomial family in place of the binomial family demonstrated that errors were underdispersed, with a dispersion factor of 0.71. Normally a binomial model is assumed to have a dispersion value of one because, by definition, the variance is equal to the mean. The spline correlogram constructed based on Pearson residuals for a random subset of 9000 observations from the working model showed evidence of spatial autocorrelation at fine scales, with only small levels of autocorrelation at coarser scales (Figure D-4a). We first used horizontal swaths of 9000 points from each of the main branches of Narragansett Bay to fine-tune the estimate of the range of spatial autocorrelation at short spatial scales. The lower limit of the 95% confidence intervals of the first x-intercept from bootstrapping spline fits ranged from 345 to 3428 meters (Figure D-4b). Note that reconstruction of spline correlograms from horizontal swaths suggested that spatial autocorrelation occurred not only at fine scales but also at intermediate and extreme distance ranges. Because these did not appear in the correlogram constructed from points sampled from the entire range, we assumed that these zones of apparent spatial autocorrelation were an artifact of subsampling.

As explained earlier, we used spline correlograms based on shoreline distance to evaluate the range of spatial autocorrelation parallel to shorelines, but used Mantel correlograms (using the R commcorrelogram package) to evaluate the range of autocorrelation along depth gradients perpendicular to the shoreline. Due to memory constraints, we ran the communitycorrelogram function for a random subset of points for one shoreline at a time. Figure D-5 shows the Mantel correlogram for Shoreline 14. Spatial autocorrelation was not significant at distances greater than 200 meters in the offshore direction.

Table D-1. Potential independent variables included in models predicting seagrass presence.

Variable	Definition	Units
fSHORLIN	Shoreline code	-99, 1-19
cSAL	Centered growing season salinity	PSU
cTEMPER	Centered growing season average water temperature	Deg C
fSEDn	Sediment type (n represents level of lumping)	1-13
cWIND	Wind wave energy	Joules/m
cPTTOC	Centered sediment percent total organic carbon	%
cSDavggtrZ	Centered growing season average Secchi depth – water depth	Meters
cSDmngtrZ	Centered growing season minimum Secchi depth – water depth	Meters
fZgtrMXZ	Depth greater than wave mixing depth (0 = FALSE, 1 = TRUE)	
cZgtrMXZ	Centered water depth greater than wave mixing depth	Meters
fISOLATED	Isolated shoreline (0 = FALSE, 1=TRUE)	
cDistHdShor	Centered distance to hardened shoreline	Meters
cDistMarina	Centered distance to nearest marina	Meters
cUSRMARIkm2	Centered unsewered residences on high infiltration soils/catchment area	#/km^2
cCG046avkm2	Centered winter 2004-2006 Canada goose density	#/km^2
fEGPA99	Historic (1999) eelgrass presence (0 = FALSE, 1 = TRUE)	
cAREA	Centered area of 1999 eelgrass patch	Meters^2
cDistEG99	Centered distance to edge of nearest 1999 eelgrass patch	Meters

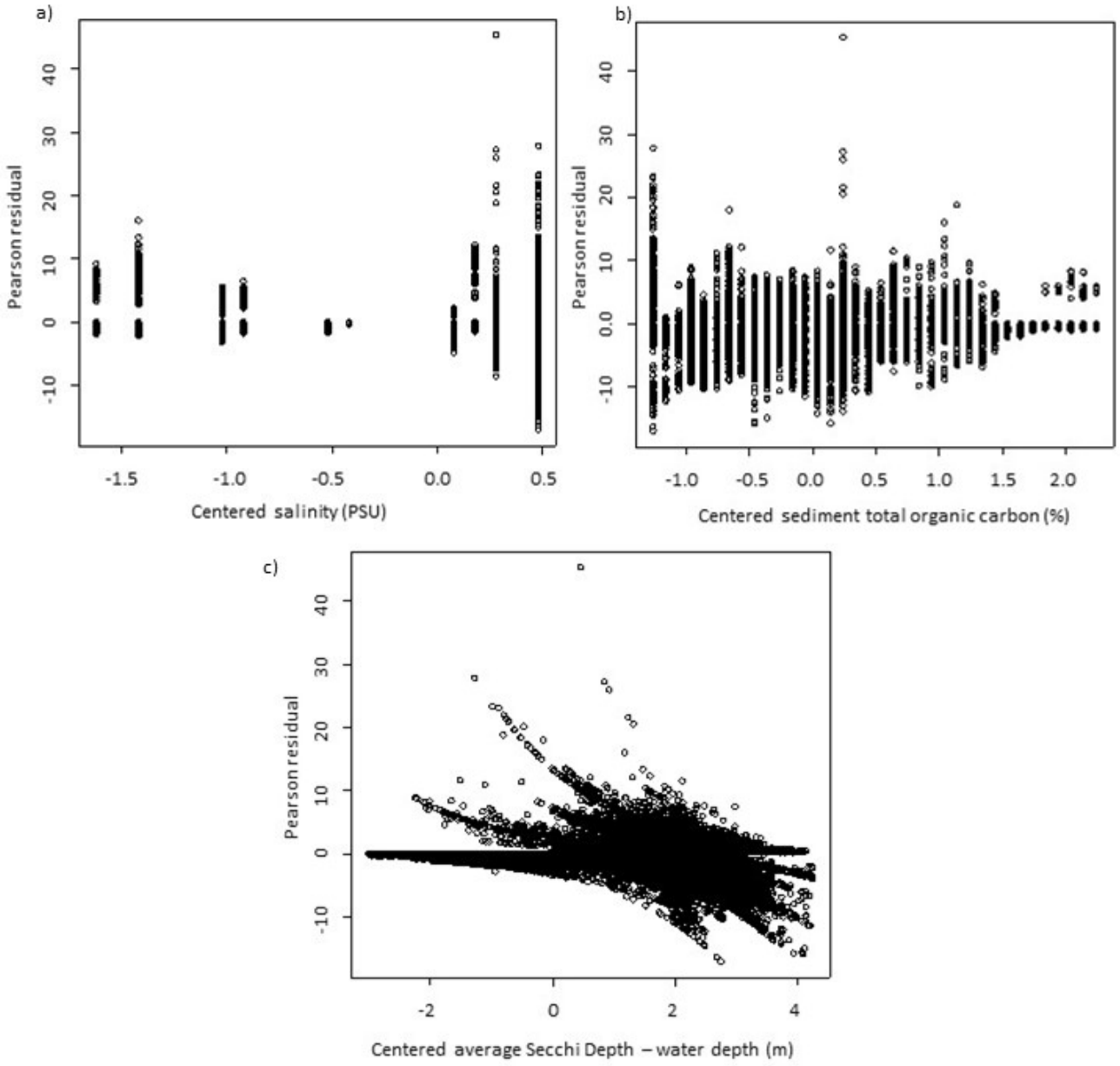

Figure D-3. Heterogeneity of variance for residuals of initial GLMM model to predict seagrass presence at the scale of 10 meter grid cells. Residuals show evidence of heterogeneity of variance.

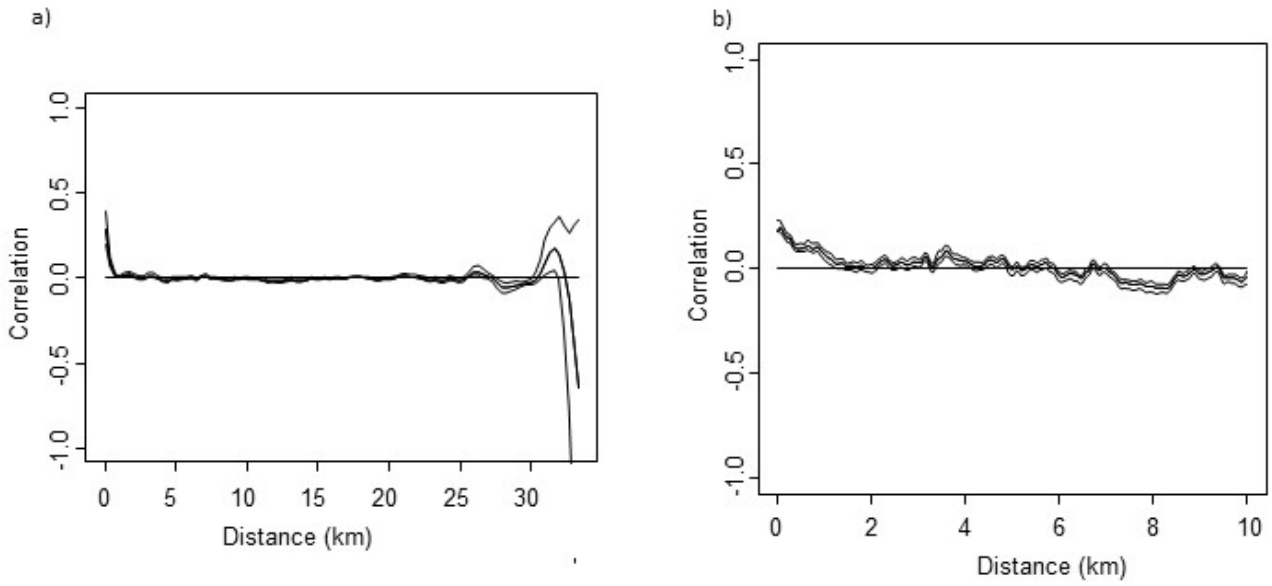

Figure D-4. Spline correlogram with 95% confidence intervals from bootstrapping based on Pearson residuals from working version of general linear mixed model. a) Random subset of 9000 points from all shorelines. b) Random subset of 9000 points from shoreline 14.

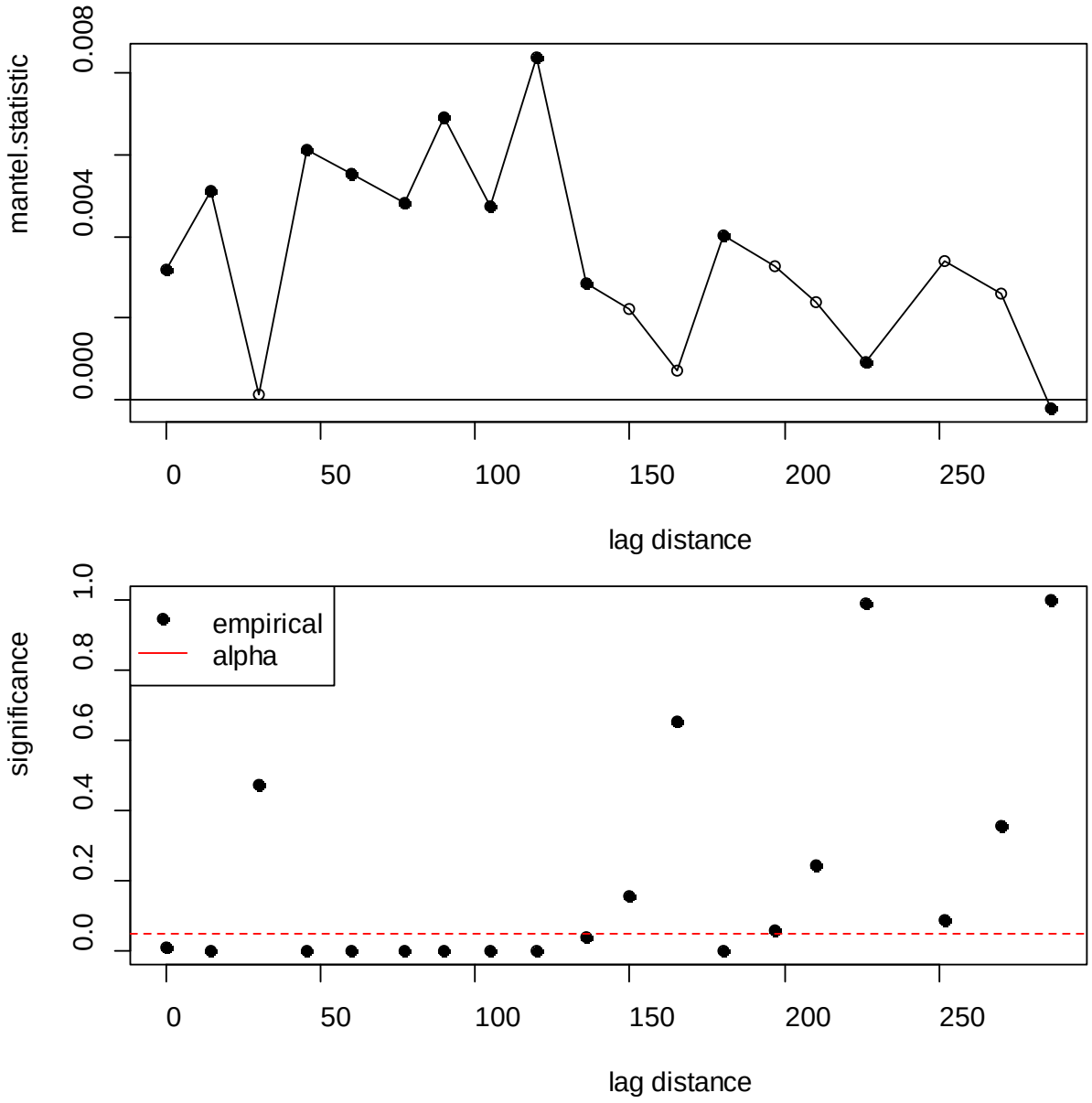

Figure D-5. Correlogram showing spatial autocorrelation for offshore distance based on residuals from model predicting seagrass presence/absence for Shoreline 14. Spatial autocorrelation is not significant at distances greater than 200 meters.

D.2. Models for Shoreline Segment Presence/Absence

D.2.1. Exploratory Analyses of Shoreline P/A Relationships

Figures D-6a-p display exploratory spine plots showing general relationships between single predictors and seagrass presence/absence by shoreline distance. Percent shoreline occupied by seagrass ranges from less than 1 percent (shoreline 3) to almost 60% (shoreline 17), with occupancy rates almost twice as high for isolated as compared to main shorelines (Figures D-6a, b). Shoreline occupancy ranges from almost 60 to over 80 percent for zones near larger 1999 eelgrass patches, and falls off exponentially with distance to nearest 1999 eelgrass patch (Figures D-6d). As a single predictor Secchi depth performs poorly along the entire gradient, although occupancy rates are greater than 60% for the highest decile of minimum transparency values (Figures D-6e, f). Shoreline occupancy shows an apparent unimodal response to salinity and percent sediment total organic carbon, but does not appear to vary with average temperature (Figures D-6g, h, i). Occupancy increases steadily as the proportion of depths greater than average wave mixing depths increases. Shoreline occupancy is low for gravel- dominated sediment classes (1-2-4 and 6) (Figure D-6l). Occupancy is highest for the top four deciles of distance to hardened shoreline (Figure D-6o).

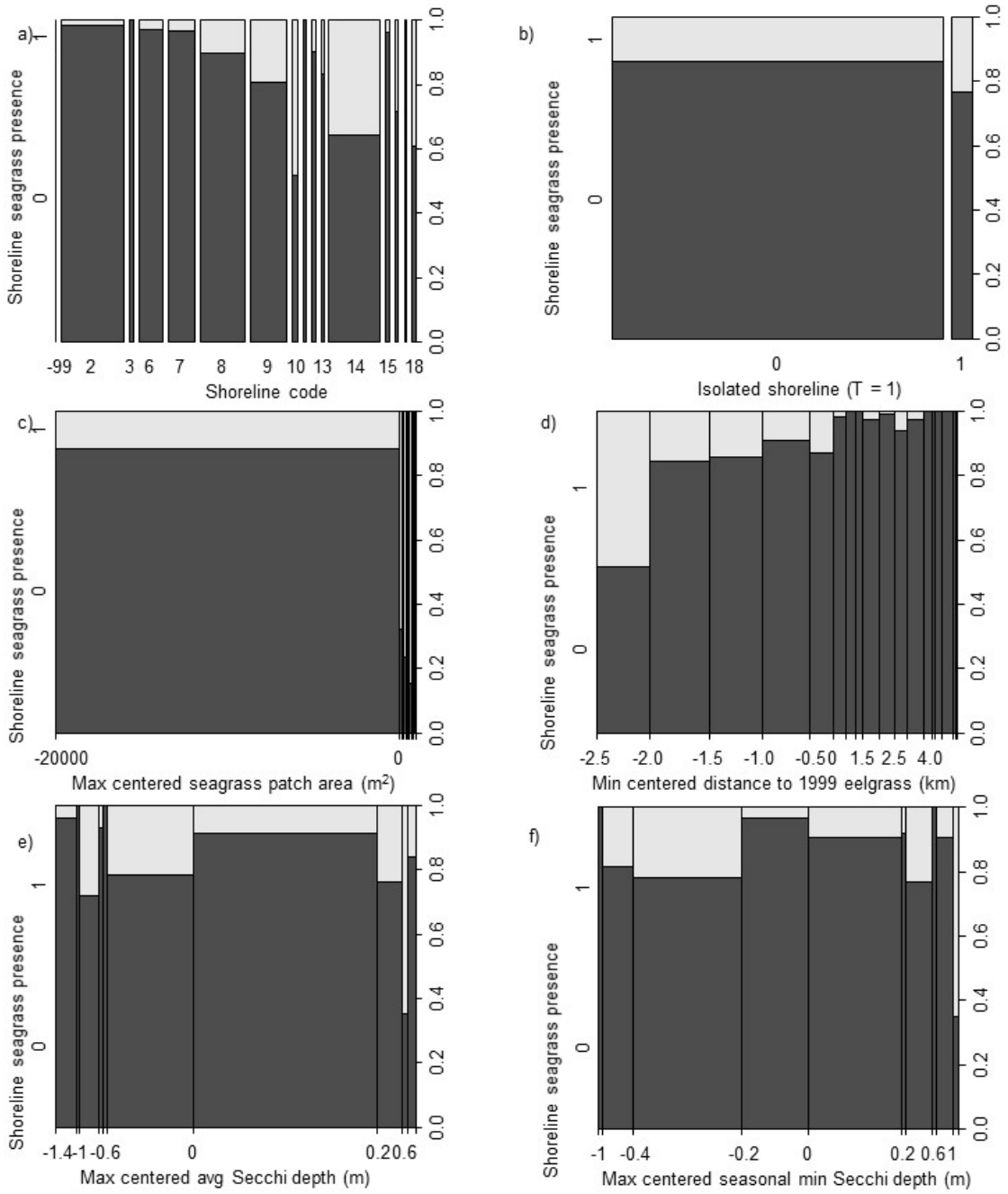

Figure D-6. Spine plots showing conditional probabilities of seagrass shoreline presence (1) versus absence (0) as a function of potential independent predictor variables: a) Shoreline (see Figure 3); b) shoreline isolation; c) centered maximum 1999 eelgrass patch size (m^2); d) centered min distance to 1999 eelgrass patch (km); e) maximum centered time-averaged Secchi depth at i (m); f) maximum centered seasonal-minimum Secchi depth (m); g) centered average salinity (PSU); h) minimum centered average temperature (deg C); i) minimum centered sediment percent total organic carbon; j) minimum centered unsewered residences on high infiltration soils/km^2 watershed area; k) minimum centered wind energy (kJoules/m); l) maximum centered depth - average wave mixing depth; m) dominant sediment particle-size class; n) minimum centered average Canada goose density (geese/km2); o) maximum centered distance to hardened shoreline; and p) maximum centered distance to marina (km). Width of band is proportional to number of observations in band. Inverted light color bands at top of plots represent relative frequency of occurrence of current (2006) eelgrass and darker gray band segments at bottom represent relative frequency of current eelgrass absence at shoreline unit i. For continuous variables, values are grouped into subsets by decile to facilitate viewing patterns.

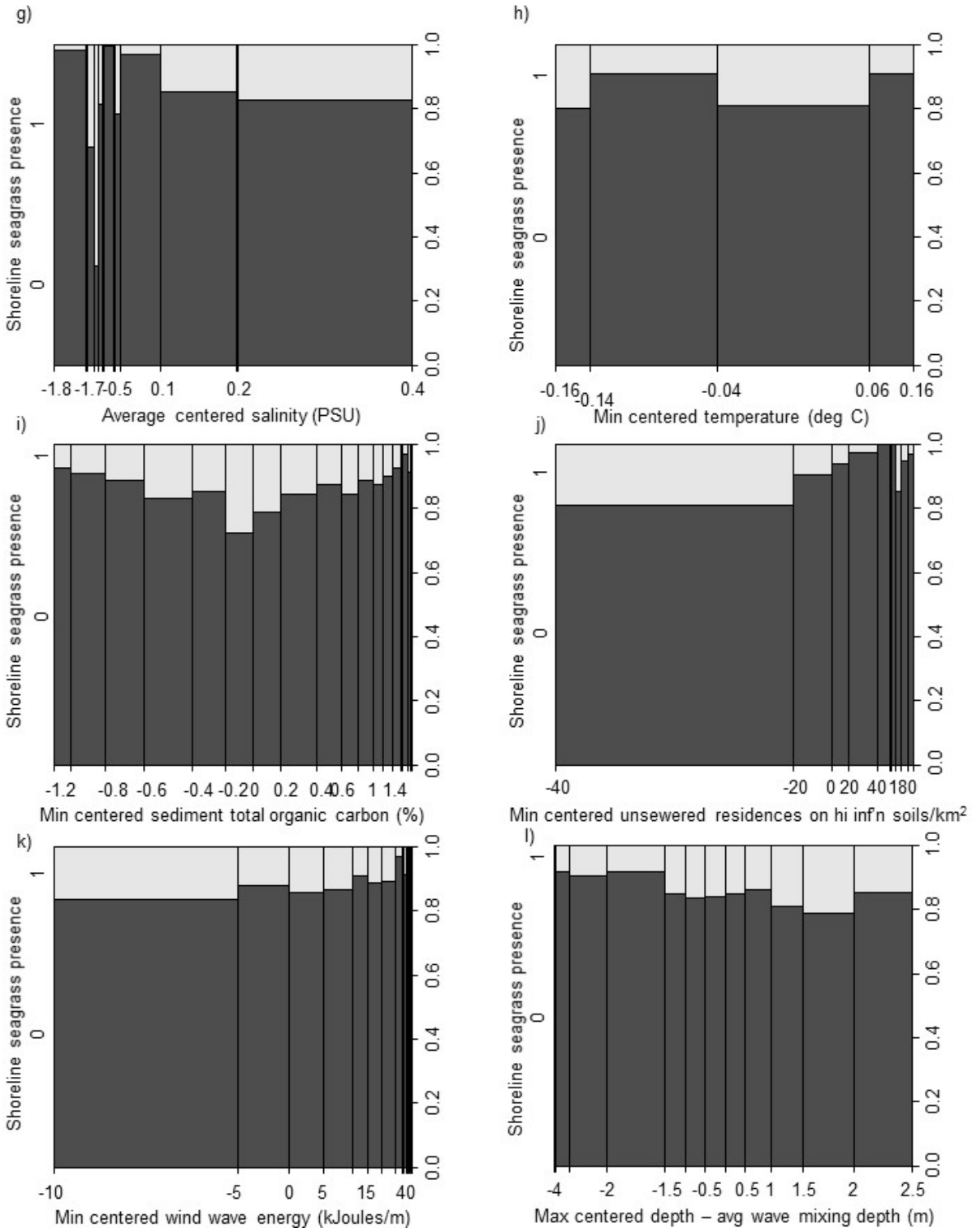

Figure D-6 (Cont'd).

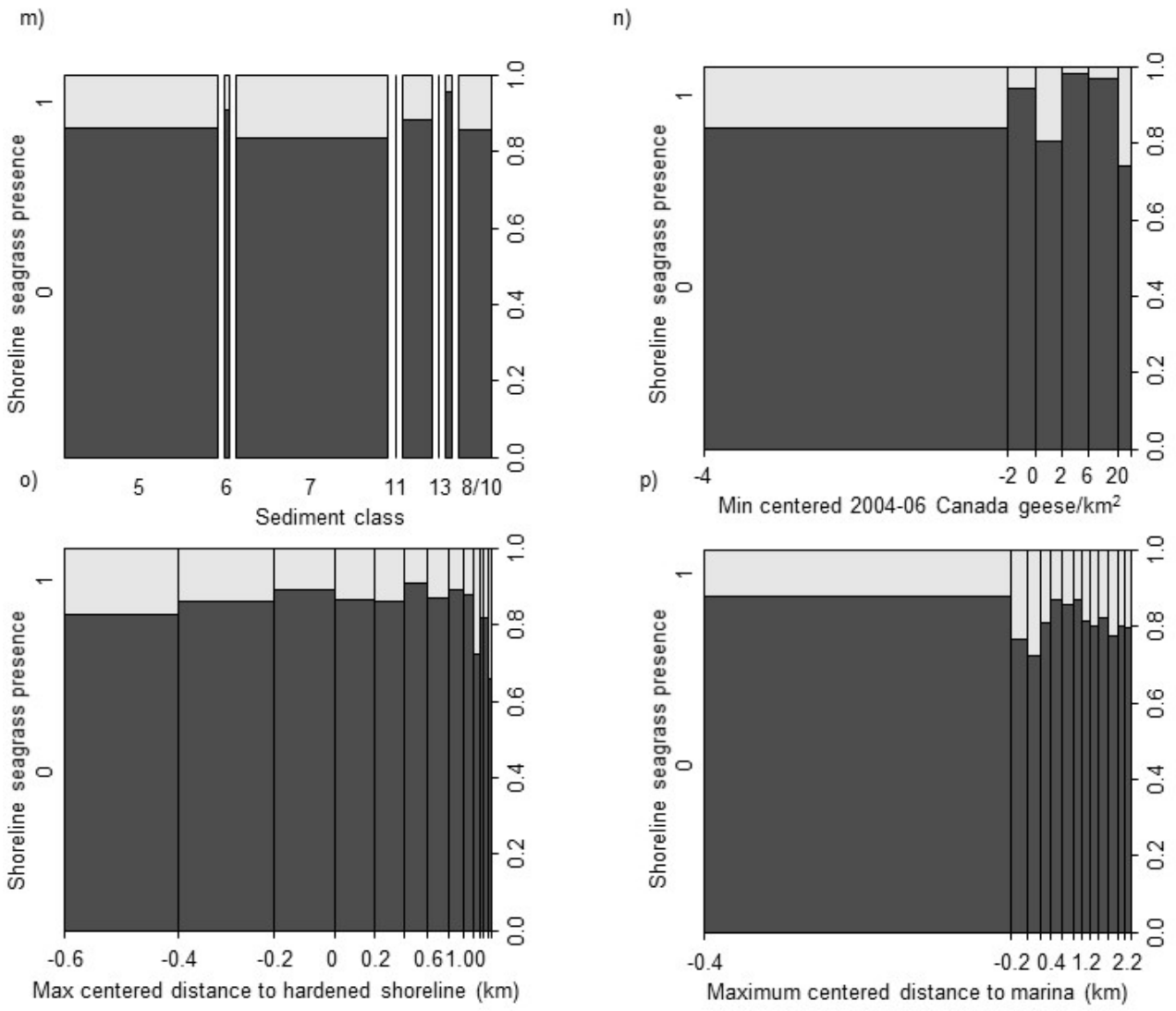

Figure D-6 (Cont'd)

D.2.2. Test of Model Assumptions During Initial Model Development for Shoreline P/A

Prior to fitting a model for shoreline occupancy, we screened predictors for potential cross-correlations. Minimum shoreline average temperature was correlated with maximum shoreline average Secchi depth and shoreline average salinity ($r > 0.70$). Shoreline average salinity was correlated with maximum shoreline average Secchi but not with maximum shoreline Secchi depth minima. Thus we dropped temperature and maximum shoreline average Secchi depth, but retained maximum shoreline Secchi depth minimum as predictors.

The original model fit after dropping insignificant terms and prior to accounting for spatial autocorrelation included the following main effects: salinity, availability of depths greater than average mixing depth, shoreline isolation, distance to hardened shorelines, distance to nearest marina, density of Canada geese, and two interaction terms: wind-generated wave energy x sediment particle- size class and Secchi depth x sediment percent total organic carbon. The density of unsewered residences on high infiltration soils was not a significant predictor. Model diagnostic tests showed evidence of heterogeneity of variance, and spline correlograms with 95% confidence interval by shoreline showed evidence of strong spatial autocorrelation (Figures D-7a, b). The minimum x-intercept, corresponding to the minimum range at which spatial autocorrelation of residuals was undetectable, was 1365 meters. Subsequent peaks in the correlogram have similar breadth, and probably represent autocorrelation between different seagrass patches, as compared to correlation within a given patch. In many cases, residuals appeared to be greater for classes or ranges associated with shoreline occupancy (Figures D-7c, d).

D.2.3. Test of Model Assumptions During Initial Model Development for Shoreline Relative Abundance

The final model fit to predict <u>average</u> shoreline occupancy was more complex, incorporating both higher-order terms to account for nonlinear responses and two- and three-way interaction terms:

$$\begin{aligned} savcodeAv = {} & csalAv + csalAv^2 + csalAv^3 + csecchiminAv + csecchiminAv^3 + \\ & cpttocAv^2 + cpttocAv^3 + cZgtMXZavAv + cZgtMXZavAv^2 + cDistToHdShAv + \\ & cDstoMarinaAv + cUSRMARIkm2Av + fISOLATED + csalAv \times cpttocAv + \\ & csecchiminAv \times cpttocAv + csalAv \times csecchiminAv \times cZgtMXZavAv + cwindAv \times \\ & fSED5) \end{aligned} \quad \text{(D1)}$$

with terms defined as above (Table D2) and added variables defined as:

fSED5 = sediment classes

SED5_124561213 = Gravel, Sandy gravel, Gravel-sand-silt, Sand, Gravelly sand, Sand-silt-clay, Gravel-silt-clay

SED5_7 = Silty sand

SED5_81011 = Silty, Sandy Silt, and Clay-Silt

Based on diagnostic plots of Pearson residuals versus individual predictor variables, nonlinearities appear to have been accounted for in the final model. However, the model still demonstrated heterogeneity of variance and spatial autocorrelation of residuals (Figure D-7a-d).

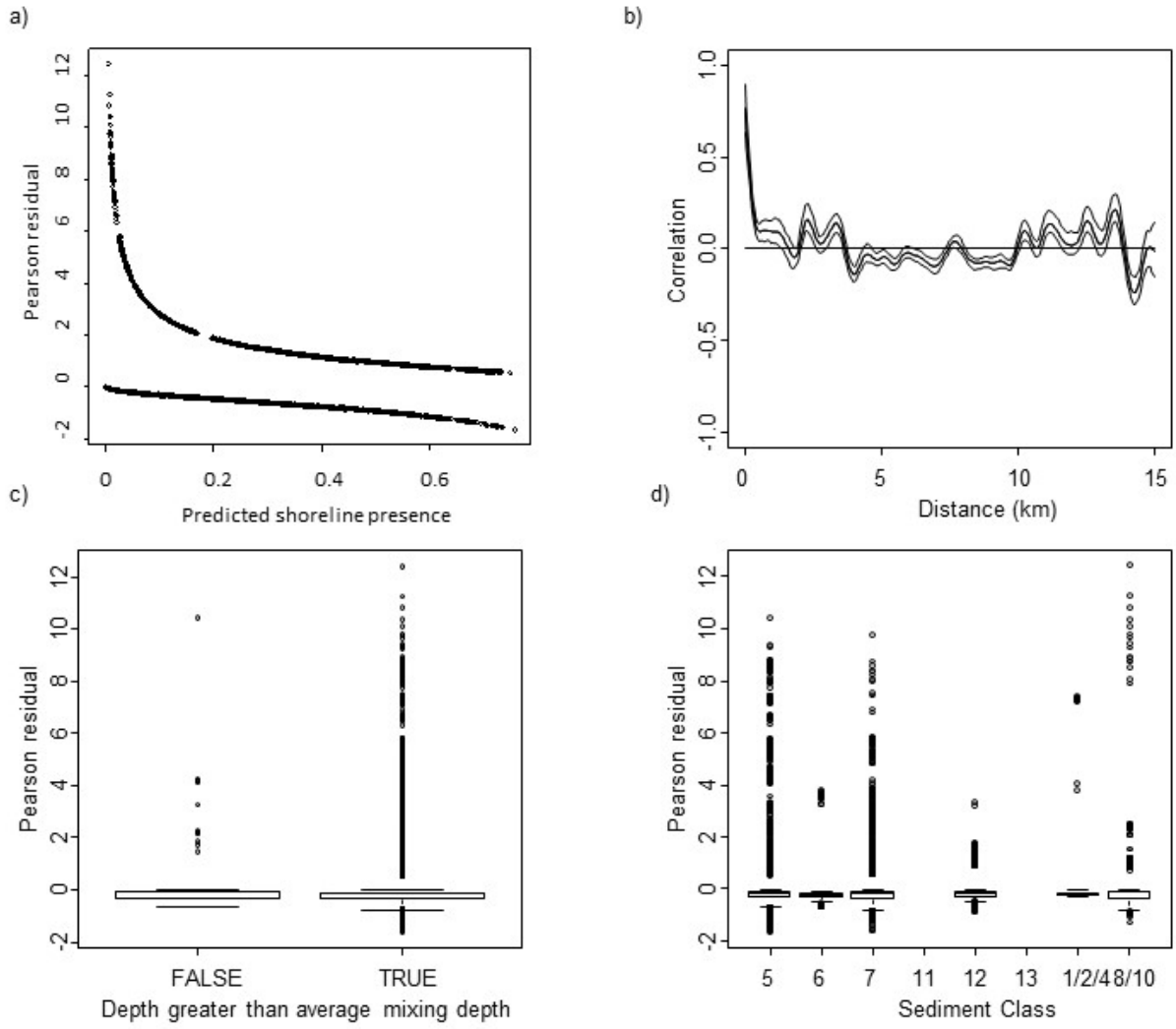

Figure D-7. Diagnostic plots for initial model 3 predicting shoreline seagrass presence/absence. a) Pearson residuals versus predicted presence/absence, b) Spline correlogram of Pearson residuals for model 3, Shoreline 14. The minimum x-intercept is 1365 meters. Pearson residuals for initial model 3 versus c) mixing depth indicator and d) sediment class. Sediment class 5 = sand, 6 = gravelly-sand, 7 = silty-sand, 11 = clay-silt, 12 = sand-silt-clay, 13 = gravel-silt-clay, 1/2/4 = gravel + sandy gravel + gravel-sand-silt, 8/10 = silty + sandy silt.

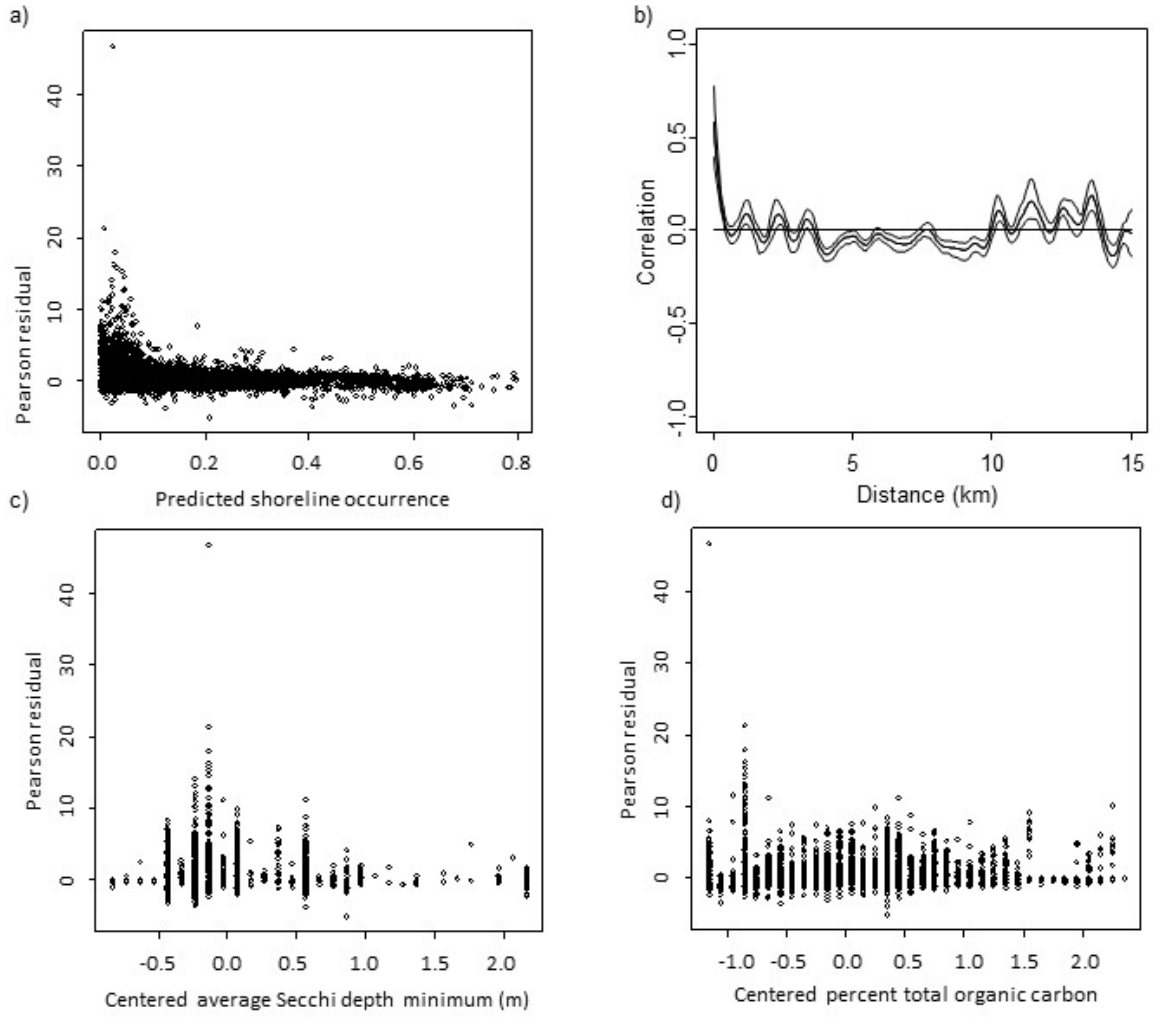

Figure D-8. Diagnostics for model 3j. a) Pearson residuals versus predicted value, b) spline correlogram of Pearson residuals for Shoreline 14. The minimum x-intercept (within 95% confidence interval) was 420 meters. Higher order terms appear to have accounted for most of the nonlinearities in relationships, as illustrated by plots of Pearson residuals versus c) centered average Secchi depth minimum (m) and d) centered shoreline average percent sediment total organic carbon.

D.2.4. Test of Model Assumptions During Initial Model Development for Maximum Seagrass Depth

Models predicting minimum and maximum depth of occurrence by shoreline index were developed for the subset of records corresponding to shoreline occupancy. Models predicting minimum depth of occurrence were created with a subset of data after excluding records with average or maximum mixing depths of zero. The restriction of model predictions to shoreline occurrence transects allowed us to drop the random shoreline effect from models and compare the fit of general linear models with general additive models because fitting GAMs is much less memory intensive than fitting GAMMs.

The initial GLM model evaluated to predict maximum depth of seagrass occurrence included main effects and both two- and three-way interaction terms for shoreline maximum Secchi depth average, shoreline maximum percent sediment total organic carbon, and density of unsewered residences on high infiltration soils:

bathymmax = csecchiavMax + cpttocMax + cUSRMARIkm2Max + csecchiavMax x cpttocMax + csecchiavMax x cUSRMARIkm2Max + cpttocMax x cUSRMARIkm2Max + csecchiavMax x cpttocMax x cUSRMARIkm2Max (D2)

After we dropped insignificant interaction terms, the resultant model was

bathymmax = csecchiavMax + cpttocMax + cUSRMARIkm2Max + csecchiavMax x cpttocMax + csecchiavMax x cUSRMARIkm2Max (D3)

Residual plots showed an increasing variance with the mean, possibly peaking at intermediate values, so we fit a generalized additive models and general linear models with second- and third-order terms to try to capture nonlinearities in response. The best GLM model with higher order terms and interactions was:

$$bathymmax = csecchiavMax^2 + csecchiavMax^3 + cpttocMax + cpttocMax^2 + cUSRMARIkm2Max + cUSRMARIkm2Max^2 + cUSRMARIkm2Max^3 + csecchiavMax \times cpttocMax \quad \text{(D4)}$$

However, this model still showed evidence of heterogeneity of variance, so we refit GLM models after log10-transformation of predictors. No terms were dropped from the log-transformed model with 3-way interactions and VIF terms were less than 10:

bathymmax = cL10secchiavMax*cL10pttocMax*cL10USRMARIkm2Max) (D5)

However, based on comparison of AIC values, a GAM model provided a superior fit compared to the GLM model with log-transformed predictors:

bathymmax = s(cL10secchiavMax) + s(cL10pttocMax) + s(cL10USRMARIkm2Max)) (D6)

where s = smoothing function

Appendix E. Quality of the Data and Limitations on Use of the Data

We used our Narragansett Bay pilot application of the statistical modelling approach to illustrate how a predictive model could be developed to assess factors affecting seagrass growth. Our goal was to distinguish between nutrient and nonnutrient factors affecting seagrass growth and survival and to elucidate different mechanisms of action for effects of nutrients on seagrass. We used the best publically available data sets to describe environmental variables that affect seagrass growth and survival in Narragansett Bay to support our predictive statistical model based on: temporal matches to 2006 seagrass maps, spatial extent, completeness, and spatial resolution. We filled in gaps in Secchi depth at the southern end of the western arm of Narragansett Bay using offshore remotely sensed estimates of light attenuation coefficients (from > 30 meters depth), assuming these were similar to nearshore values. When only point data were available, e.g., for salinity, Secchi depth, and wave energy, we created a continuous grid by interpolation through use of Theissen polygons, the Euclidean function in ArcMap (filling in gaps along shoreline), or inverse distance weighting. Some variables were not available so we substituted indicators. This included use of density of coastal residences on high infiltration soils as an indicator of potential groundwater N inputs, use of distance to nearest marina as an indicator of potential physical disturbance from mooring beds, and use of salinity gradients as an indicator of total N gradients in final scenarios. Data on some variables potentially affecting seagrass distribution were simply not available, e.g., sulfide concentrations in sediment porewater, actual measurements of tidal current velocity as compared to estimated values, and measurements of turbidity or light attenuation near the sediment interface (as opposed to upper water column). Any of the limitations to data availability or completeness of model inputs could have influenced the accuracy of our model predictions, but are unlikely to have produced biased model results. We based model projections of future condition following nutrient load reductions on the assumption that nutrient concentrations will decline in proportion to load reductions and that space-for-time substitutions are appropriate for model development. The model can be improved in the future as more complete data or modelled estimates become available, e.g., for tidal currents and for dissolved inorganic N and total N concentrations across the bay.

EPA
United States
Environmental Protection
Agency

PRESORTED STANDARD
POSTAGE & FEES PAID
EPA
PERMIT NO. G-35

Office of Research and Development (8101R)
Washington, DC 20460