Species Distribution Modelling and Kernel Density Analysis of Benthic Ecologically and Biologically Significant Areas (EBSAs) and Other Benthic Fauna in the Maritimes Region

L. Beazley, E. Kenchington, and C. Lirette

Ocean and Ecosystem Sciences Division
Maritimes Region
Fisheries and Oceans Canada

Bedford Institute of Oceanography
PO Box 1006
Dartmouth, Nova Scotia
Canada B2Y 4A2

2017

Canadian Technical Report of Fisheries and Aquatic Sciences 3204

Fisheries and Oceans Canada Pêches et Océans Canada

Canada

Canadian Technical Report of Fisheries and Aquatic Sciences

Technical reports contain scientific and technical information that contributes to existing knowledge but which is not normally appropriate for primary literature. Technical reports are directed primarily toward a worldwide audience and have an international distribution. No restriction is placed on subject matter and the series reflects the broad interests and policies of Fisheries and Oceans Canada, namely, fisheries and aquatic sciences.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in the data base *Aquatic Sciences and Fisheries Abstracts*.

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page.

Numbers 1-456 in this series were issued as Technical Reports of the Fisheries Research Board of Canada. Numbers 457-714 were issued as Department of the Environment, Fisheries and Marine Service, Research and Development Directorate Technical Reports. Numbers 715-924 were issued as Department of Fisheries and Environment, Fisheries and Marine Service Technical Reports. The current series name was changed with report number 925.

Rapport technique canadien des sciences halieutiques et aquatiques

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui ne sont pas normalement appropriés pour la publication dans un journal scientifique. Les rapports techniques sont destinés essentiellement à un public international et ils sont distribués à cet échelon. II n'y a aucune restriction quant au sujet; de fait, la série reflète la vaste gamme des intérêts et des politiques de Pêches et Océans Canada, c'est-à-dire les sciences halieutiques et aquatiques.

Les rapports techniques peuvent être cités comme des publications à part entière. Le titre exact figure au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la base de données *Résumés des sciences aquatiques et halieutiques.*

Les rapports techniques sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page du titre.

Les numéros 1 à 456 de cette série ont été publiés à titre de Rapports techniques de l'Office des recherches sur les pêcheries du Canada. Les numéros 457 à 714 sont parus à titre de Rapports techniques de la Direction générale de la recherche et du développement, Service des pêches et de la mer, ministère de l'Environnement. Les numéros 715 à 924 ont été publiés à titre de Rapports techniques du Service des pêches et de la mer, ministère des Pêches et de l'Environnement. Le nom actuel de la série a été établi lors de la parution du numéro 925.

Canadian Technical Report of
Fisheries and Aquatic Sciences 3204

2017

Species Distribution Modelling and Kernel Density Analysis of Benthic Ecologically and Biologically Significant Areas (EBSAs) and Other Benthic Fauna in the Maritimes Region

by

L. Beazley, E. Kenchington, C. Lirette

Ocean and Ecosystem Sciences Division
Maritimes Region
Fisheries and Oceans Canada
Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, N.S.
Canada B2Y 4A2

Cat. No. Fs97-6/3204E-PDF ISBN 978-0-660-07759-8 ISSN 1488-5379

Correct citation for this publication:

Beazley, L., Kenchington, E., and Lirette, C. 2017. Species Distribution Modelling and Kernel Density Analysis of Benthic Ecologically and Biologically Significant Areas (EBSAs) and Other Benthic Fauna in the Maritimes Region. Can. Tech. Rep. Fish. Aquat. Sci. 3204: vi + 159p.

TABLE OF CONTENTS

ABSTRACT

Beazley, L., Kenchington, E., and Lirette, C. 2017. Species Distribution Modelling and Kernel Density Analysis of Benthic Ecologically and Biologically Significant Areas (EBSAs) and Other Benthic Fauna in the Maritimes Region. Can. Tech. Rep. Fish. Aquat. Sci. 3204: vi + 159p.

We present random forest species distribution models and kernel density estimation (KDE)-derived significant areas for benthic Ecologically and Biologically Significant Areas (EBSAs) and other benthic taxa in Fisheries and Oceans Canada's Maritimes Region: Horse Mussel (*Modiolus modiolus*) Beds, Stalked Tunicate Fields, Sand Dollar Beds, Soft Coral Gardens, and *Flabellum* cup corals. Using a suite of sixty-six environmental variables derived from various sources and native spatial resolutions, random forest was employed to predict the probability of occurrence and biomass distribution of these taxa using data collected from DFO multispecies trawl surveys, DFO scallop stock assessment surveys, and targeted *in situ* benthic camera and/or video surveys. Random forest presence-absence models had excellent predictive capacity, with cross-validated Area Under the Receiver Operating Characteristic Curve (AUC) values ranging from 0.868 to 0.965. Areas of suitable habitat generated from the presence-absence models were compared to the KDE-derived significant area polygons with relatively good congruence. Such a process could be formally conducted in the future to refine the outer boundaries of these significant areas. Significant concentrations (as defined by KDE) of these taxa are considered conservation priorities in the Scotian Shelf Bioregional Marine Protected Area (MPA) Network Design Strategy, and are currently being used in exploratory conservation planning analyses.

RÉSUMÉ

Beazley, L., Kenchington, E., and Lirette, C. 2017. Modélisation de la répartition des espèces et analyse du noyau de densité des zones benthiques d'importance écologique et biologique et de la faune benthique dans la région des Maritimes. Can. Tech. Rep. Fish. Aquat. Sci. 3204: vi + 159p.

Nous présentons des zones importantes dérivées de modèles de répartition des espèces au moyen de forêts aléatoires et de l'estimation de la densité par la méthode du noyau pour les zones benthiques d'importance écologique et biologique et d'autres taxons benthiques dans la région des Maritimes de Pêches et Océans Canada (MPO) : gisements de modioles (*Modiolus modiolus*), zones de tuniciers lobés, gisements de clypéastres, jardins de coraux mous, et madrépolaires *Flabellum*. À l'aide d'un ensemble de soixante-six variables environnementales provenant de diverses sources et de résolutions spatiales natives, le modèle de forêts aléatoires a été employé pour prédire la probabilité d'occurrence et la répartition de la biomasse de ces taxons grâce aux données recueillies à partir des relevés plurispécifiques au chalut du MPO, des évaluations des stocks de pétoncles par le MPO et des relevés benthiques ciblés sur place (par caméra ou vidéo). Les modèles de forêts aléatoires sur la présence et l'absence avaient une excellente efficacité de prévision selon des valeurs contre-validées de l'aire sous la courbe de la fonction d'efficacité du récepteur variant de 0,868 à 0,965. Les zones constituant un habitat convenable générées à partir des modèles sur la présence et l'absence ont été comparées aux polygones de zones importantes dérivés de l'estimation de la densité par la méthode du noyau et dont la concordance est relativement bonne. Un tel processus pourrait être officiellement réalisé ultérieurement pour préciser les limites extérieures de ces zones importantes. Des concentrations importantes (comme définies par l'estimation de la densité par la méthode du noyau) de ces taxons sont considérées comme des priorités en matière de conservation dans la stratégie de conception du réseau d'aires marines protégées de la biorégion du plateau néo-écossais, et elles sont déjà utilisées dans le cadre d'analyses exploratoires de la planification de la conservation.

INTRODUCTION

When Canada ratified the United Nations Convention on Biological Diversity (CBD) in 1992, it adopted various biodiversity conservation and protection commitments, including the commitment to establish a network of marine protected areas (MPAs). In October 2010, Canada and other parties to the Convention on Biological Diversity agreed to the CBD's Strategic Plan for Biodiversity 2011-2020. The Strategic Plan consists of 20 new biodiversity targets for 2020, termed the 'Aichi Biodiversity Targets' (SCBD, 2010). Canada's commitment under the Strategic Plan includes protecting 5% of coastal and marine areas by 2017, and 10% by 2020. To meet this commitment, Fisheries and Oceans Canada (DFO), with consultation from other government departments, organizations, and stakeholders, is leading the development of a national MPA network at the regional level. Guidance for the design of the network is provided in the 2011 National Framework for Canada's Network of MPAs (Government of Canada, 2011), which recommends that the scientific guidance for selecting areas to establish a representative network of MPAs outlined in Annex II of CBD Decision IX/20 be followed. This guidance states that effective MPA networks should include 1) Ecologically and Biologically Significant Areas (EBSAs), 2) ecological representation (or representivity), 3) connectivity, 4) replicated ecological features, 5) adequate and viable sites.

Since 2004, DFO has made significant efforts to identify Ecologically and Biologically Significant Areas in Canadian waters. Delineation of an EBSA does not impart immediate conservation status. However, it does draw attention to an area that has particularly high ecological or biological significance, information that is useful during broader oceans planning and management processes including MPA network design (DFO, 2004; 2014b). In 2014, eighteen EBSAs were described and delineated in the offshore component of the Scotian Shelf bioregion (the Scotian Shelf bioregion is approximately equal to DFO's Maritimes Region boundary). To help evaluate and identify these EBSAs a total of 149 ecological or biological data layers were compiled or created, which were organized under the following major themes: areas of high biological productivity or biomass, areas of high fish and invertebrate diversity, and important habitats for fishes and invertebrates, coral and sponge occurrences, Critical Habitat for species at risk, important areas for seabird functional guilds, and distinct physical conditions (DFO, 2014b). Concurrently, Kenchington (2014) compiled information on marine benthic species and habitats occurring on the Scotian Shelf that are recognized in other jurisdictions as meeting EBSA or similar criteria. Fourteen structure-forming, biogenic habitats were identified and could be used to identify EBSAs in the region. Delineating EBSAs based on the spatial distribution of individual species or species groups represents a challenge, particularly if the full spatial distribution of the species is not known. In these instances, species distribution or habitat suitability modelling can be a useful tool to help delineate the potential habitat of a species from which boundaries can be drawn. Such models have already been developed for deep-water corals and sponges for the entire spatial extent of the Maritimes Region (see Beazley et al., 2016a).

In March 2016, a Canadian Science Advisory Secretariat (CSAS) meeting was held in Halifax, Nova Scotia to identify Significant Benthic Areas of deep-water corals and sponges across eastern Canada. In this process, Significant Benthic Areas (SBAs) were defined as a regional habitat that contains corals and/or sponges as a dominant and defining feature (Kenchington et al., 2016a). According to DFO's Ecological Risk Assessment Framework for Coldwater Corals and Sponge Dominated Communities, significance of an area is determined "through guidance provided by DFO-lead processes based on current knowledge of such species, communities, and ecosystems" (http://www.dfo-mpo.gc.ca/fm-gp/peches-fisheries/fish-ren-peche/sff-cpd/risk-ecolo-risque-eng.htm). The primary tool used for the identification of Significant Benthic Areas of deep-water corals and sponges in Canada is kernel density estimation (KDE). This tool was first applied in the Northwest Atlantic Fisheries Organization (NAFO) Regulatory Area to identify significant concentrations of corals and sponges from their broader distribution using research vessel (RV) trawl survey catch data (Kenchington et al., 2009). KDE was first applied domestically in 2010 to delineate coral and sponge concentrations from RV trawl data in all five biogeographic zones of eastern Canada (Kenchington et al., 2010). This analysis was re-run in 2016 to include the most recent catch data (Kenchington et al., 2016b). Kenchington et al. (2016b) recommended that the boundaries of these significant area polygons be refined using the distribution of null catches and species distribution models, which incorporate environmental data. Such a process was conducted at the March 2016 CSAS meeting, where Significant Benthic Areas for the coral and sponge taxa were delineated by comparing the spatial congruence between the KDE-derived significant area polygons, the species distribution models, and data observations (presences and nulls) (Kenchington et al., 2016a). In the Scotian Shelf bioregion there was a high degree of spatial congruence between the KDE-derived polygons, the areas predicted as suitable habitat by the model, and the presence data observations and therefore no refinement to the KDE-derived significant area polygons was required. Additional Significant Benthic Areas were created from the SDM outputs for those taxa whose distribution was not fully sampled by the RV trawl survey, particularly along the slope.

Here we present species distribution models and KDE-derived significant area polygons for several EBSA taxa in DFO's Maritimes Region using records primarily from the DFO multispecies trawl survey. For many of the EBSA taxa identified in Kenchington (2014) there were an insufficient number of records from the DFO trawl surveys to produce adequate distribution models, due in part to the ineffectiveness of trawl gear at catching certain fauna (e.g. xenophyophores and tube-dwelling anemones). As a result, species distribution models and KDE analyses were conducted only for the following four taxa: Horse Mussel (*Modiolus modiolus*) Beds, Stalked Tunicate Fields, Sand Dollar Beds, and Soft Coral Gardens. Of these, horse mussel reefs, stalked tunicate *Boltenia ovifera*, and soft coral *Gersemia rubiformis* have been recognized as possible Sensitive Benthic Areas in the Bay of Fundy (DFO, 2015). *Flabellum* cup corals, which are not considered an EBSA or vulnerable marine ecosystem (VME) indicator in international waters, was also included in this report as they reach locally high densities on the Scotian Shelf and Slope (Cogswell et al., 2009) and can be relevant to conservation in different contexts (Kenchington et al., 2016b). Significant concentrations (as delineated by KDE) of these taxa are considered conservation priorities in the Scotian Shelf Bioregional Marine Protected Area Network Design Strategy (King et al., in prep.). Although the KDE-derived polygons and species distribution model outputs are not formally compared to delineate Significant Benthic Areas following Kenchington et al. (2016a), we show the spatial congruence between both outputs so that such a task could easily be conducted in the future.

METHODOLOGY

Study Area

The Maritimes Region, one of DFO's six administrative regions across Canada, was used as the boundary for species distribution modelling in this report (Figure 1). This study area encompasses the entire Scotian Shelf and Bay of Fundy and is delimited by the Canadian Maritime Boundary to the west in Gulf of Maine, the 200 nautical mile Exclusive Economic Zone (EEZ) in the south, the Placentia Bay-Grand Bank Large Ocean Management Area in the east, and the Gulf Region MPA Network Planning Boundary in the north. A 5-km buffer was placed around all land to avoid its inclusion in the models. The total area covered in the study extent is approximately 459,139 km^2 based on a NAD 1983 UTM Zone 20N projection.

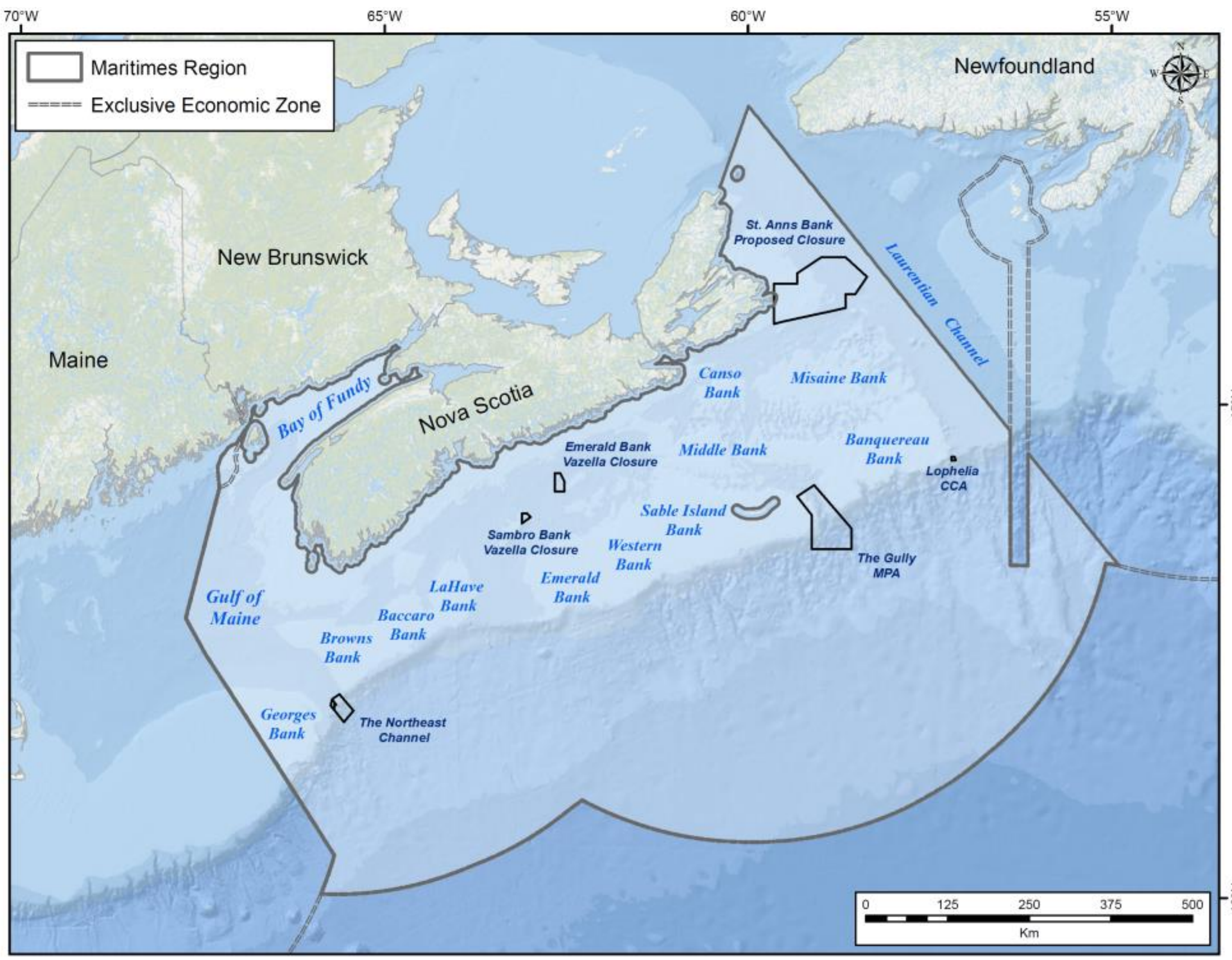

Figure 1. Extent of the DFO Maritimes Region boundary used for species distribution modelling. Place names and location of the Gully MPA, St. Anns Bank Proposed Closure, and other areas closed to protect corals and sponges are indicated.

Environmental Data

Sixty-six environmental variables derived from various sources and native spatial resolutions were used as predictor variables in the random forest models (Table 1). Variables were chosen based on their availability and assumed relevance to the distribution of benthic fauna. Bathymetry was derived from the Canadian Hydrographic Service (CHS) Atlantic Bathymetry Compilation (ABC). This data is the highest resolution bathymetry available for the entire study area, with a horizontal resolution of up to 64 m in some areas. In the Maritimes Region the data are resolved to 15 arc-seconds, which is equivalent to approximately 500 m. Slope in degrees was derived from the depth raster using the 'Slope' tool in ArcMap's Spatial Analyst toolbox, ArcMap version 10.2.2 (ESRI, 2011). All other environmental variables were derived from long-term modelled oceanographic or remote-sensing data and were spatially interpolated across the study area using ordinary kriging in ArcMap. Specific details on data sources and methodology used for the spatial interpolation of these variables are documented in a separate technical report (Beazley et al., in prep, although see Beazley et al., 2016b for information on the same environmental data sources and variables for the Estuary and Gulf of St. Lawrence). Only variables that were spatially interpolated with reasonable confidence were used in this report, and as a result a number of available data layers (e.g. dissolved oxygen) were not considered. All predictor layers were displayed in raster format with geographic coordinates using the WGS 1984 datum and a ~0.012º cell size (approximately equal to 1 km in the Maritimes Region).

Response Data

Species composition of the five taxonomic groups modelled in this report is presented in Table 2. Records from the DFO research vessel multispecies trawl surveys were considered the main source of data given the wide spatial coverage of the trawl surveys and their stratified (by depth) random design (DFO, 2014a). Trawl surveys in the Maritimes Region are conducted on the CCGS *Alfred Needler*, *Wilfred Templeman*, or *Teleost*. Tows are conducted using primarily Western IIA trawl gear, although other gear types (i.e. Campelen and US 4 seam bridle 3 trawls) are also used in the region. Only those tows conducted using Western IIA gear were extracted to avoid mixing records with different catchabilities. The invertebrate catch data from the surveys are accessed through the Maritimes Region Virtual Data Centre (VDC) (http://marvdc.bio.dfo.ca/pls/vdc/mwmfdweb.auth). Data from 1999 to March 2015 were extracted from the VDC for all taxonomic groups, coinciding with the year that selected invertebrates were recorded more systematically in the surveys (Tremblay et al., 2007). For each taxonomic group, absence records were created from null (zero) catches that occurred in the same surveys.

Table 1. Summary of the 66 environmental variables used as predictor variables in random forest modelling. N/A = Not applicable.

Variable	Data source	Temporal range	Unit	Native resolution
Depth	CHS-ABC	N/A	metres	15 arc-sec.
Slope	CHS-ABC	N/A	degrees	15 arc-sec.
Bottom Salinity Mean	GLORYS2V1	1993 - 2011	N/A	¼ º
Bottom Salinity Average Minimum	GLORYS2V1	1993 - 2011	N/A	¼ º
Bottom Salinity Average Maximum	GLORYS2V1	1993 - 2011	N/A	¼ º
Bottom Salinity Average Range	GLORYS2V1	1993 - 2011	N/A	¼ º
Bottom Temperature Mean	GLORYS2V1	1993 - 2011	ºC	¼ º
Bottom Temperature Average Minimum	GLORYS2V1	1993 - 2011	ºC	¼ º
Bottom Temperature Average Maximum	GLORYS2V1	1993 - 2011	ºC	¼ º
Bottom Temperature Average Range	GLORYS2V1	1993 - 2011	ºC	¼ º
Bottom Current Mean	GLORYS2V1	1993 - 2011	$m\ s^{-1}$	¼ º
Bottom Current Average Minimum	GLORYS2V1	1993 - 2011	$m\ s^{-1}$	¼ º
Bottom Current Average Maximum	GLORYS2V1	1993 - 2011	$m\ s^{-1}$	¼ º
Bottom Current Average Range	GLORYS2V1	1993 - 2011	$m\ s^{-1}$	¼ º
Bottom Shear Mean	GLORYS2V1	1993 - 2011	Pa	¼ º
Bottom Shear Average Minimum	GLORYS2V1	1993 - 2011	Pa	¼ º
Bottom Shear Average Maximum	GLORYS2V1	1993 - 2011	Pa	¼ º
Bottom Shear Average Range	GLORYS2V1	1993 - 2011	Pa	¼ º
Surface Salinity Mean	GLORYS2V1	1993 - 2011	N/A	¼ º
Surface Salinity Average Minimum	GLORYS2V1	1993 - 2011	N/A	¼ º

Surface Salinity Average Maximum	GLORYS2V1	1993 - 2011	N/A	¼ º
Surface Salinity Average Range	GLORYS2V1	1993 - 2011	N/A	¼ º
Surface Temperature Mean	GLORYS2V1	1993 - 2011	ºC	¼ º
Surface Temperature Average Minimum	GLORYS2V1	1993 - 2011	ºC	¼ º
Surface Temperature Average Maximum	GLORYS2V1	1993 - 2011	ºC	¼ º
Surface Temperature Average Range	GLORYS2V1	1993 - 2011	ºC	¼ º
Surface Current Speed Mean	GLORYS2V1	1993 - 2011	m s^{-1}	¼ º
Surface Current Speed Average Minimum	GLORYS2V1	1993 - 2011	m s^{-1}	¼ º
Surface Current Speed Average Maximum	GLORYS2V1	1993 - 2011	m s^{-1}	¼ º
Surface Current Speed Average Range	GLORYS2V1	1993 - 2011	m s^{-1}	¼ º
Maximum Average Fall Mixed Layer Depth	GLORYS2V1	1993 - 2011	metres	¼ º
Maximum Average Winter Mixed Layer Depth	GLORYS2V1	1993 - 2011	metres	¼ º
Maximum Average Spring Mixed Layer Depth	GLORYS2V1	1993 - 2011	metres	¼ º
Maximum Average Summer Mixed Layer Depth	GLORYS2V1	1993 - 2011	metres	¼ º
Fall Chlorophyll *a* Mean	MODIS Level I	2002 - 2012	mg m^{-3}	2 km
Fall Chlorophyll *a* Minimum	MODIS Level I	2002 – 2012	mg m^{-3}	2 km
Fall Chlorophyll *a* Maximum	MODIS Level I	2002 – 2012	mg m^{-3}	2 km
Fall Chlorophyll *a* Range	MODIS Level I	2002 – 2012	mg m^{-3}	2 km
Spring Chlorophyll *a* Mean	MODIS Level I	2002 – 2012	mg m^{-3}	2 km
Spring Chlorophyll *a* Minimum	MODIS Level I	2002 – 2012	mg m^{-3}	2 km
Spring Chlorophyll *a* Maximum	MODIS Level I	2002 – 2012	mg m^{-3}	2 km
Spring Chlorophyll *a* Range	MODIS Level I	2002 – 2012	mg m^{-3}	2 km
Summer Chlorophyll *a* Mean	MODIS Level I	2002 – 2012	mg m^{-3}	2 km

Summer Chlorophyll *a* Minimum	MODIS Level I	2002 – 2012	mg m^{-3}	2 km
Summer Chlorophyll *a* Maximum	MODIS Level I	2002 – 2012	mg m^{-3}	2 km
Summer Chlorophyll *a* Range	MODIS Level I	2002 – 2012	mg m^{-3}	2 km
Annual Chlorophyll *a* Mean	MODIS Level I	2002 – 2012	mg m^{-3}	2 km
Annual Chlorophyll *a* Minimum	MODIS Level I	2002 – 2012	mg m^{-3}	2 km
Annual Chlorophyll *a* Maximum	MODIS Level I	2002 – 2012	mg m^{-3}	2 km
Annual Chlorophyll *a* Range	MODIS Level I	2002 – 2012	mg m^{-3}	2 km
Fall Primary Production Mean	SeaWiFS Level-3 with other input parameters	2006 – 2010	mg C m^{-2} day^{-1}	9 km
Fall Primary Production Average Minimum	SeaWiFS Level-3 with other input parameters	2006 – 2010	mg C m^{-2} day^{-1}	9 km
Fall Primary Production Average Maximum	SeaWiFS Level-3 with other input parameters	2006 – 2010	mg C m^{-2} day^{-1}	9 km
Fall Primary Production Average Range	SeaWiFS Level-3 with other input parameters	2006 – 2010	mg C m^{-2} day^{-1}	9 km
Spring Primary Production Mean	SeaWiFS Level-3 with other input parameters	2006 – 2010	mg C m^{-2} day^{-1}	9 km
Spring Primary Production Average Minimum	SeaWiFS Level-3 with other input parameters	2006 – 2010	mg C m^{-2} day^{-1}	9 km
Spring Primary Production Average Maximum	SeaWiFS Level-3 with other input parameters	2006 – 2010	mg C m^{-2} day^{-1}	9 km
Spring Primary Production Average Range	SeaWiFS Level-3 with other input parameters	2006 – 2010	mg C m^{-2} day^{-1}	9 km
Summer Primary Production Mean	SeaWiFS Level-3 with other input parameters	2006 – 2010	mg C m^{-2} day^{-1}	9 km

Summer Primary Production Average Minimum	SeaWiFS Level-3 with other input parameters	2006 – 2010	mg C m^{-2} day^{-1}	9 km
Summer Primary Production Average Maximum	SeaWiFS Level-3 with other input parameters	2006 – 2010	mg C m^{-2} day^{-1}	9 km
Summer Primary Production Average Range	SeaWiFS Level-3 with other input parameters	2006 – 2010	mg C m^{-2} day^{-1}	9 km
Annual Primary Production Mean	SeaWiFS Level-3 with other input parameters	2006 – 2010	mg C m^{-2} day^{-1}	9 km
Annual Primary Production Average Minimum	SeaWiFS Level-3 with other input parameters	2006 – 2010	mg C m^{-2} day^{-1}	9 km
Annual Primary Production Average Maximum	SeaWiFS Level-3 with other input parameters	2006 – 2010	mg C m^{-2} day^{-1}	9 km
Annual Primary Production Average Range	SeaWiFS Level-3 with other input parameters	2006 – 2010	mg C m^{-2} day^{-1}	9 km

Table 2. Species composition of each EBSA/taxonomic group from catch data collected from the DFO multispecies trawl surveys, the main source of data used in random forest and KDE analyses. Also shown are the number of presence records included in the random forest models and the VDC code used for data entry into the VDC. Composition shown is for the purposes of re-extracting the data from the VDC and should not be considered as taxonomically certain. *Invalid; now accepted as *Gersemia rubiformis* in WoRMS (2017).

EBSA/Taxonomic Group	Species/Taxon	Number of Presences	VDC Taxon Code
Horse Mussel (*Modiolus modiolus*) Beds	*Modiolus modiolus*	93	4332
Stalked Tunicate Fields	*Boltenia* sp. (*Boltenia ovifera*)	385	1823
Sand Dollar Beds	*Echinarachnius parma*	816	6511
	Clypeasteroida O.	544	6500
Soft Coral Gardens	*Anthomastus grandiflorus*	4	8328
	*Eunephthya rubiformis**	160	8324
	Soft Coral Unidentified	73	8327
***Flabellum* Cup Corals**	*Flabellum alabastrum*	3	8362
	Flabellum sp.	50	8335

As the DFO multispecies trawl surveys are largely limited to the shelf and do not sample the majority of the slope or areas containing rugged bottom, species distribution models for slope species based solely on data from these surveys may be poor predictors of presence if the environmental conditions are largely different in those areas. Ideally the models will contain data from the environmental envelope that the species occurs in. This will reduce the area of extrapolation and give more confidence to model predictions. For such taxonomic groups whose distribution we felt was not fully sampled by the multispecies trawl surveys or when the number of trawl records was insufficient for producing accurate predictions of distribution (e.g. horse mussels, stalked tunicates), random forest models were run using trawl survey data augmented with data from other sources: 1) DFO scallop stock assessment surveys, and 2) *in situ* benthic imagery observations from DFO scientific surveys. Note that absence records were not generated for the benthic imagery observations. Combining data from different gear types introduces bias through differences in catchability and may drive the resulting presence probability prediction surfaces. However, for presence-absence models with data matching the above conditions, we felt that the use of mixed data collection methods was justified in order to expand sampling of the environmental niche of these taxa. We did not extend this to models of biomass where catchability was deemed to more of an issue (see below).

Records from other data sources, such as local ecological knowledge and museum records (see Gass, 2002), the NOAA Deep-Sea Coral Data Portal, and other scientific missions were also considered for each taxonomic group and used for model validation in some cases. Details on the data sources used for random forest modelling are provided separately for each taxonomic group in the Results section below. Considering that there are likely differences in catchability between the records used to train the model and those used for validation, caution must be taken when interpreting the results of model validation.

The presence-absence records used in each random forest model were filtered so that only one presence or absence occurred within a single environmental data raster cell (~1 km). Presence records took precedence over an absence record when both occurred within the same raster cell. For some taxonomic groups, multiple species would often occur in the same raster cell. Since only one presence record was chosen per cell, the species composition listed in Table 2 therefore does not reflect the true number of records for each species found in the VDC.

Biomass (kg) data associated with the DFO multispecies trawl survey records were also extracted from the VDC. To avoid introducing any bias related to differences in catchability between gear types, only biomass data obtained from a single gear type (Western IIA trawl) were used in the random forest models. For each taxonomic group, weights were averaged across multiple tows occurring within the same environmental raster cell.

Species Distribution Modelling using Random Forest

Random forest (Breiman, 2001), a non-parametric machine learning technique, was used to generate probability of occurrence and biomass models for the five taxonomic groups in this report. The random forest modelling methods used in Beazley et al. (2016a) were followed and are not further described here. For taxonomic groups with a highly imbalanced number of presences and absences, models were generated using all the available presence and absence records (i.e., unbalanced species prevalence) and a threshold equal to species prevalence only based on the results of Beazley et al. (2016a). This approach has shown to produce the most realistic presence probability prediction surfaces and highest model accuracy measures compared to models generated using absence data that was randomly down-sampled across the study area (i.e., a balanced species prevalence). However, for the Sand Dollar Beds group, which had a relatively high prevalence, an additional model was generated using a balanced number of presences and absences to serve as a comparison. The results of this model are shown in the Results section below.

Kernel Density Estimation (KDE)

Kernel density estimation (KDE) was used to identify significant concentrations of each taxonomic group from its catch biomass. Only tows made using Western IIA trawl gear were considered. The methodology for the KDE analyses is described in Kenchington et al. (2010; 2014; 2016b) and is not outlined here. Tow position of the significant catches for each taxonomic group can be found in Appendix A. For all groups, the resulting significant area polygons were overlaid on the random forest model outputs to determine the congruence between areas predicted as suitable habitat and the significant concentrations from the catch biomass.

RESULTS

Horse Mussel (*Modiolus modiolus*) Beds

Data Sources and Distribution

The horse mussel *Modiolus modiolus* forms dense bioherms in the Bay of Fundy that have not yet been observed in other areas of the Maritimes Region (Kenchington, 2014; DFO, 2015). In Bay of Fundy, approximately 1500 mussel beds were mapped and measured through the examination of multibeam bathymetry and backscatter strength maps (Kostylev et al., 2009). A high density of horse mussels were found off Margaretsville, north of Digby. This area has since been identified as an EBSA for horse mussel reefs (DFO, 2012). The boundaries of this EBSA are under further consideration and may be revised after targeted surveys conducted in 2017 (D. Fenton, Oceans and Coastal Management Division at BIO, DFO, pers. comm.). Figure 2 shows the distribution of available horse mussel (*Modiolus modiolus*) records in the Maritimes Region. Records are mainly from DFO multispecies trawl and scallop dredge surveys. Several horse mussel records on Western Bank were collected using an Engel 145 otter trawl and video-grab sampler system during a field experiment to determine the effects of trawling on the benthos of Western Bank (Kenchington et al., 2006). Other records on Sable Island, Western, and Emerald Banks were collected between 2003 and 2005 using a video-grab sampler to determine the influence of benthic macrofauna as a structuring agent for juvenile haddock (Rincón and Kenchington, 2016).

Given the relatively low number of presences from the DFO multispecies survey, particularly in the Bay of Fundy area where horse mussels are known to occur, the random forest presence-absence model was generated on both the DFO multispecies trawl survey (19 presences and 1701 absences) and DFO scallop dredge survey records (74 presences and 181 absences). See Table 3 for the data separated by survey and year. The combined dataset consisted of 93 presences and 1882 absences, with a prevalence of 0.05.

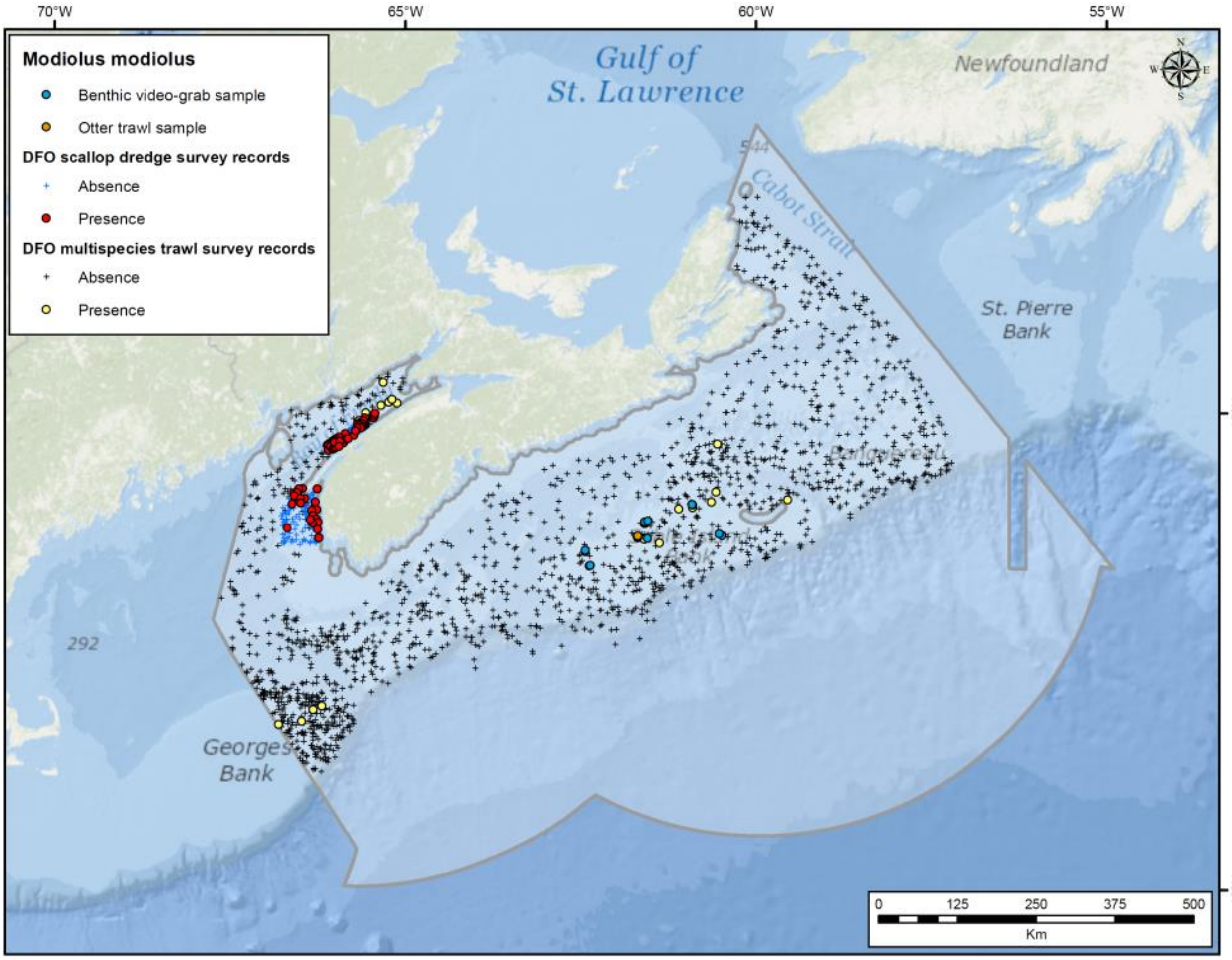

Figure 2. Presence and absence records of *Modiolus modiolus* catch recorded from DFO multispecies trawl surveys conducted between 1999 and 2015, scallop dredge stock assessment surveys conducted in 1997 and 2007, and DFO science records collected using a video-grab sampler between 2003 and 2005, and an otter trawl between 1997 and 1999 in the Maritimes Region.

Table 3. Number of presence and absence records of *Modiolus modiolus* catch in the Maritimes Region from DFO scallop dredge surveys in 1997 and 2007, and DFO multispecies trawl surveys from 1999 to 2015. N/A indicates that the presence data from the survey in this year was outside the study extent, but absences are considered valid and used in the model.

Survey	Year	Total number of presences	Total number of absences
DFO scallop dredge survey	1997	45	91
	2007	29	90
DFO multispecies trawl survey	1999	1	104
	2000	1	95
	2005	NA	44
	2009	4	264
	2010	3	254
	2011	3	253
	2012	2	266
	2013	5	354
	2014	NA	34
	2015	NA	33
	Total:	**93**	**1882**

Random Forest Model Results

The random forest model on the presence-absence *Modiolus modiolus* data had an excellent performance rating, with an average AUC of 0.918 ± 0.074 SD (Table 4). Class error for both the presence and absence classes was low, however the presence records were incorrectly classified more than the absence records. Sensitivity and specificity of this model were high (0.839 and 0.874, respectively).

The predicted presence probability surface of *M. modiolus* from the model is shown in Figure 3. Much of the Maritimes Region was predicted to have either low or zero presence probability of horse mussels. Small patches of high presence probability occurred along the Nova Scotia coast in Bay of Fundy. These areas corresponded to the location of the DFO scallop dredge survey records (Figure 4). Small patches of elevated presence probability occurred on Sable Island Bank and along the Scotian Slope. The deep waters beyond the Scotian Shelf were considered areas of extrapolation by the model. Areas considered extrapolated by the model also occurred in small patches on the Scotian Shelf, but did not do not overlap with areas of elevated presence probability.

Figure 5 depicts the predicted distribution of *M. modiolus* based on a prevalence threshold of 0.05. In this map, all presence probability values greater than 0.05 were classified as presence, while values less than 0.05 were classed as absence. The majority of the study area was classified as unsuitable habitat for *M. modiolus*. Much of the coast of Nova Scotia and Bay of Fundy was predicted as suitable habitat for *M. modiolus*.

Environmental variables that measure the oceanographic conditions and primary production in surface waters performed well in this model (see Figure 6). Annual Chlorophyll *a* Minimum was predicted as the top environmental variable. Prior to spatial interpolation this variable displayed a highly right-skewed spatial distribution (Beazley et al., in prep.). Examination of the Q-Q plot revealed a strong spatial pattern to those data points over- and under-predicted by a normal distribution, with over-predicted values located tightly along the coast of Nova Scotia and in the deepest portion of the study area, and under-predicted values located inshore along the width of the study extent. Annual Chlorophyll *a* Minimum was followed in importance by Surface Salinity Average Range and two other sea surface chlorophyll *a* variables. The partial dependence of *M. modiolus* on the top 6 most important variables is shown in Figure 7. Presence probability was highest at Annual Chlorophyll *a* Minimum values of 2 mg m^{-3} and higher. These values are associated with those values along the coast of Nova Scotia and in Bay of Fundy that were over-predicted by a normal distribution (see Beazley et al., in prep.). Values in this range were well-predicted by the kriging model.

Table 4. Accuracy measures and confusion matrix from 10-fold cross validation of a random forest model of presence and absence of *Modiolus modiolus* within the Maritimes Region. Observ. = Observations; Sensit. = Sensitivity, Specif. = Specificity.

Model Fold	**AUC**	**Observ.**	**Predictions**		**Total n**	**Class error**	**Sensit.**	**Specif.**
1	0.901		**Absence**	**Presence**				
2	0.968	**Absence**	1645	237	1882	0.123	0.839	0.874
3	0.761	**Presence**	15	78	93	0.161		
4	0.971							
5	0.978							
6	0.868							
7	0.995							
8	0.927							
9	0.848							
10	0.968							
Mean	**0.918**							
SD	**0.074**							

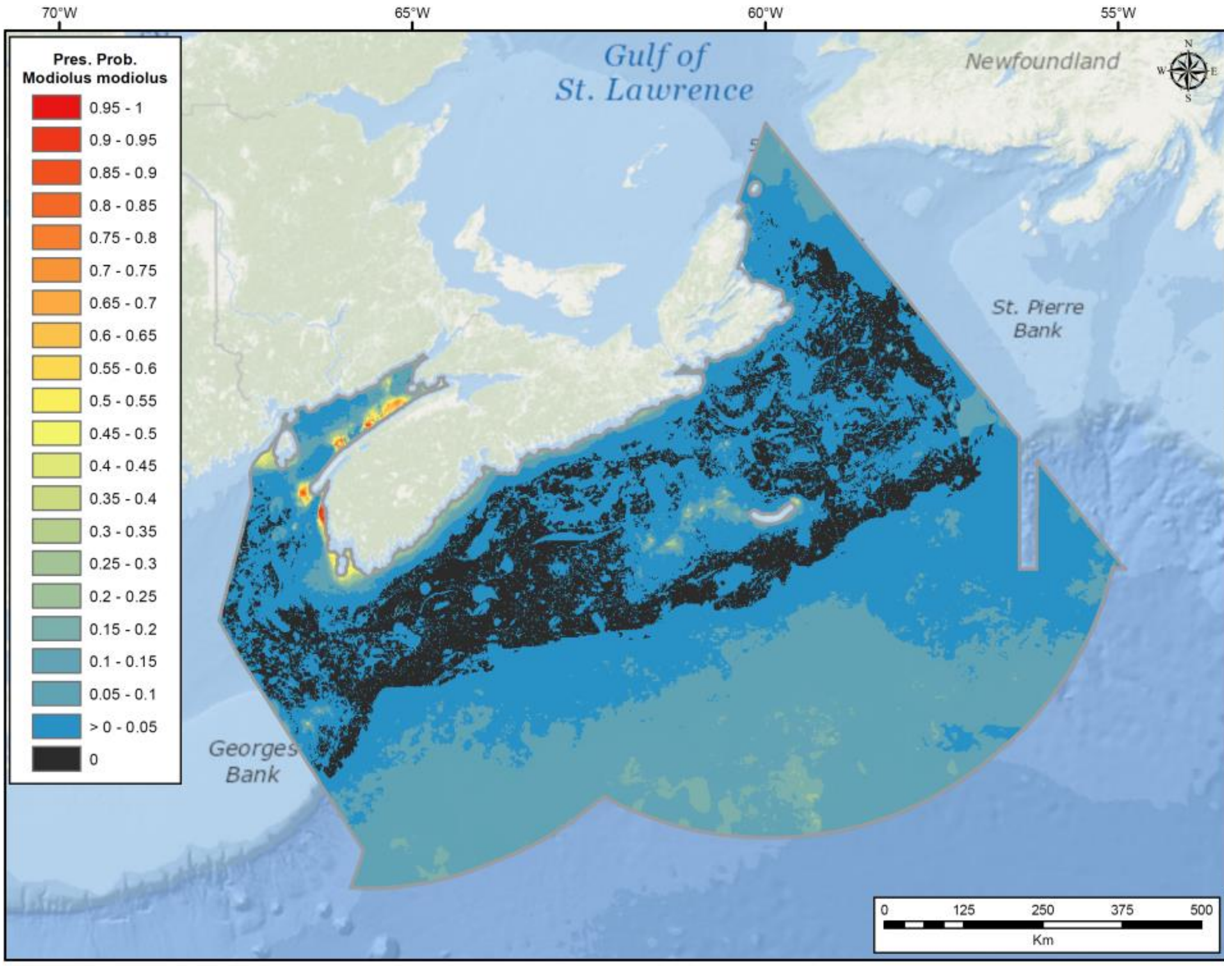

Figure 3. Predictions of presence probability (Pres. Prob.) of *Modiolus modiolus* based on a random forest model on *M. modiolus* presence and absence data collected from DFO multispecies trawl surveys and DFO scallop dredge conducted in the Maritimes Region between 1997 and 2015.

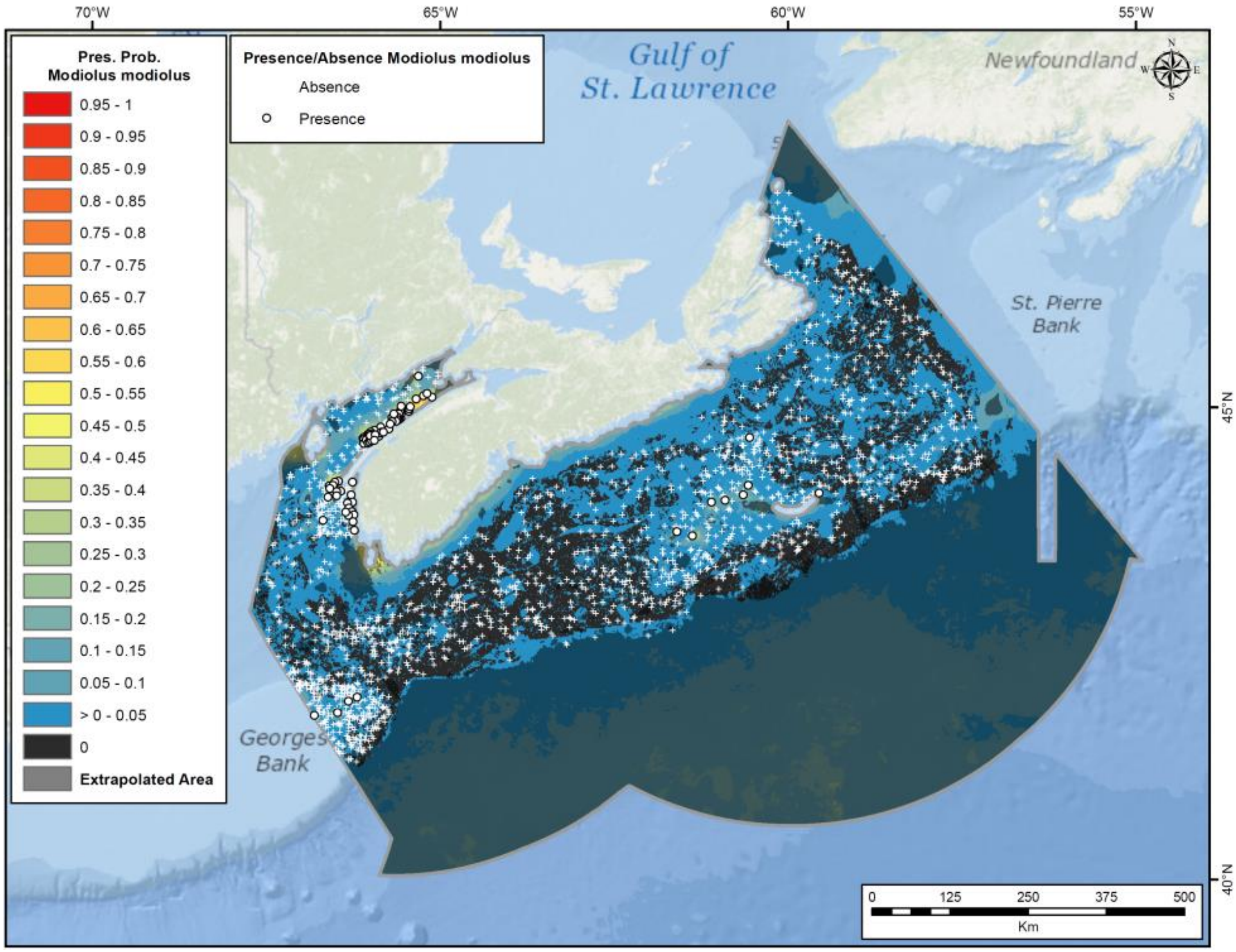

Figure 4. Presence and absence observations and predictions of presence probability (Pres. Prob.) of *Modiolus modiolus* based on a random forest model on presence and absence *M. modiolus* catch data collected from DFO multispecies trawl surveys and DFO scallop dredge conducted in the Maritimes Region between 1997 and 2015. Also shown are the areas of model extrapolation.

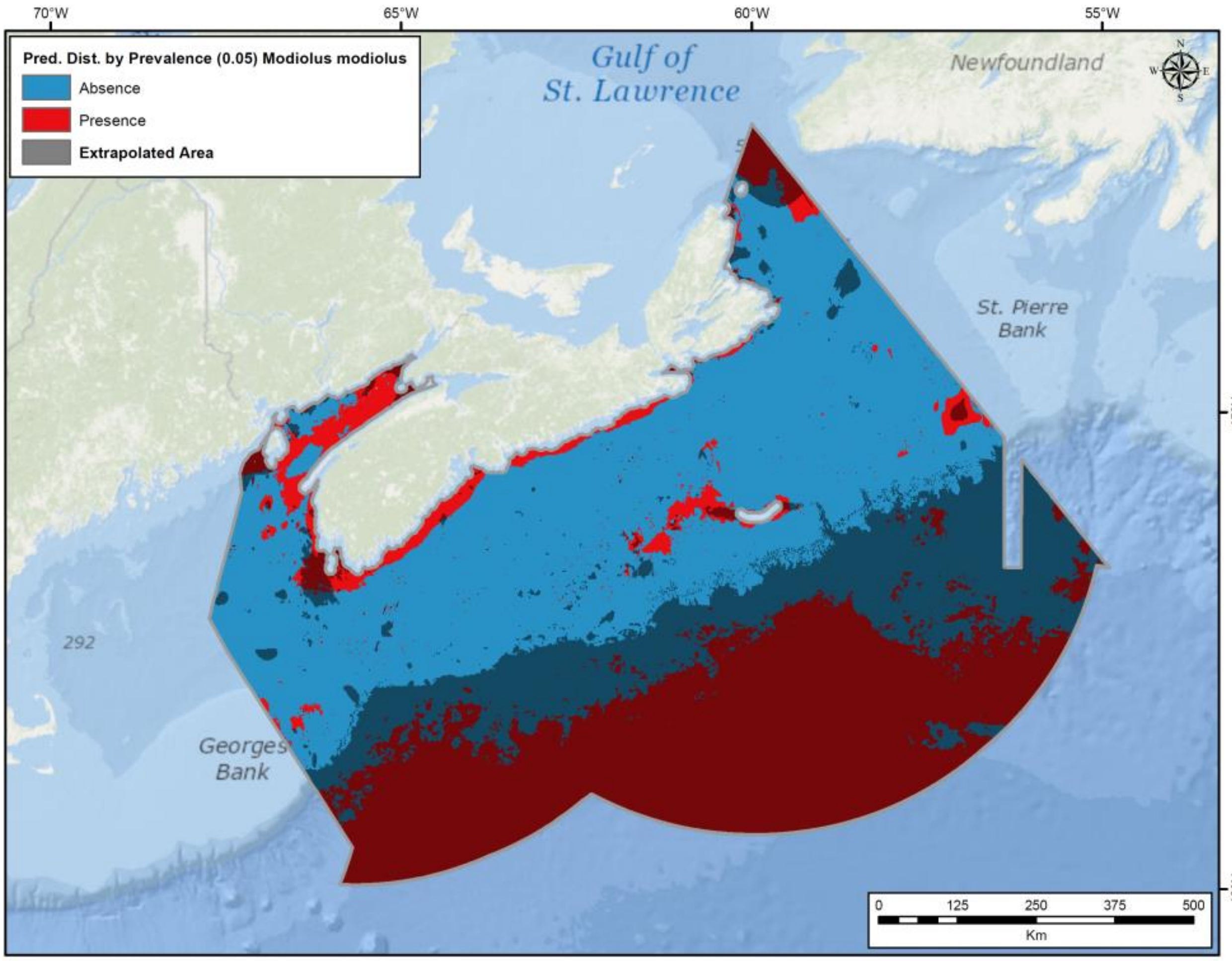

Figure 5. Predicted distribution (Pred. Dist.) of *Modiolus modiolus* in the Maritimes Region based on the prevalence threshold (0.05) of *M. modiolus* presence and absence data used in the random forest model. Also shown are the areas of model extrapolation (grey polygon may appear red or blue).

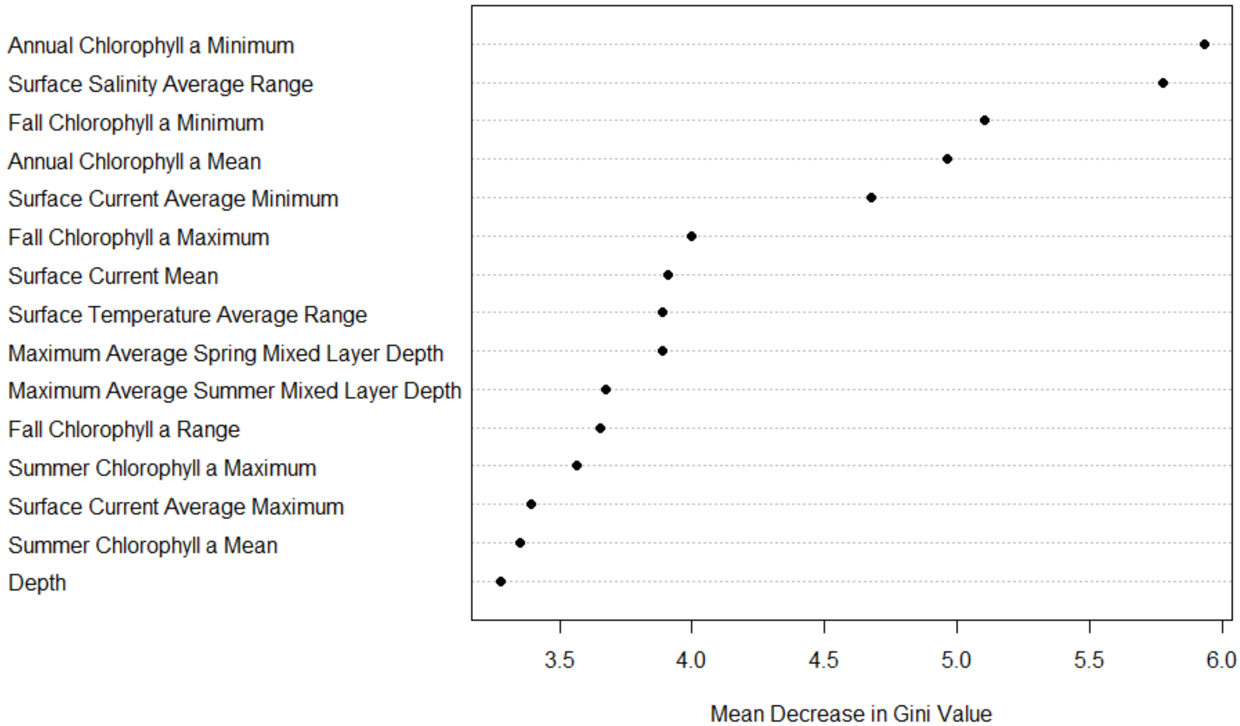

Figure 6. Importance of the top 15 predictor variables measured as the Mean Decrease in Gini value of the random forest model on *Modiolus modiolus* presence and absence data in the Maritimes Region. The higher the Mean Gini value the more important the variable is for predicting the response data.

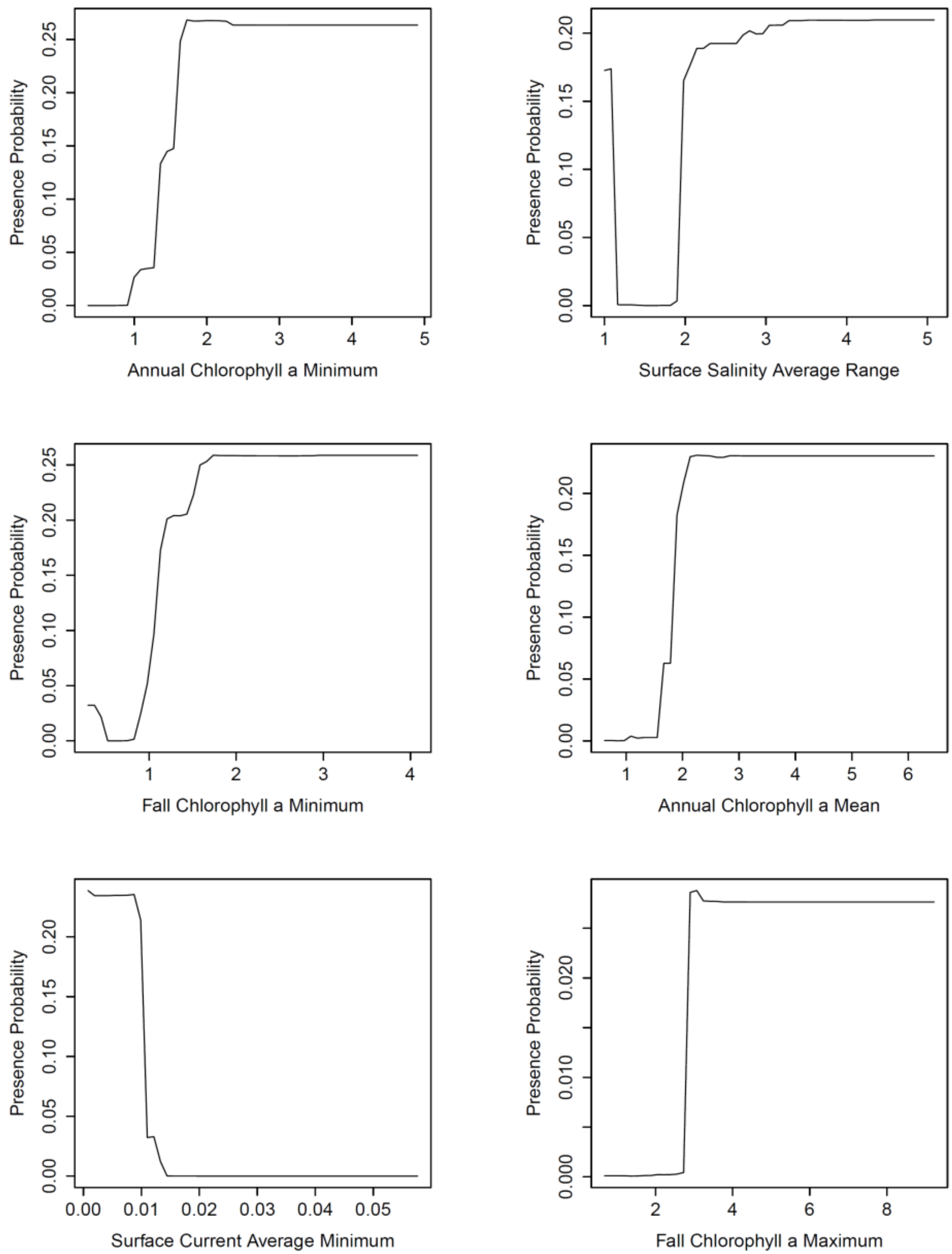

Figure 7. Partial dependence plots of the top six predictors from the random forest model of *Modiolus modiolus* presence and absence data collected within the Maritimes Region, ordered left to right from the top. Presence probability is shown on the *y*-axis.

Validation of Model Using Independent Data

Additional *Modiolus modiolus* records from DFO video-grab samples collected between 2003 and 2005 on Sable Island, Western, and Emerald Banks, and otter trawl samples collected between 1997 and 1999 on Western Bank were used for validating the prevalence results (see Figure 8). These records were located west of Sable Island and in the vicinity of DFO multispecies trawl survey records used to train the random forest model. Nonetheless, most were predicted to occur in areas of unsuitable habitat (as 'absence' by the model). Of the 15 video-grab samples, 11 (73%) were predicted as absence based on the prevalence threshold of 0.05. These records were distributed in large areas of unsuitable habitat, but were surrounded with small pockets of 'presence'. The 4 records predicted as 'presence' were distributed in small patches of suitable habitat on Sable Island and Western Banks. Of the 6 *M. modiolus* records from the experimental otter trawl survey, 4 (67%) were predicted as 'absence' by the model (see inset map of Figure 8). Many of these records occurred along the edge of areas predicted as suitable habitat.

The high number of records predicted as absence do not necessarily invalidate the model. *M. modiolus* can form semi-infaunal reefs as well as infaunal reefs, the latter associated with coarser sediments and strong currents (http://www.ukmarinesac.org.uk/communities/biogenic-reefs/br1_2_3.htm). There is a range of habitats produced depending upon the environment and the age of the habitat. With time, both can develop into bioherms. Grab samples will capture infaunal mussels that would not be captured with trawl gear if bioherms were not present. Therefore it is not surprising that grab samples would show false absence in this validation. Further, RF models often show very convoluted boundaries between presence and absence at small scales that should not be over-interpreted (Kenchington et al., 2016a). As in Figure 8, most of the otter trawl catches used for validation cluster around one of these areas of small scale imprecision and should not be considered to invalidate the model results.

As a means of model validation, the horse mussel reef EBSA identified in the Bay of Fundy off Margaretsville, N.S. (DFO, 2012) was overlaid on the prevalence surface generated from the random forest model (see Figure 9). The entire EBSA fell over a large area predicted as suitable habitat (presence of horse mussels) based on species prevalence.

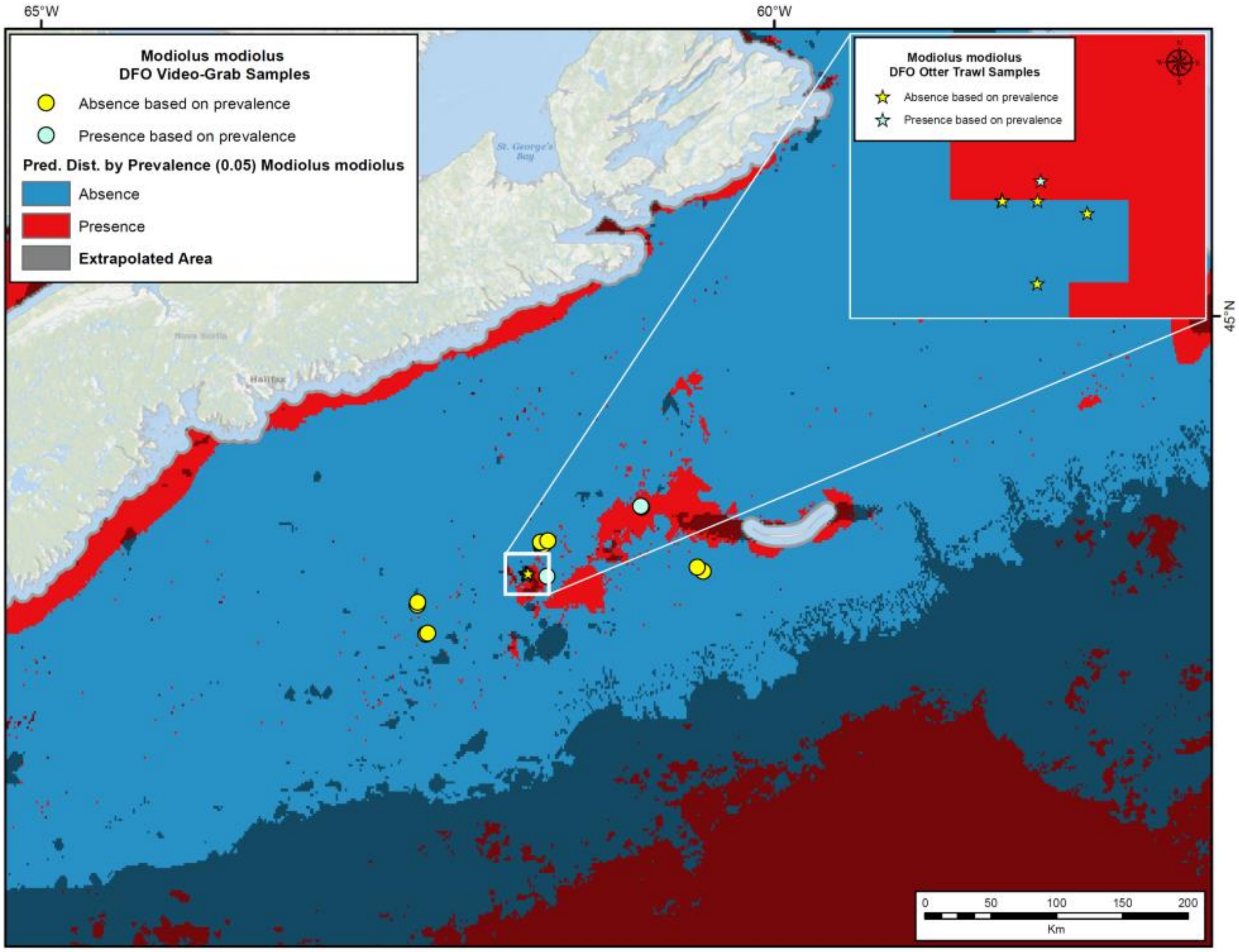

Figure 8. Validation of the prevalence surface generated from random forest on the *Modiolus modiolus* presence-absence records using DFO video-grab samples collected between 2003 and 2005 (large background map), and DFO otter trawl records collected between 1997 and 1999 (stars in inset map) on the Scotian Shelf.

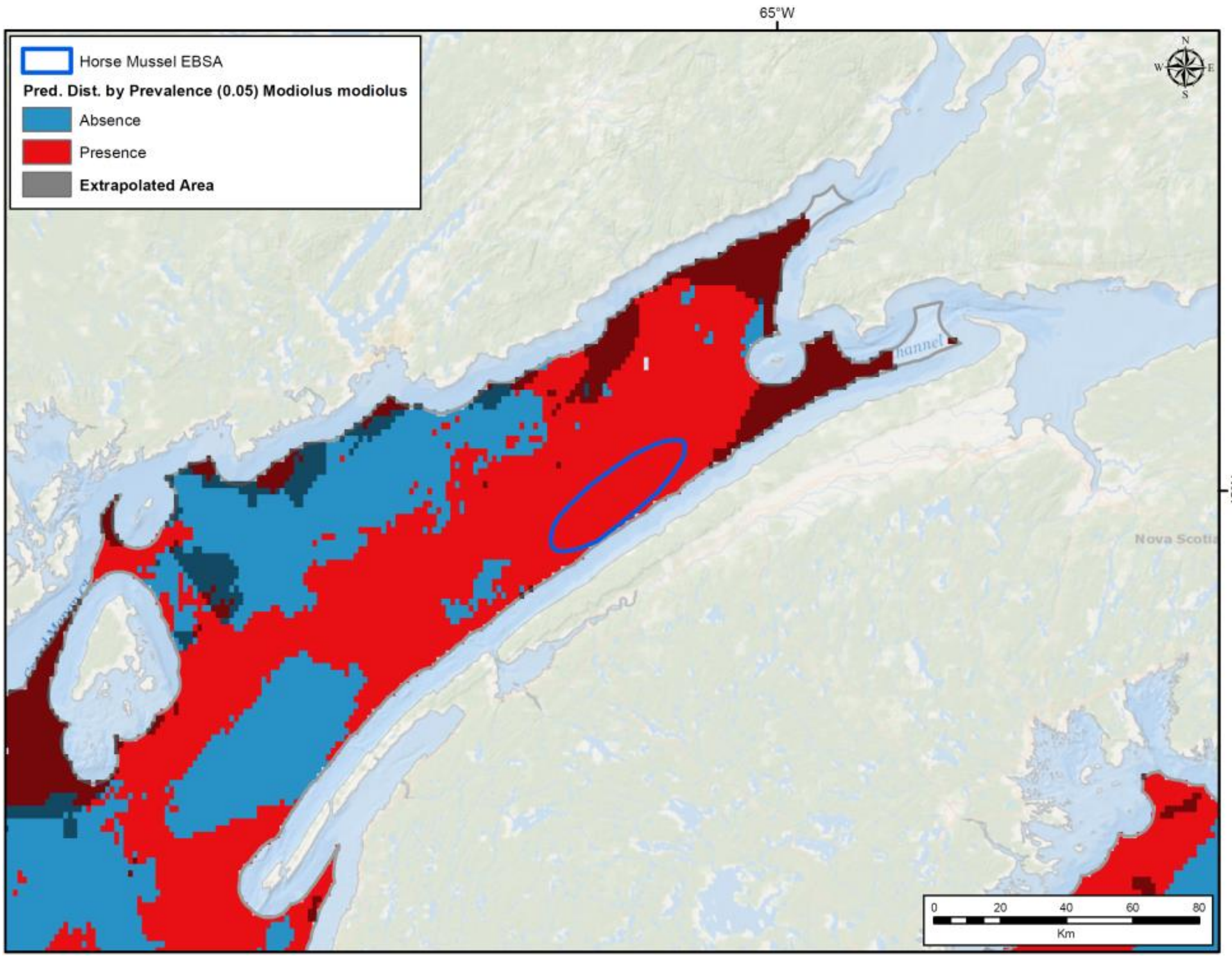

Figure 9. Horse mussel reef EBSA located off Margaretsville, N.S. in the Bay of Fundy (DFO, 2012) overlaid on the prevalence surface generated from the random forest model on *Modiolus modiolus* presence-absence records.

Prediction of *Modiolus modiolus* Biomass Using Random Forest

The accuracy measures of the regression random forest model on mean *Modiolus modiolus* biomass per grid cell from DFO multispecies trawl surveys are presented in Table 5. The highest R^2 was 0.004 while the average was 9.020 x 10^{-4} ± 0.001 SD, indicating poor model performance. The average Normalized Root-Mean-Square Error was 0.015 ± 0.023 SD. The high standard deviation indicates high variability between model folds. Percent variance explained was negative.

Figures 10 and 11 show the predicted biomass surface of *M. modiolus*. The majority of the spatial extent was predicted to have zero biomass of *M. modiolus*. Small pockets of high (up to 14.69 kg) biomass were predicted on Sable Island Bank and off southwestern Nova Scotia, that were coincident with the location of the large horse mussel catches there (see Figure 11).

The top 15 most important environmental variables for predicting *M. modiolus* biomass are shown in Figure 12. Slope and Depth (two non-interpolated variables) were the top predictors in the model. These were followed more distantly by Bottom Salinity Average Minimum and the other variables in the model. The partial dependence plots are shown in Figure 13. *M. modiolus* predicted biomass was highest at the lowest slopes (< 1˚) and shallowest depths (< 100 m).

Table 5. Accuracy measures from 10-fold cross validation of a random forest model of average *Modiolus modiolus* biomass (kg) per grid cell recorded from DFO multispecies trawl surveys in the Maritimes Region. RMSE = Root-Mean-Square Error; NRMSE = Normalized Root-Mean-Square Error.

Model Fold	**R^2**	**RMSE**	**NRMSE**	**Percent (%) variance explained**
1	8.150 x 10^{-5}	0.290	0.010	-9.08
2	0.004	0.323	0.012	-10.76
3	1.179 x 10^{-4}	0.118	0.004	-8.18
4	3.434 x 10^{-4}	0.168	0.006	-7.84
5	NA	0.103	0.004	-9.37
6	0.003	2.166	0.078	-10.39
7	8.231 x 10^{-6}	0.303	0.011	-5.74
8	1.354 x 10^{-4}	0.157	0.006	-8.82
9	3.992 x 10^{-4}	0.307	0.011	-7.17
10	0.001	0.123	0.004	-8.47
Mean	**9.020 x 10^{-4}**	**0.406**	**0.015**	**-8.58**
SD	**0.001**	**0.625**	**0.023**	**1.48**

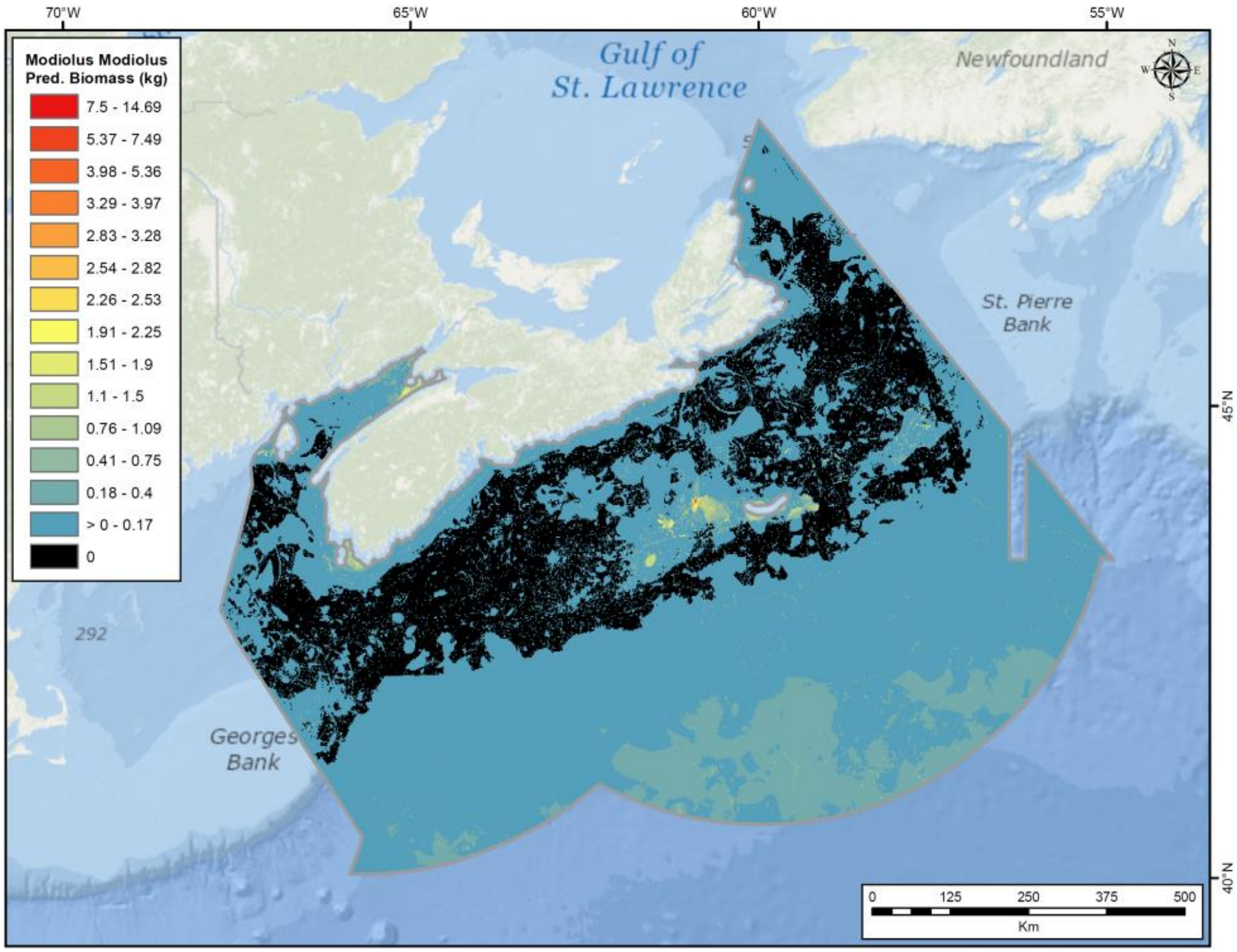

Figure 10. Predictions of biomass (kg) per grid cell of *Modiolus modiolus* from catch data recorded in DFO multispecies trawl surveys conducted in the Maritimes Region between 1999 and 2015.

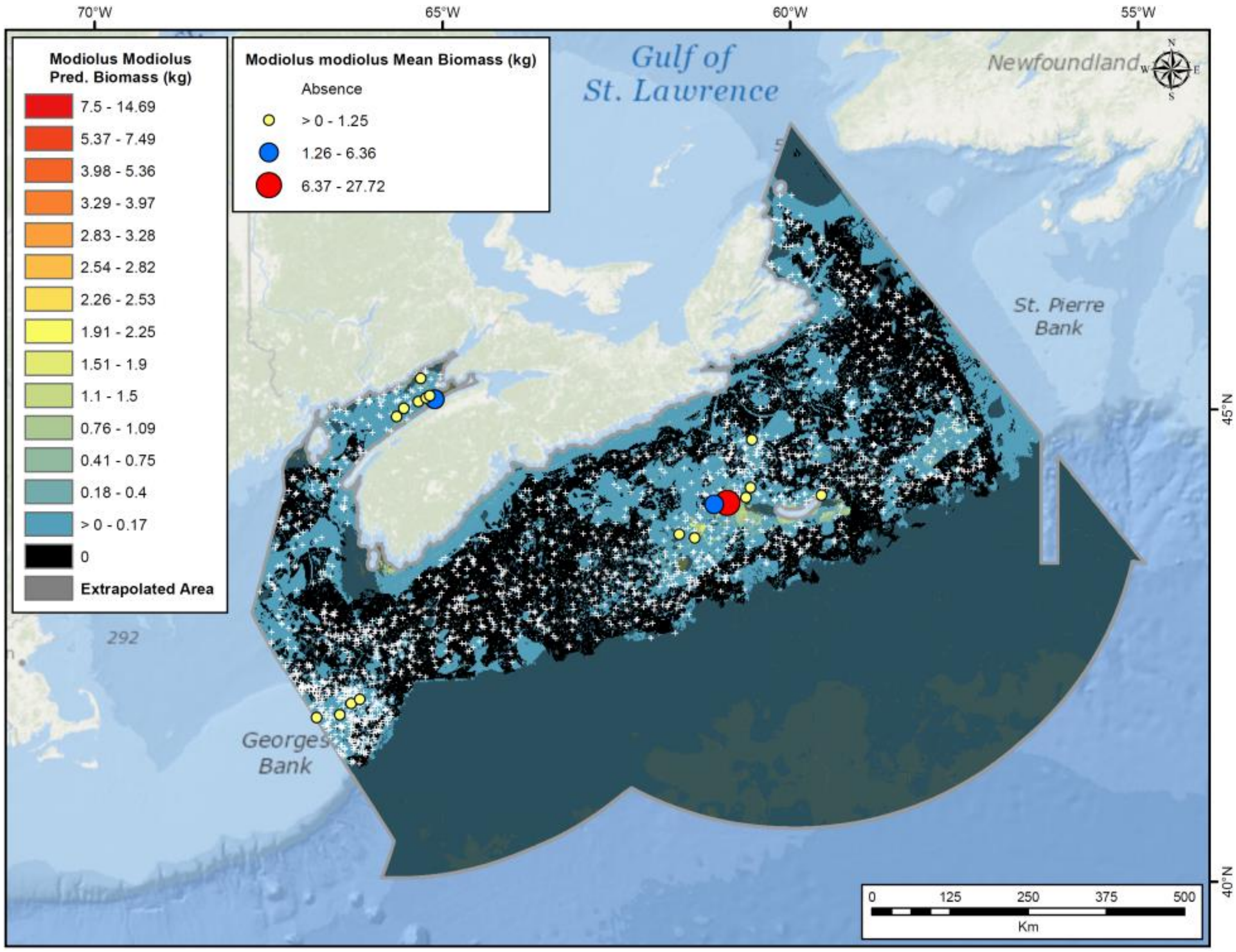

Figure 11. Predictions of biomass (kg) per grid cell of *Modiolus modiolus* from catch data recorded in DFO multispecies trawl surveys conducted in the Maritimes Region between 1999 and 2015. Also shown are mean biomass values per grid cell and areas of model extrapolation.

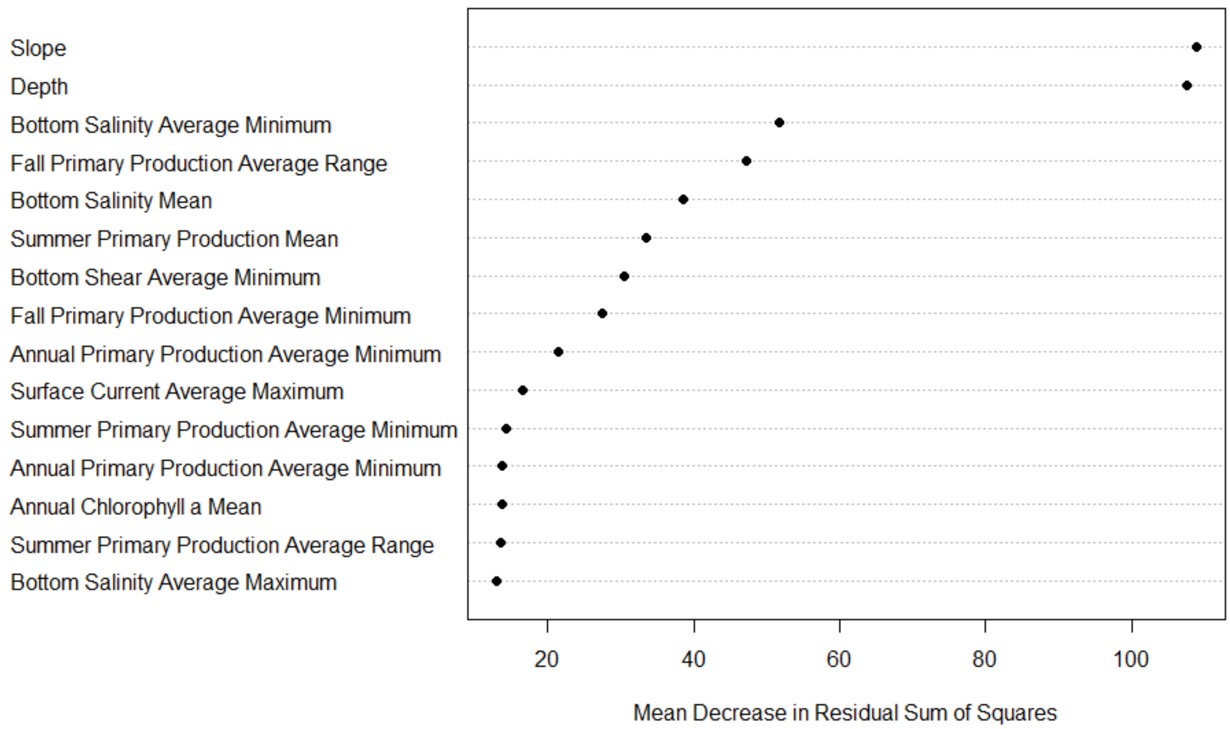

Figure 12. Importance of the top 15 predictor variables measured as the Mean Decrease in Residual Sum of Squares of the regression random forest model on *Modiolus modiolus* mean biomass. The higher the Mean Decrease in Residual Sum of Squares the more important the variable is for predicting the response data.

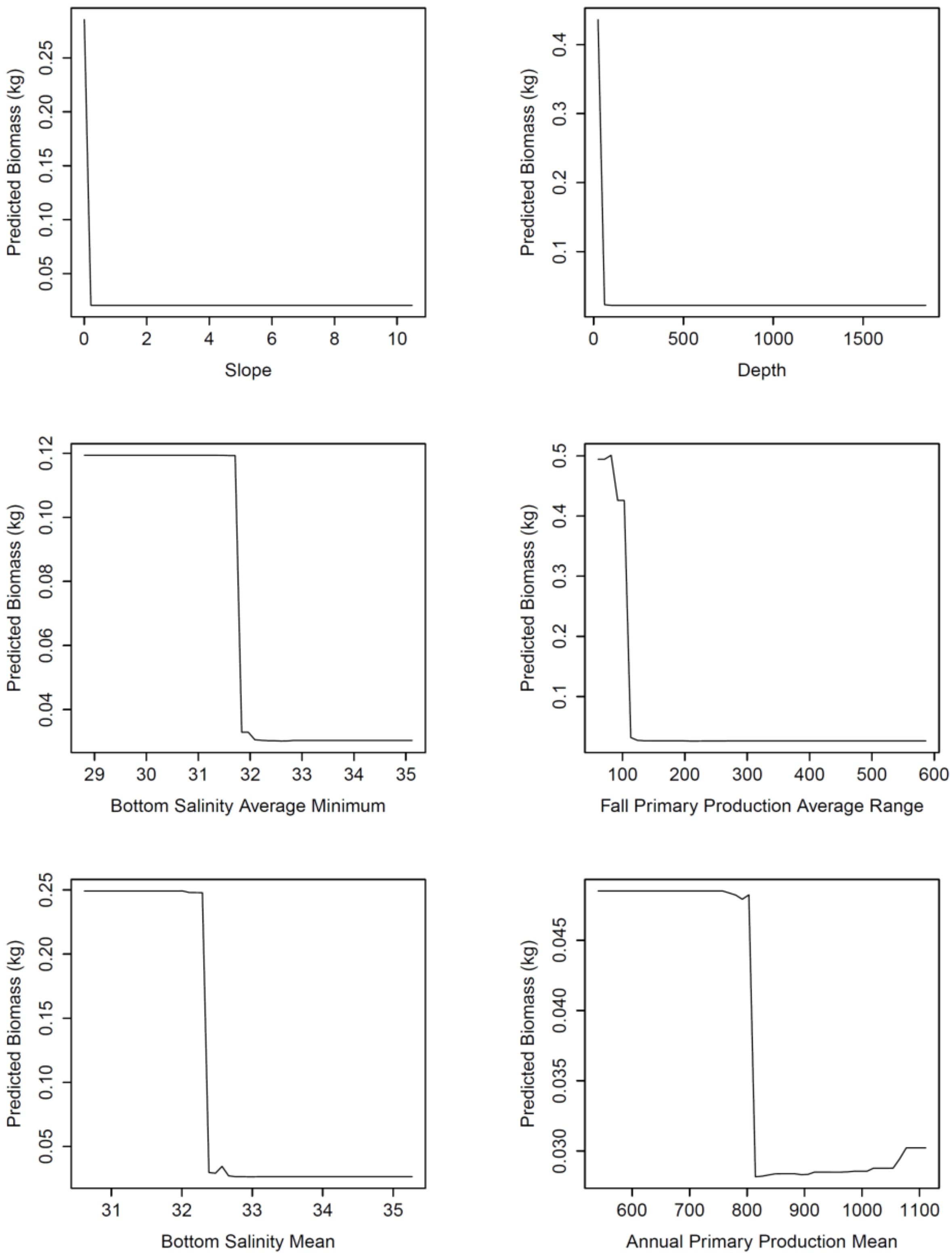

Figure 13. Partial dependence plots of the top six predictors from the random forest model of *Modiolus modiolus* biomass data collected within the Maritimes Region, ordered left to right from the top. Predicted biomass (kg) is shown on the *y*-axis.

Kernel Density Analysis of *Modiolus modiolus* Catch Data

Kernel density analysis of the *Modiolus modiolus* biomass data showed highly discontinuous concentrations on the eastern Scotian Shelf, with smaller concentrations in the Bay of Fundy and on Georges Bank. The largest concentrations were observed on Sable Island Bank (Figure 14), but this may represent differences in catchability on different substrates. The KDE search radius for this group was 12.9 km. Polygons encompassing each of 9 catch thresholds showed that the greatest increase in area between successive weights, after the initial mapping out of the habitat, occurred between 0.25 kg and 0.1 kg, where the area increased by 303% (Table 6, Figure 15). There were only 5 data points contributing to this expansion but this was the greatest number for any potential threshold as the sample size was small (Table 6, Figure 16). Consequently 0.25 kg was considered as the threshold delineating the *M. modiolus* beds. These beds occupy 417 km^2 of seabed, mostly in the Bay of Fundy, with single catches on the eastern Scotian Shelf. However, the catch data are not considered reliable given that this species can bury in soft sediment. The mussels in the Bay of Fundy are known to produce reefs, whereas no reefs have been identified on the Scotian Shelf. KDE analyses may not be appropriate for defining habitats for this species and *in situ* validation is needed. Figure 17 shows the final KDE significant area polygons, the location of significant and non-significant catches, and null records of *M. modiolus* from the DFO multispecies trawl surveys.

All KDE polygons occurred in areas predicted as suitable habitat (i.e. presence of *M. modiolus*) when overlain on the predicted distribution by prevalence surface (Figure 18, upper panel). The area predicted as suitable habitat by the RF model was much larger in spatial extent than the KDE, particularly in Bay of Fundy. The KDE polygons were not compared to the RF biomass results for this species due to the poor performance of that model.

The large significant area polygon in Bay of Fundy only partially overlapped with the horse mussel EBSA identified in DFO (2012) (Figure 18, lower panel). The boundaries of this polygon are considered inaccurate and will undergo revision after targeted *in situ* data collection surveys in 2017 (D. Fenton, pers. comm.).

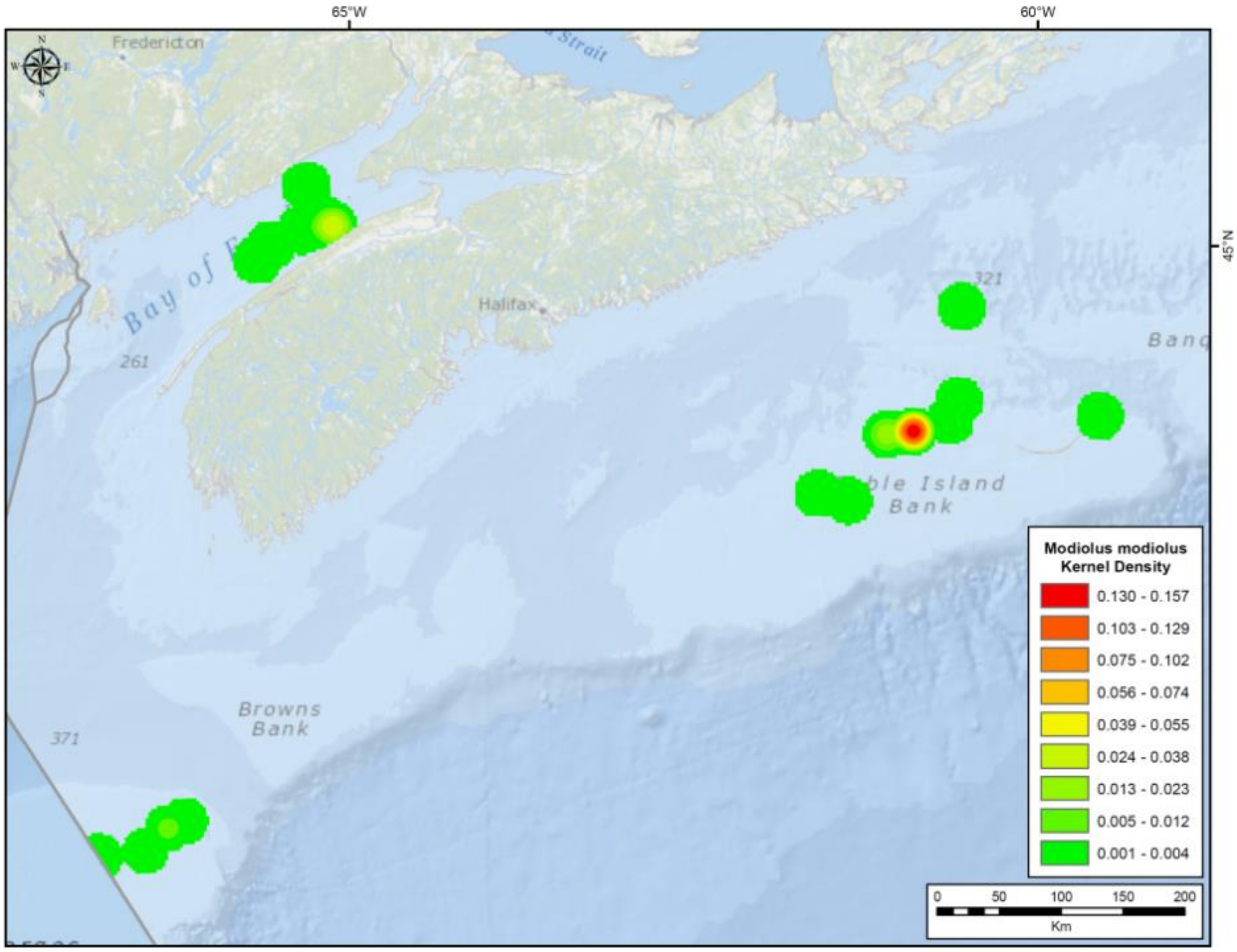

Figure 14. Kernel density analysis of *Modiolus modiolus* catches in the Maritimes Region.

Table 6. Summary characteristics for nine *Modiolus modiolus* catch weight thresholds. The threshold corresponding to the habitat delineation is indicated by shading.

Number of Points in Polygon	Area (km^2)	Catch Threshold (kg)	Percent Change in Area (%)
2	2	5	79.94
3	3	3	58.41
4	5	1	0.00
4	5	.75	8684.99
8	410	.50	1.67
11	417	.25	303.24
16	1682	.10	38.03
18	2322	.075	0.71
19	2339	.050	

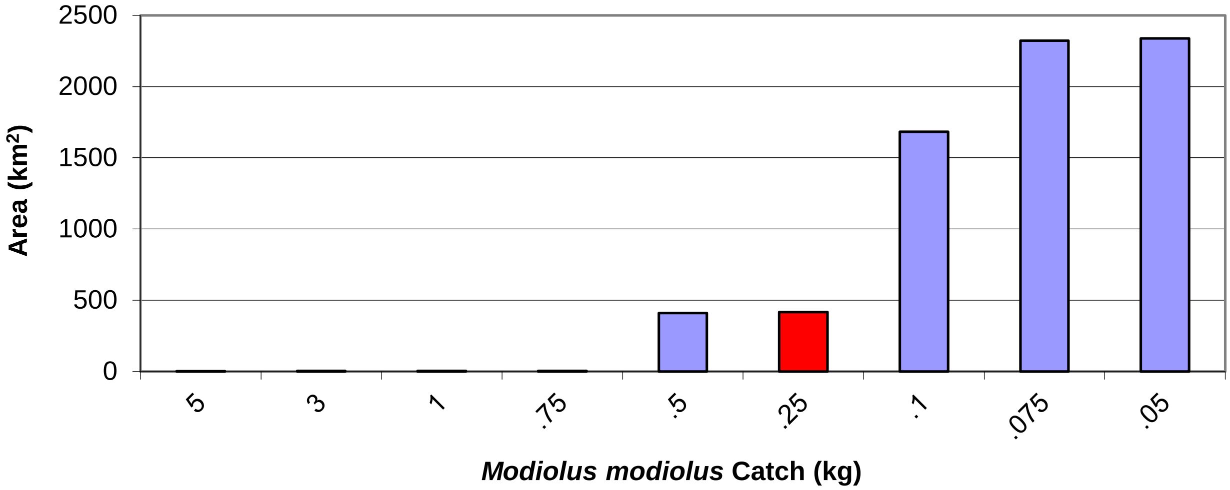

Figure 15. Area occupied by polygons encompassing catches of horse mussels greater than or equal to each of 9 catch thresholds (5 kg to 0.05 kg).

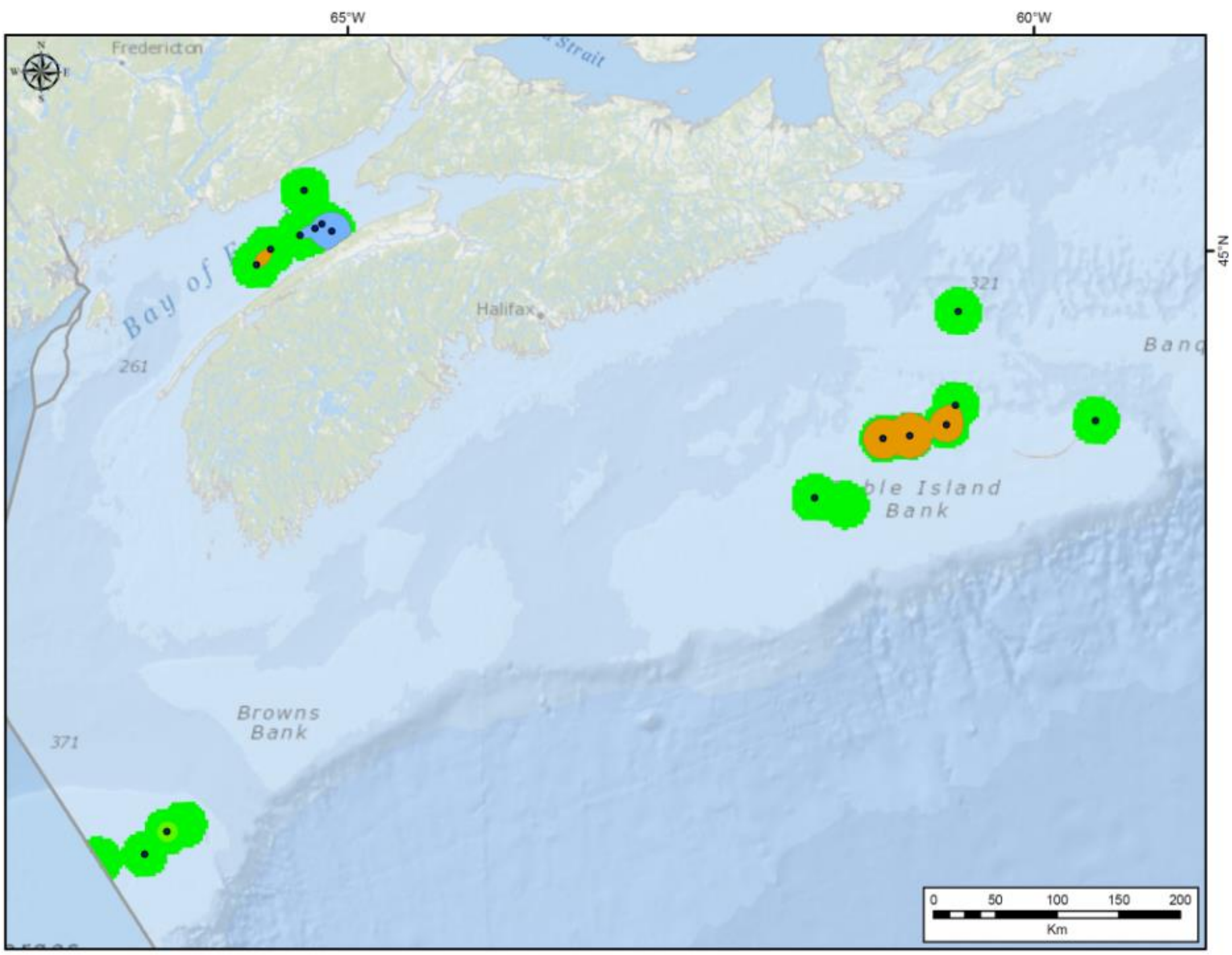

Figure 16. Polygons encompassing *Modiolus modiolus* catch weights greater than or equal to 0.25 kg (blue) and 0.1 kg (orange) overlain on the kernel density surface. Tow locations associated with biomass values above these weight thresholds are also shown. The 0.25 kg threshold was used to delineate significant concentrations.

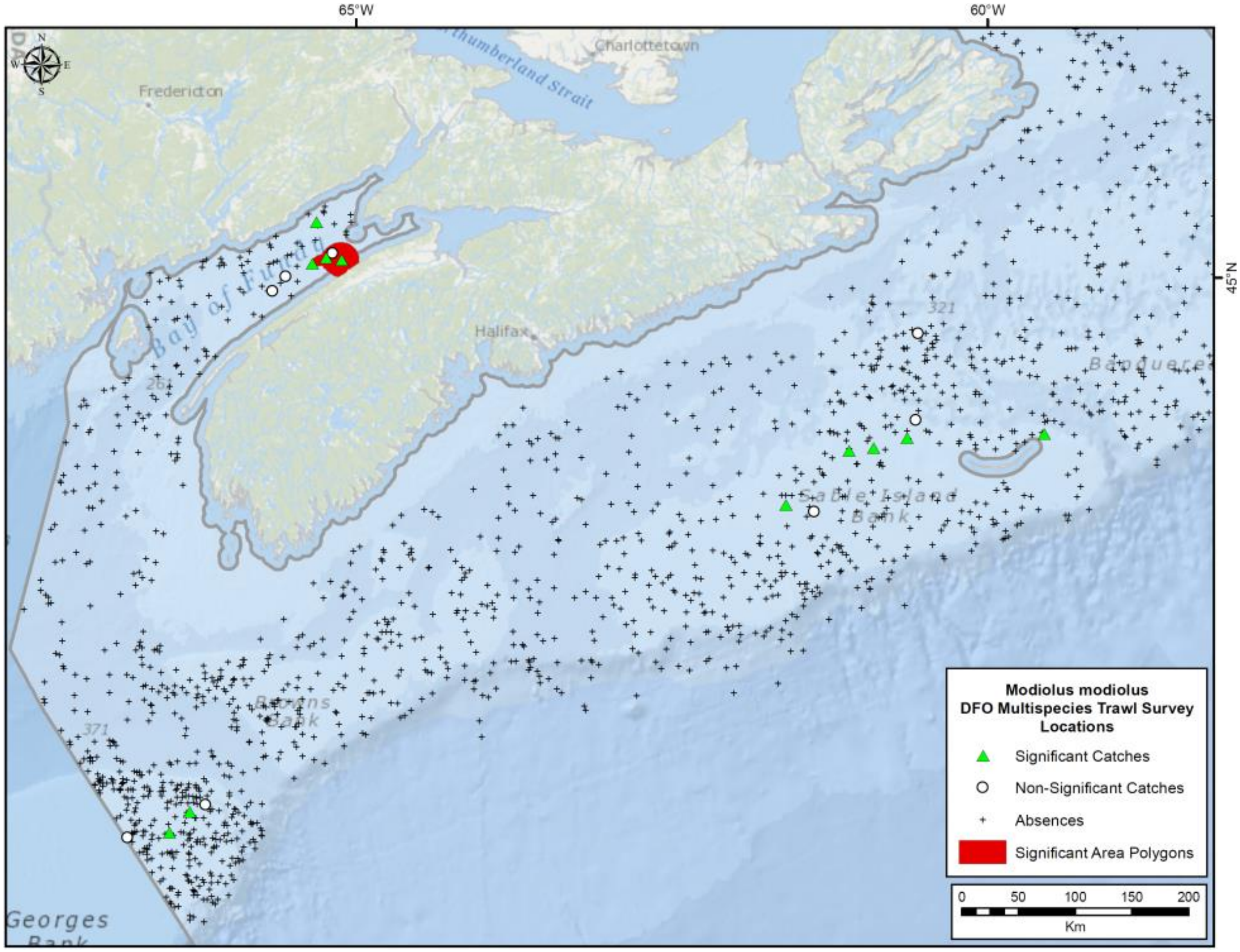

Figure 17. Location of polygons identifying significant *Modiolus modiolus* concentrations in the Maritimes Region. Also shown are the significant and non-significant catches of *M. modiolus* and absence data (nulls) from the DFO multispecies trawl surveys.

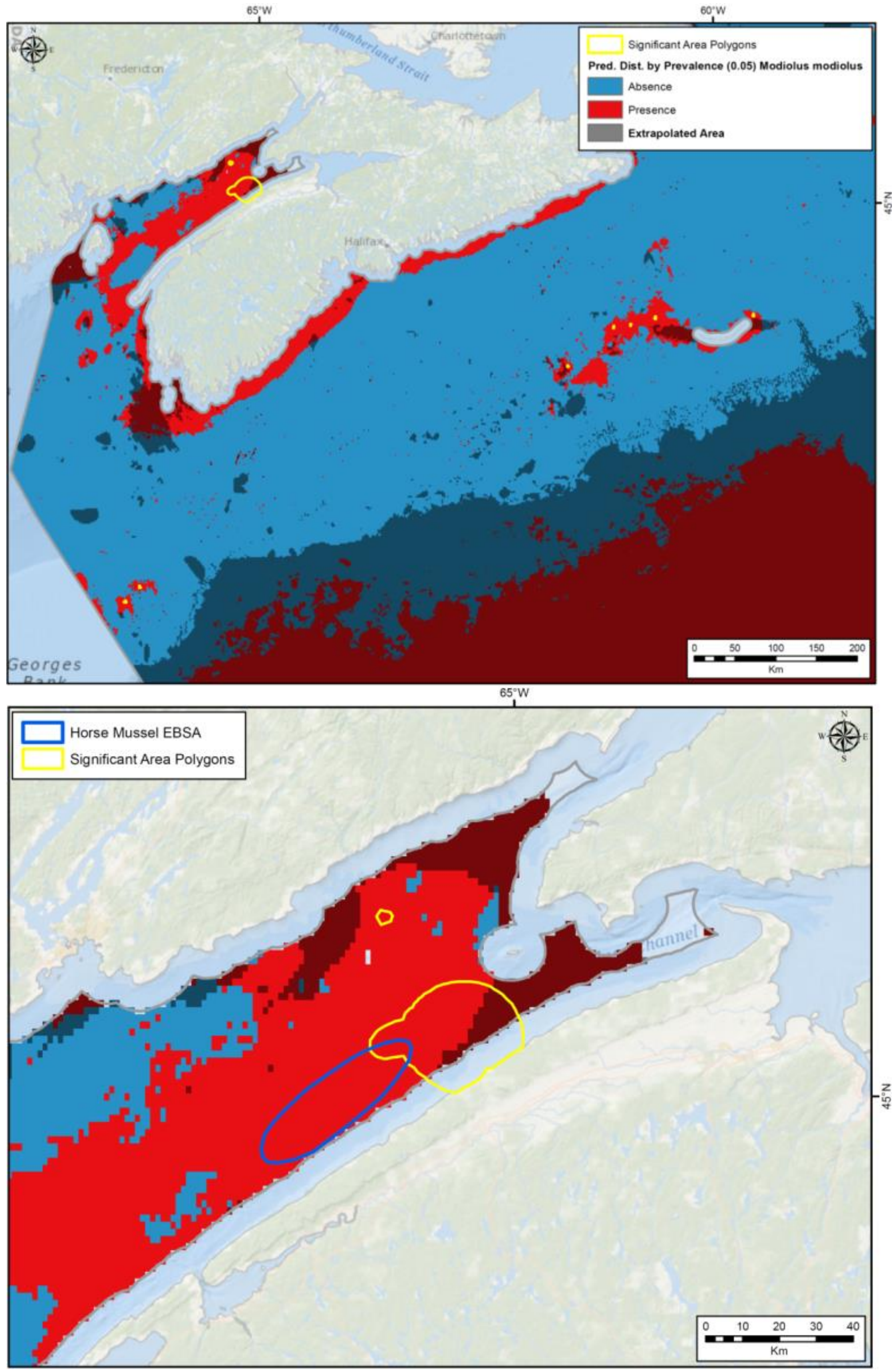

Figure 18. KDE polygons denoting significant concentrations of *Modiolus modiolus* overlain on the predicted distribution by prevalence (upper panel) of *M. modiolus* generated from the random forest model. Lower panel shows the overlap between the KDE significant area polygons in Bay of Fundy (yellow) and the horse mussel reef EBSA (blue) identified in DFO (2012).

Stalked Tunicate Fields (*Boltenia ovifera*)

Data Sources and Distribution

In the DFO multispecies trawl surveys, tunicates are not easily identified below the subphylum level (Tunicata). One species of stalked tunicate is caught in the trawls and is identified as *Boltenia* sp. As only one stalked *Boltenia* species (*Boltenia ovifera*) has been identified from the region, we are confident that all samples from the multispecies trawl surveys are *B. ovifera*, and are referred to as such herein.

Boltenia ovifera is widely distributed in the Arctic and North Atlantic between 10 and 300 m depth (Plough, 1969). This species forms significant concentrations in the Bay of Fundy and the shallow waters off Halifax Nova Scotia (Francis et al., 2014; DFO, 2015), reaching densities of up to 100 individuals per m^{-2} in some areas (Francis et al., 2014). In a localized area off Halifax, generalized additive models (GAMs) predicted that depth, substrate type, and benthic algal type were strong determinants of *B. ovifera* abundance (Francis et al., 2014). The current works represents the first attempt at modelling the distribution of this species on the scale of the entire Maritimes Region. Species distribution models using both random forest and GAMs were developed for *B. ovifera* in the Gulf of St. Lawrence with very good success (see Murillo et al., 2016).

Figure 19 shows the distribution of available *B. ovifera* records in the Maritimes Region. Presence and absence data are from the DFO multispecies trawl surveys and DFO scallop dredge surveys only; there are no *in situ* records of *B. ovifera* for the region. DFO trawl survey records were concentrated in the eastern portion of the study extent off Cape Breton and on Misaine Bank. Several other records occurred on Georges Bank and in LaHave Basin. Given the relatively low number of presences of *B. ovifera* from the DFO multispecies survey, particularly in the Bay of Fundy area where *B. ovifera* is known to occur, random forest presence-absence models were generated using both the DFO multispecies trawl survey (361 presences and 2083 absences) and DFO scallop dredge survey records (24 presences and 225 absences) in order to widen the environmental niche represented in the model for this species. See Table 7 for the data separated by survey and year. The use of records from the different gear types will introduce some catchability bias into the model.

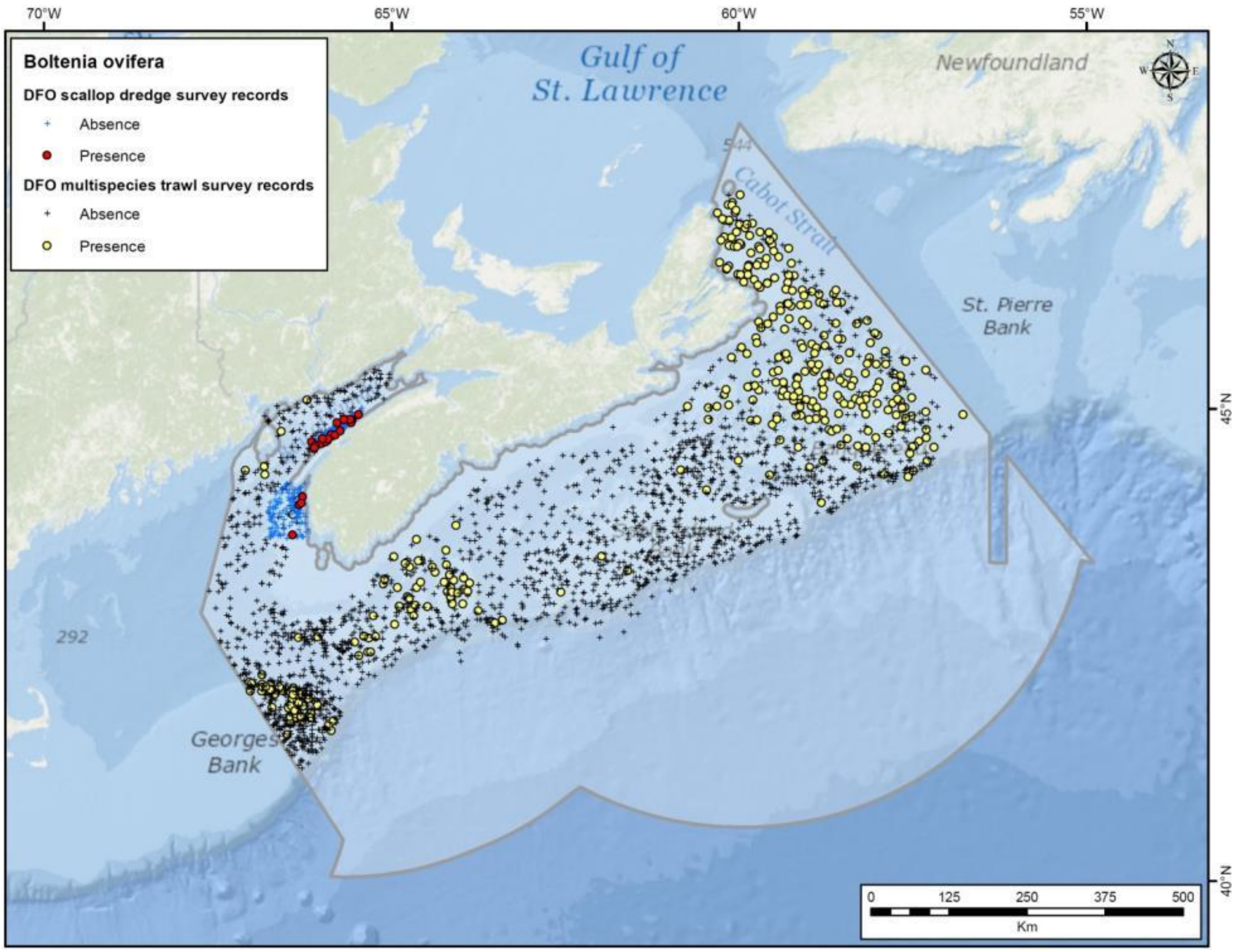

Figure 19. Presence and absence records of *Boltenia ovifera* catch recorded from DFO multispecies trawl surveys conducted between 2007 and 2015 and scallop dredge stock assessment surveys conducted in 1997 and 2007 in the Maritimes Region.

Table 7. Number of presence and absence records of *Boltenia ovifera* catch in the Maritimes Region from DFO scallop dredge surveys in 1997 and 2007, and DFO multispecies trawl surveys from 2007 to 2015.

Survey	Year	Total number of presences	Total number of absences
DFO scallop dredge survey	1997	22	119
	2007	2	106
DFO multispecies trawl survey	2007	41	265
	2008	68	222
	2009	33	236
	2010	54	284
	2011	47	266
	2012	34	232
	2013	49	303
	2014	28	249
	2015	7	26
	Total:	**385**	**2308**

Random Forest Model Results

The random forest model on the presence-absence *Boltenia ovifera* data had a very good performance rating, with an average AUC of 0.868 ± 0.031 SD (Table 8). Class error of the presence and absence classes was similar. Sensitivity and specificity of this model were good (0.787 and 0.795, respectively).

The predicted presence probability surface of *B. ovifera* from the model is shown in Figure 20. The area off northern Cape Breton and Misaine Bank were predicted to have a high probability of occurrence of *B. ovifera*. Smaller pockets of high presence probability also occurred on Georges Bank and in LaHave Basin. Areas of elevated presence probability corresponded to the occurrence of presence data points (Figure 21). This model appears to under-predict data points that occur in lower densities. Only small pockets of extrapolated area occurred on the Scotian Shelf and did not overlap with high presence probability areas.

Figure 22 depicts the classification of *B. ovifera* presence probability into discrete presence and absence classes based on the prevalence threshold of 0.14. In this map, all presence probability values greater than 0.14 were classified as presence, while values less than 0.14 were classified as absence. The area along the coast and much of the eastern portion of the study extent was classified as suitable habitat for *B. ovifera*. Bay of Fundy, Gulf of Maine, and the central Scotian Shelf were predicted as unsuitable habitat for this species.

Surface Temperature Mean was the most important variable for the classification of the *B. ovifera* presence-absence data (see Figure 23). This was followed by Bottom Temperature Mean and other surface and bottom temperature variables. Bottom Temperature Mean was the most influential variable in predicting the distribution of *B. ovifera* in the Gulf Region (Murillo et al., 2016). The partial dependence plots of the top 6 environmental variables are shown in Figure 24. Presence probability of *B. ovifera* was highest at Surface Temperature Mean values of less than 8˚C, and Bottom Temperature Mean values of less than 4˚C. A similar pattern was observed along the gradient in Bottom Temperature Mean in the Gulf Region, with high presence probability occurring at temperatures below 4˚C (Murillo et al., 2016).

Table 8. Accuracy measures and confusion matrix from 10-fold cross validation of a random forest model of presence and absence of *Boltenia ovifera* within the Maritimes Region. Observ. = Observations; Sensit. = Sensitivity, Specif. = Specificity.

Model Fold	**AUC**	**Observ.**	**Predictions**		**Total n**	**Class error**	**Sensit.**	**Specif.**
1	0.863		**Absence**	**Presence**				
2	0.808	**Absence**	1836	472	2308	0.205	0.787	0.795
3	0.886	**Presence**	82	303	385	0.213		
4	0.921							
5	0.888							
6	0.880							
7	0.842							
8	0.846							
9	0.871							
10	0.880							
Mean	**0.868**							
SD	**0.031**							

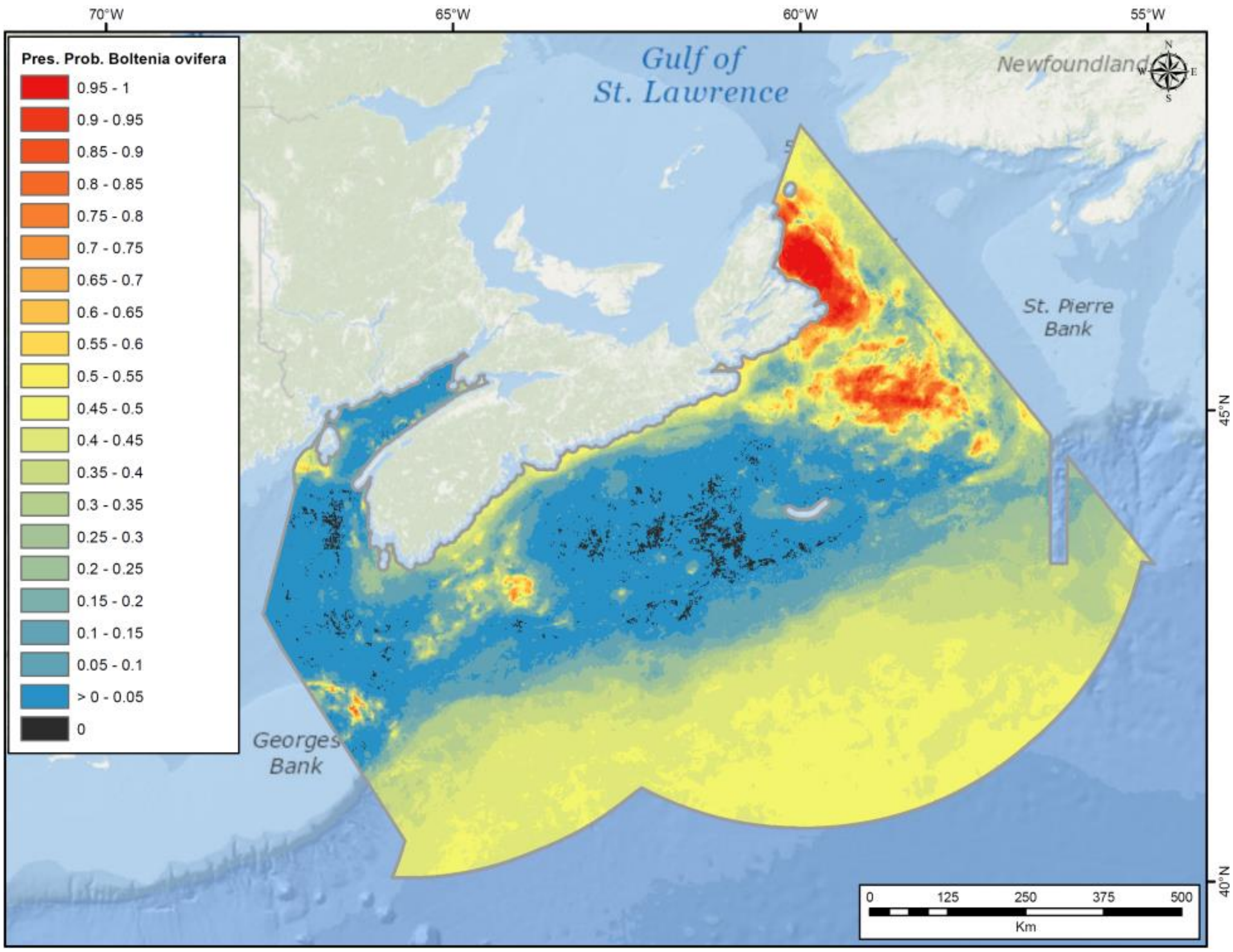

Figure 20. Predictions of presence probability (Pres. Prob.) of *Boltenia ovifera* based on a random forest model on presence and absence *B. ovifera* catch data collected from DFO multispecies trawl surveys and DFO scallop dredge surveys conducted in the Maritimes Region between 1997 and 2015.

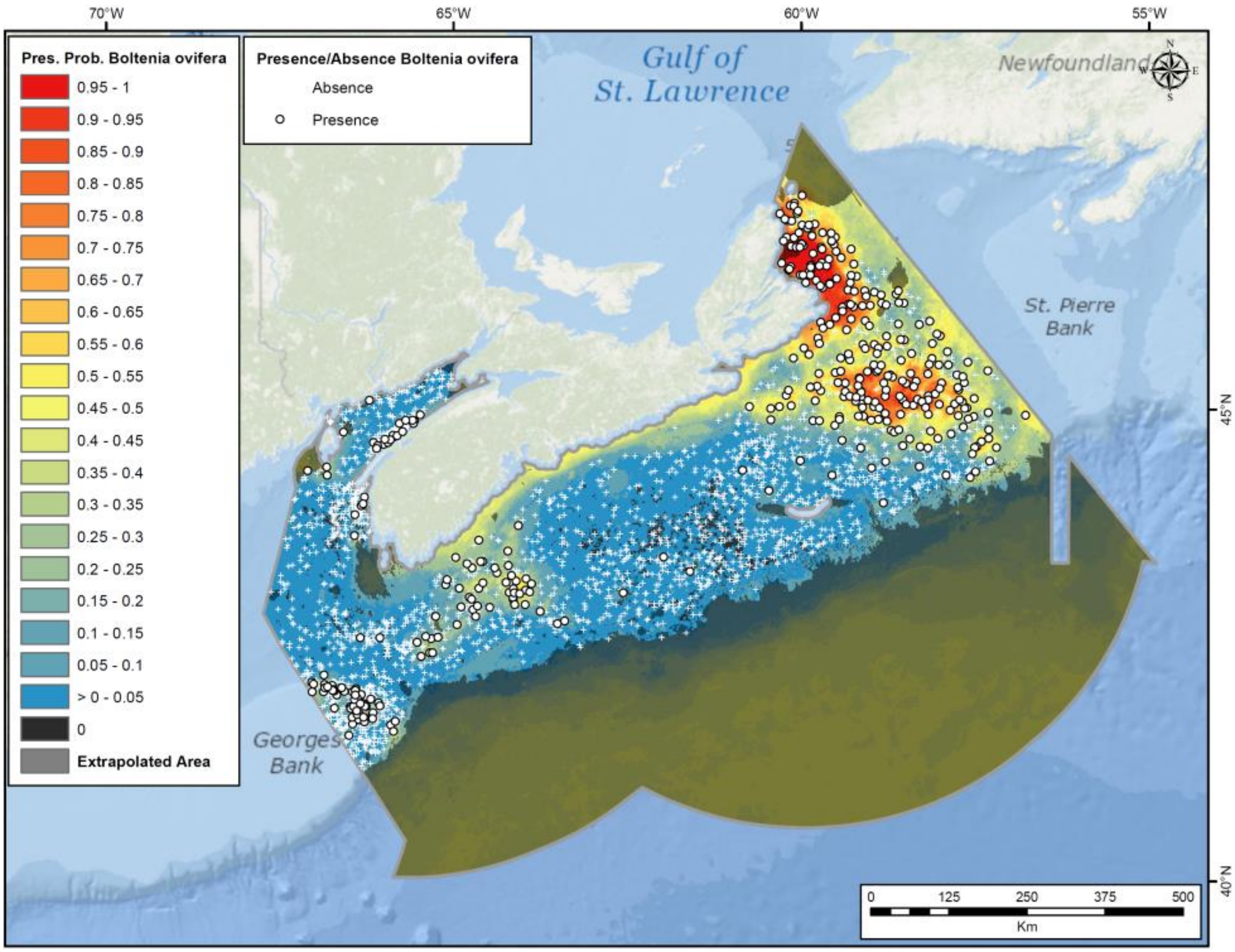

Figure 21. Presence and absence observations and predictions of presence probability (Pres. Prob.) of *Boltenia ovifera* based on a random forest model on presence and absence *B. ovifera* catch data collected from DFO multispecies trawl surveys and DFO scallop dredge surveys conducted in the Maritimes Region between 1997 and 2015. Also shown are the areas of model extrapolation.

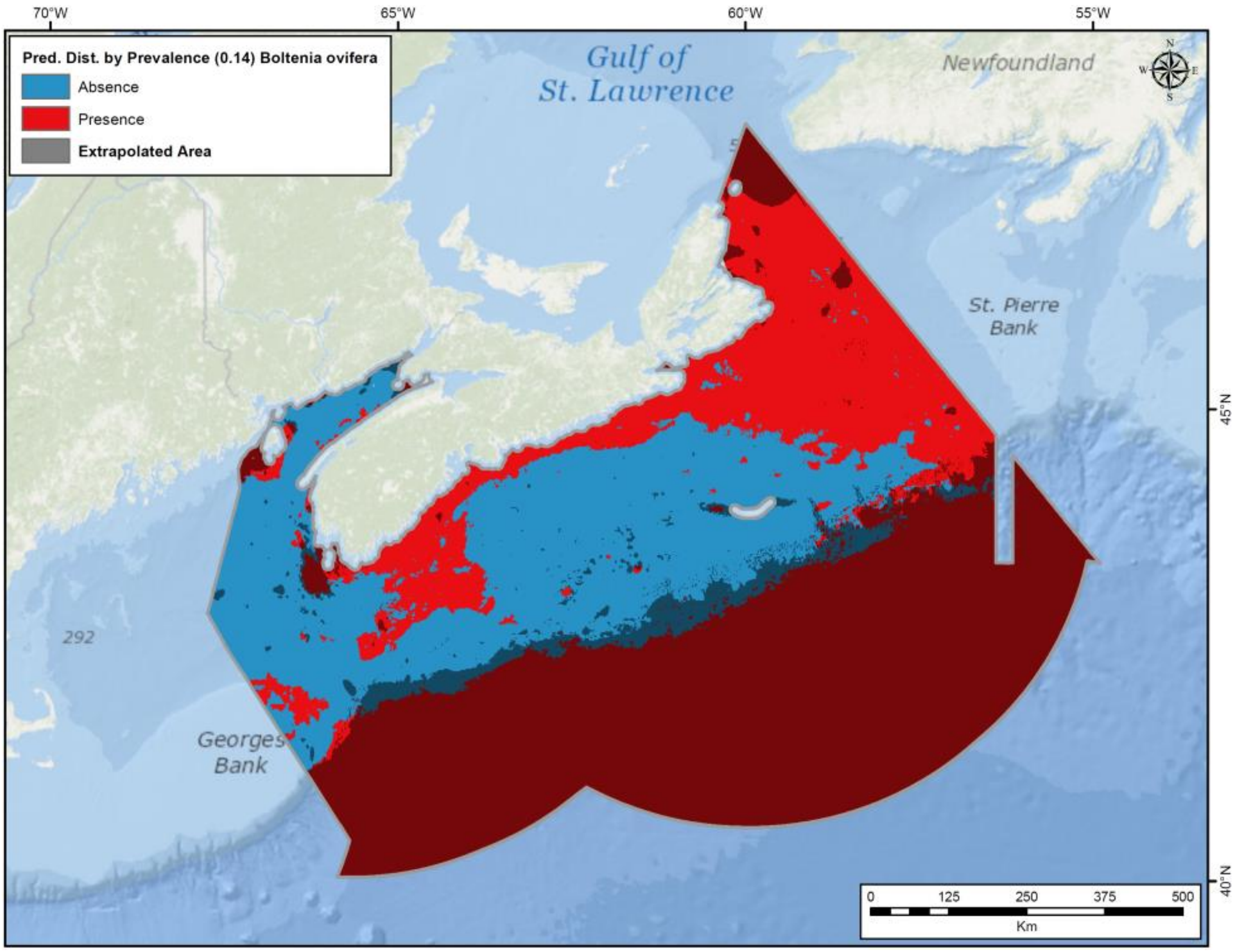

Figure 22. Predicted distribution (Pred. Dist.) of *Boltenia ovifera* in the Maritimes Region based on the prevalence threshold of 0.14 of *B. ovifera* presence and absence data used in the random forest model. Also shown are the areas of model extrapolation (grey polygon may appear red or blue when overlain on the prevalence surface).

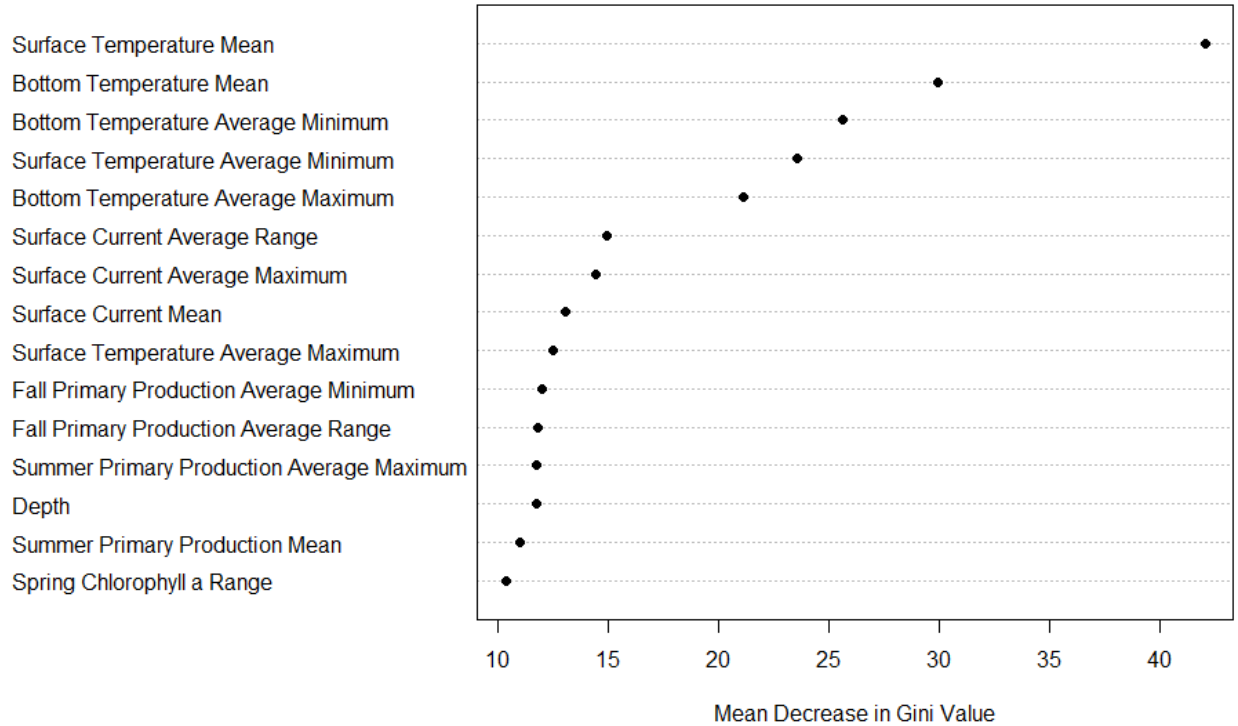

Figure 23. Importance of the top 15 predictor variables measured as the Mean Decrease in Gini value of the random forest model on *Boltenia ovifera* presence and absence data within the Maritimes Region. The higher the Mean Gini value the more important the variable is for predicting the response data.

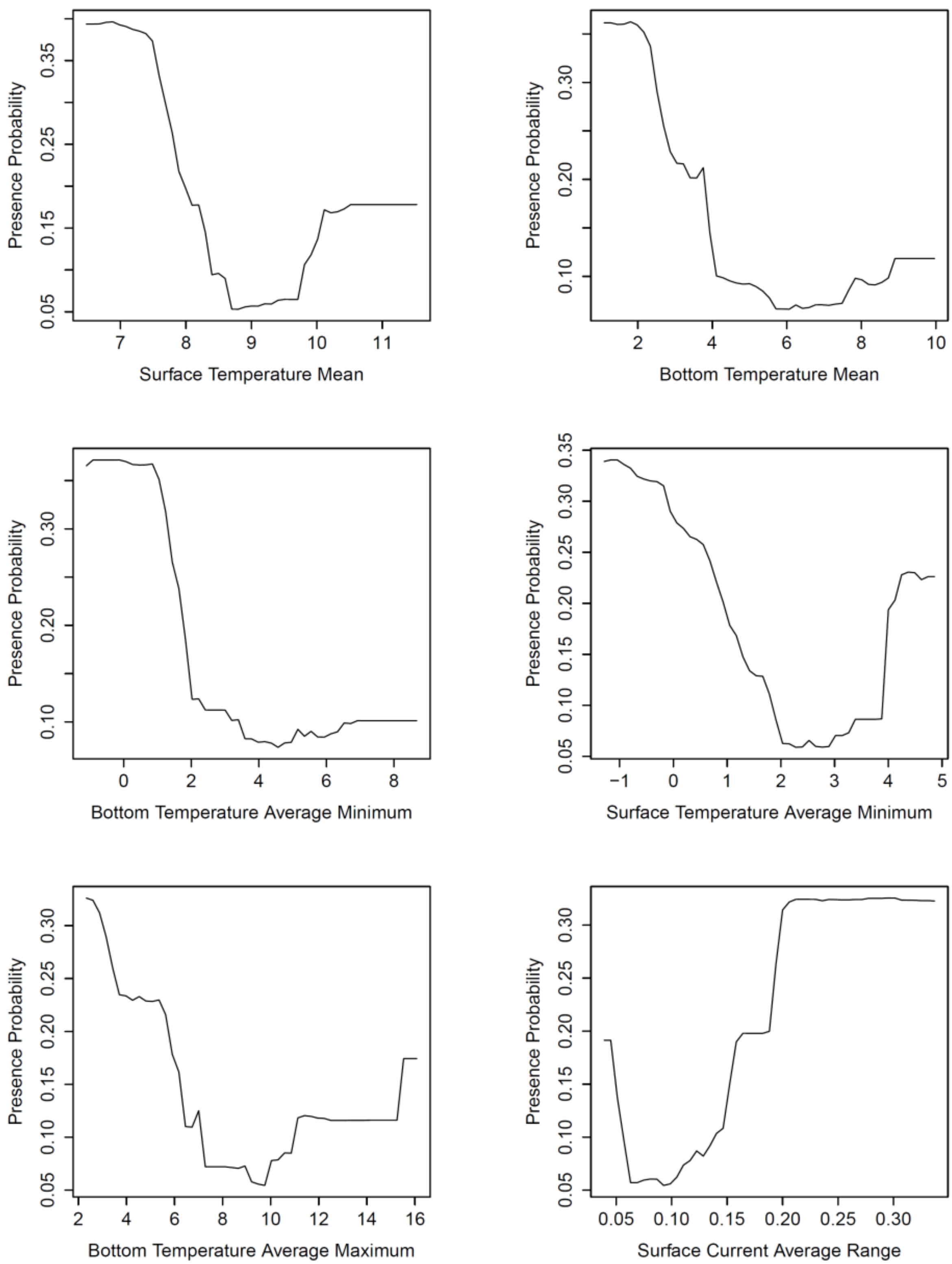

Figure 24. Partial dependence plots of the top six predictors from the random forest model of *Boltenia ovifera* presence and absence data collected within the Maritimes Region, ordered left to right from the top. Presence probability is shown on the *y*-axis.

Validation of Selected Model Using Independent Data

No additional records of *Boltenia ovifera* were available for the Maritimes Region, therefore this model could not be validated at the time this report was produced.

Prediction of *Boltenia ovifera* Biomass using Random Forest

The accuracy measures of the regression random forest model on mean *Boltenia ovifera* biomass per grid cell from DFO multispecies trawl surveys are presented in Table 9. The highest R^2 value was 0.372, while the average was 0.142 ± 0.109 SD, indicating fair model performance. The average Normalized Root-Mean-Square Error (NRMSE) was 0.014 ± 0.019 SD. The high standard deviation indicates high variability between model folds. The average percent variance explained was -3.20% ± 1.64 SD.

Figures 25 and 26 show the predicted biomass surface of *B. ovifera*. The majority of the spatial extent was predicted to have a low (0 – 0.24 kg) *B. ovifera* biomass. Small pockets of high biomass were predicted to occur on the banks of the eastern Scotian Shelf. These areas coincided with the highest biomass records (see Figure 26). Overall the biomass of *B. ovifera* was under-predicted by the model, with a maximum predicted value of 31.14 kg, just over half the empirical maximum value in the raw data (51 kg).

Slope (a non-interpolated variable) was the top predictor of the *B. ovifera* biomass data (see Figure 27). This was followed distantly by Fall Primary Production Average Minimum. The partial dependence plots showed that predicted biomass was highest on slopes greater than 4° (Figure 28).

Table 9. Accuracy measures from 10-fold cross validation of a random forest model of average *Boltenia ovifera* biomass (kg) per grid cell recorded from DFO multispecies trawl surveys in the Maritimes Region. RMSE = Root-Mean-Square Error; NRMSE = Normalized Root-Mean-Square Error.

Model Fold	**R^2**	**RMSE**	**NRMSE**	**Percent (%) variance explained**
1	0.049	1.639	0.032	-1.87
2	0.016	3.254	0.064	-0.20
3	0.216	0.165	0.003	-3.54
4	0.100	0.229	0.004	-4.18
5	0.372	0.239	0.005	-4.55
6	0.103	0.390	0.008	-4.44
7	0.185	0.237	0.005	-4.76
8	0.217	0.212	0.004	-4.67
9	0.029	0.485	0.010	-1.19
10	0.128	0.319	0.006	-2.60
Mean	**0.142**	**0.717**	**0.014**	**-3.20**
SD	**0.109**	**0.992**	**0.019**	**1.64**

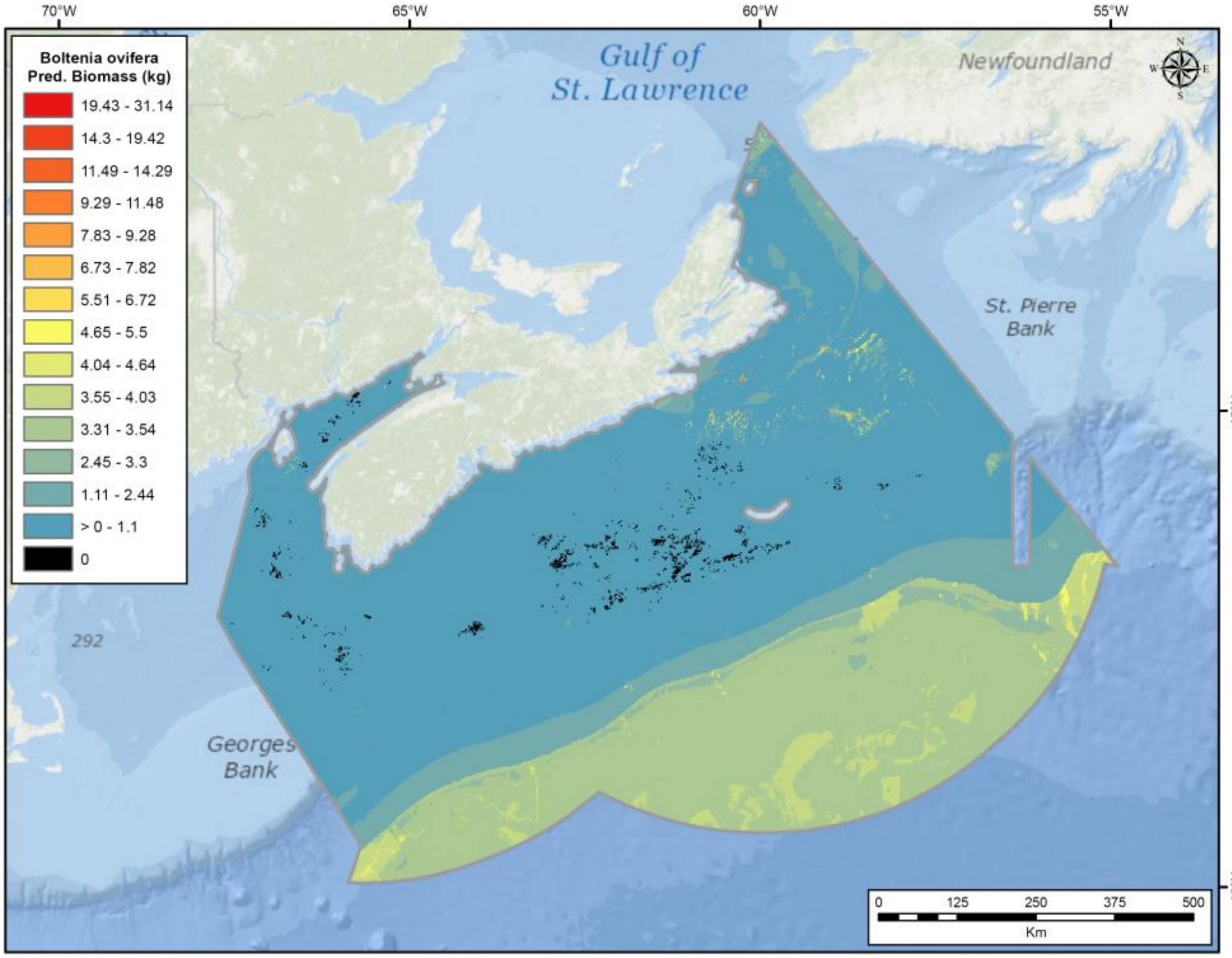

Figure 25. Predictions of biomass (kg) of *Boltenia ovifera* from catch data recorded in DFO multispecies trawl surveys in the Maritimes Region between 2007 and 2015.

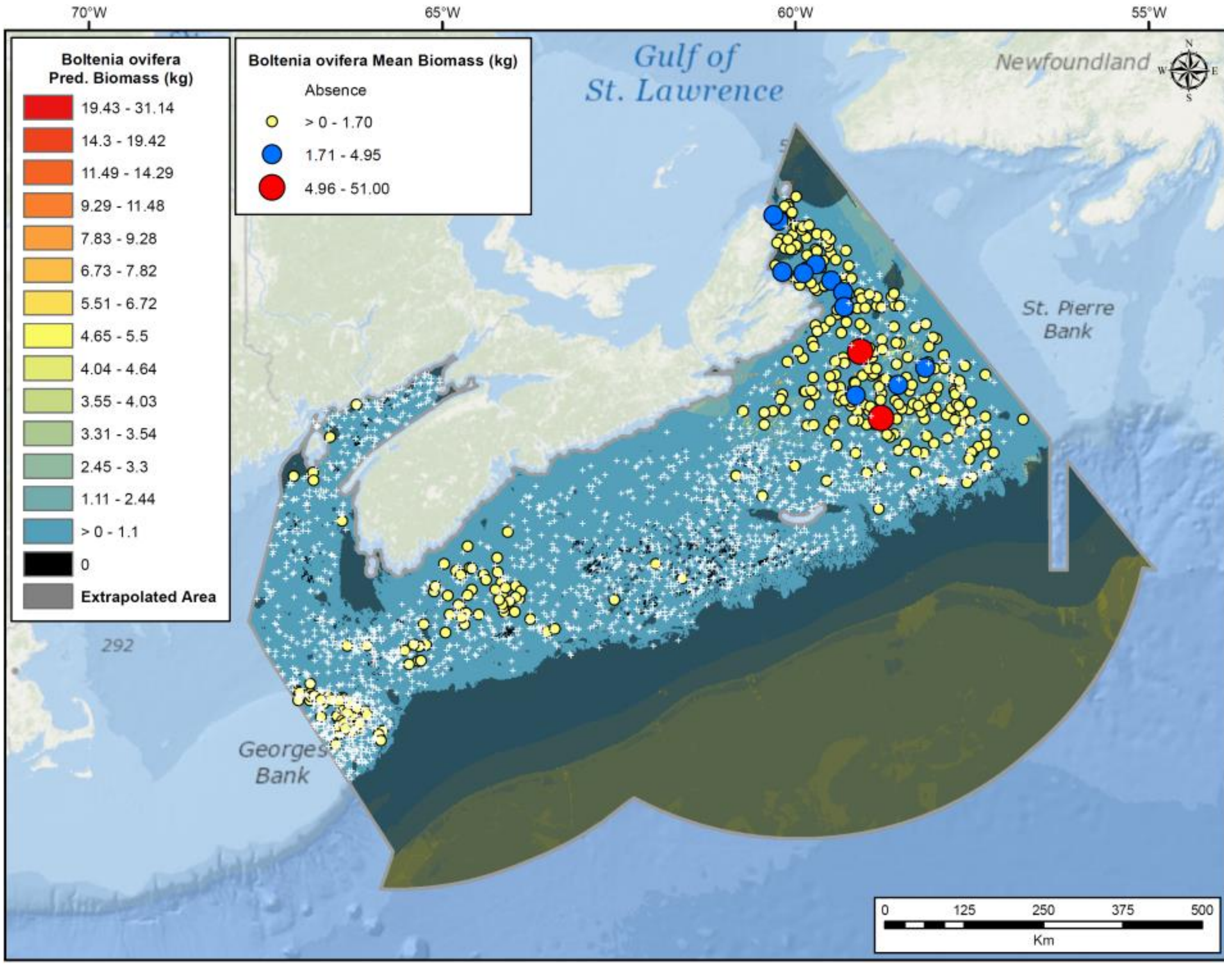

Figure 26. Predicted biomass surface of *Boltenia ovifera* based on DFO multispecies trawl survey records collected in the Maritimes Region between 2007 and 2015. Also shown are the mean biomass values per grid cell and areas of model extrapolation.

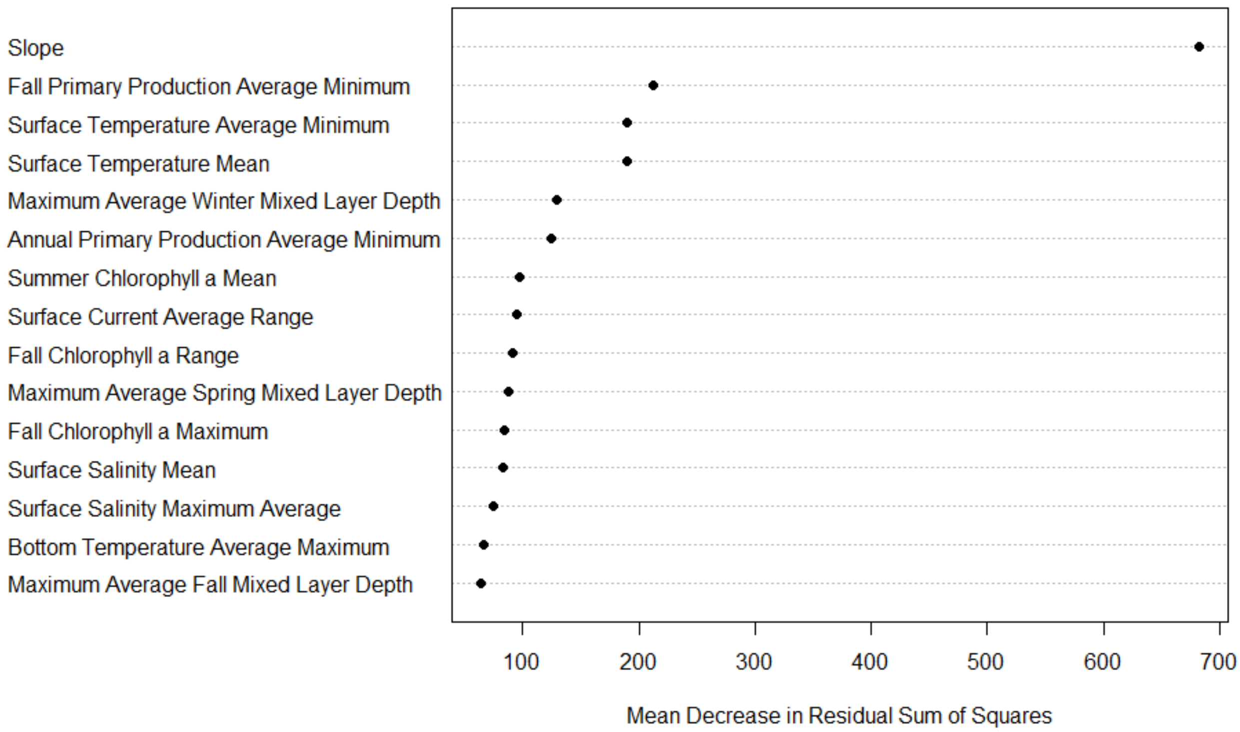

Figure 27. Importance of the top 15 predictor variables measured as the Mean Decrease in Residual Sum of Squares of the regression random forest model on *Boltenia ovifera* mean biomass data. The higher the Mean Decrease in Residual Sum of Squares the more important the variable is for predicting the response data.

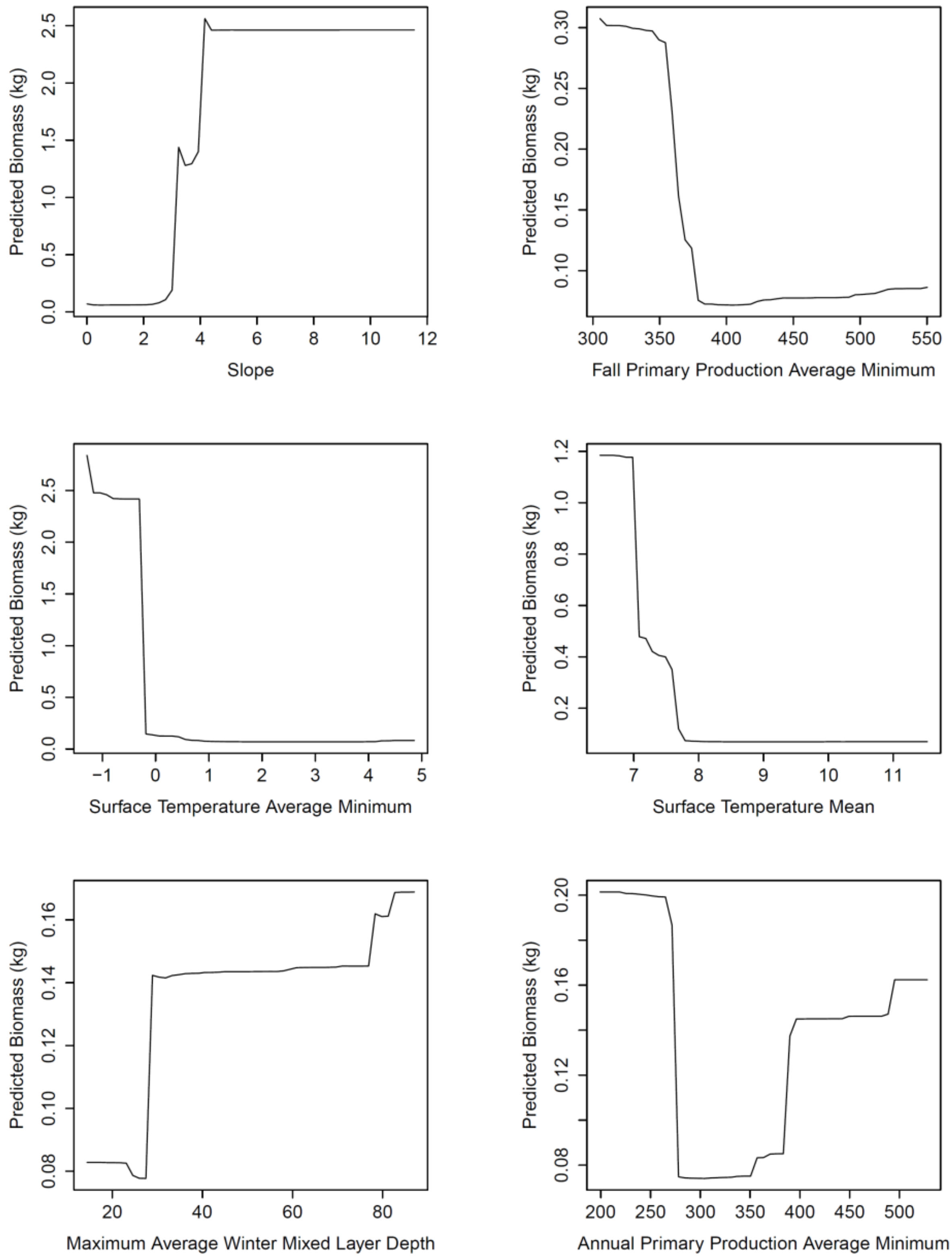

Figure 28. Partial dependence plots of the top six predictors from the random forest model of *Boltenia ovifera* biomass data collected within the Maritimes Region, ordered left to right from the top. Predicted biomass (kg) is shown on the *y*-axis.

Kernel Density Analysis of *Boltenia ovifera* Catch Data

Kernel density analysis of the *Boltenia ovifera* biomass data showed concentrations of these stalked tunicates on the eastern Scotian Shelf, with smaller densities in Bay of Fundy and on Browns and Georges Banks (Figure 29). The largest concentrations were observed on Banquereau Bank. The KDE search radius for this group was 20.5 km. Polygons encompassing each of 9 catch thresholds showed that the greatest increase in area between successive weights, after the initial mapping out of the habitat, occurred between 1 kg and 0.5 kg, where the area increased by 108% (Table 10, Figure 30). There were 24 data points contributing to this expansion and most of the increase was due to grouping of polygons on the outer shelf with few data supporting the groupings there (Table 10, Figure 31). Consequently 1 kg was considered as the threshold delineating *Boltenia ovifera* fields. These fields occupy 5,666 km^2 of seabed. However, it is likely that these stalked tunicates have a very patchy distribution within these areas. Similar KDE polygons for tunicates on the Tail of Grand Bank were not validated with on-bottom camera surveys in the NAFO Regulatory Area (NAFO, 2015). It was concluded that these species likely have patch sizes less than the survey trawl distance (< 1 km) and the individual tows above the 1 kg threshold might be more indicative of habitat location, although the density of observations is much higher in this region than on the Tail of Grand Bank, particularly in the area off Cape Breton. Figure 32 shows the final KDE significant area polygons, the location of significant and non-significant catches, and null records of *B. ovifera* from the DFO multispecies trawl surveys.

All KDE polygons occurred in areas predicted as suitable habitat (i.e. presence of *B. ovifera*) when overlain on the predicted distribution by prevalence surface (Figure 33). The area of suitable habitat was much larger in spatial extent than that covered by the KDE polygons, and the KDE polygons did not occur along the edge of suitable habitat. However, they did overlap with small pockets of model extrapolation. The KDE polygons were not compared to the RF biomass results for this species due to the poor performance of that model.

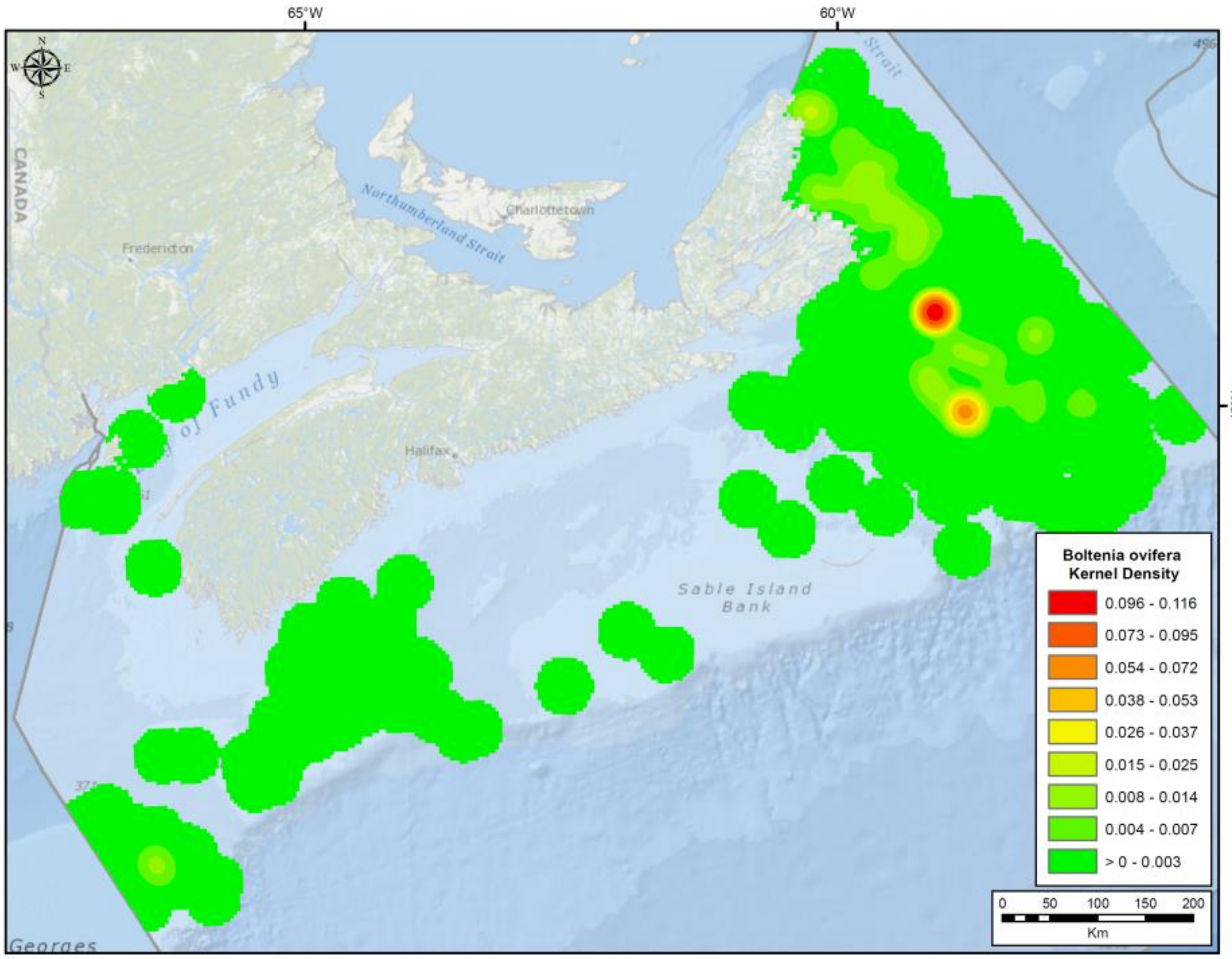

Figure 29. Kernel density analysis of *Boltenia ovifera* catches in the Maritimes Region.

Table 10. Summary characteristics for nine *Boltenia ovifera* catch weight thresholds. The threshold corresponding to the habitat delineation is indicated by shading.

Number of Points in Polygon	Area (km^2)	Catch Threshold (kg)	Percent Change in Area (%)
5	143	3	664.85
11	1090	2	419.71
31	5666	1	107.50
55	11756	.5	56.46
87	18394	.25	61.59
159	29722	.1	46.10
214	43424	.05	51.69
321	65869	.01	7.18
360	70595	.001	

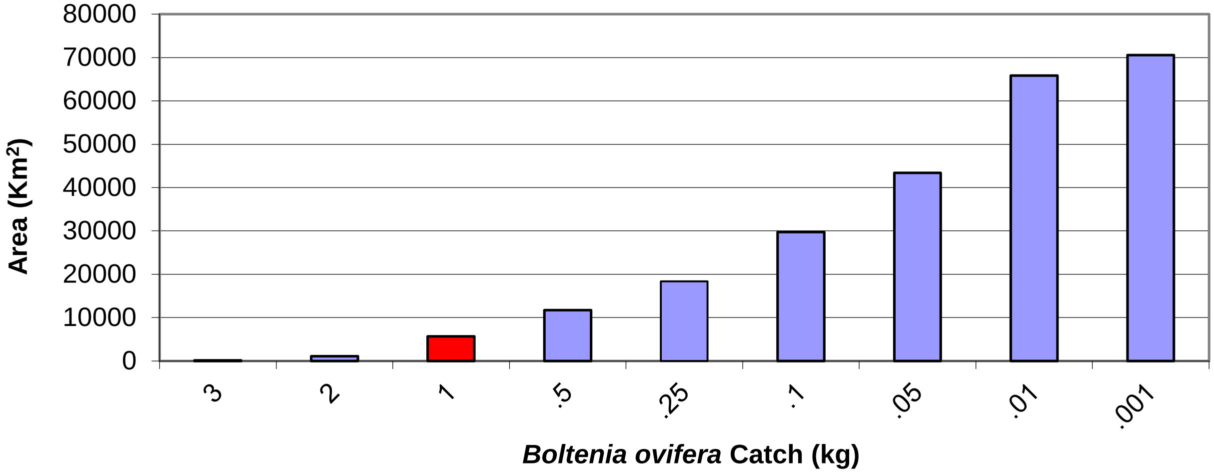

Figure 30. Area occupied by polygons encompassing catches of *Boltenia ovifera* greater than or equal to each of 9 catch thresholds (3 kg to 0.001 kg). The red bar indicates the catch threshold where the percent area increased the greatest between adjacent bins after the initial delineation of the tunicate fields.

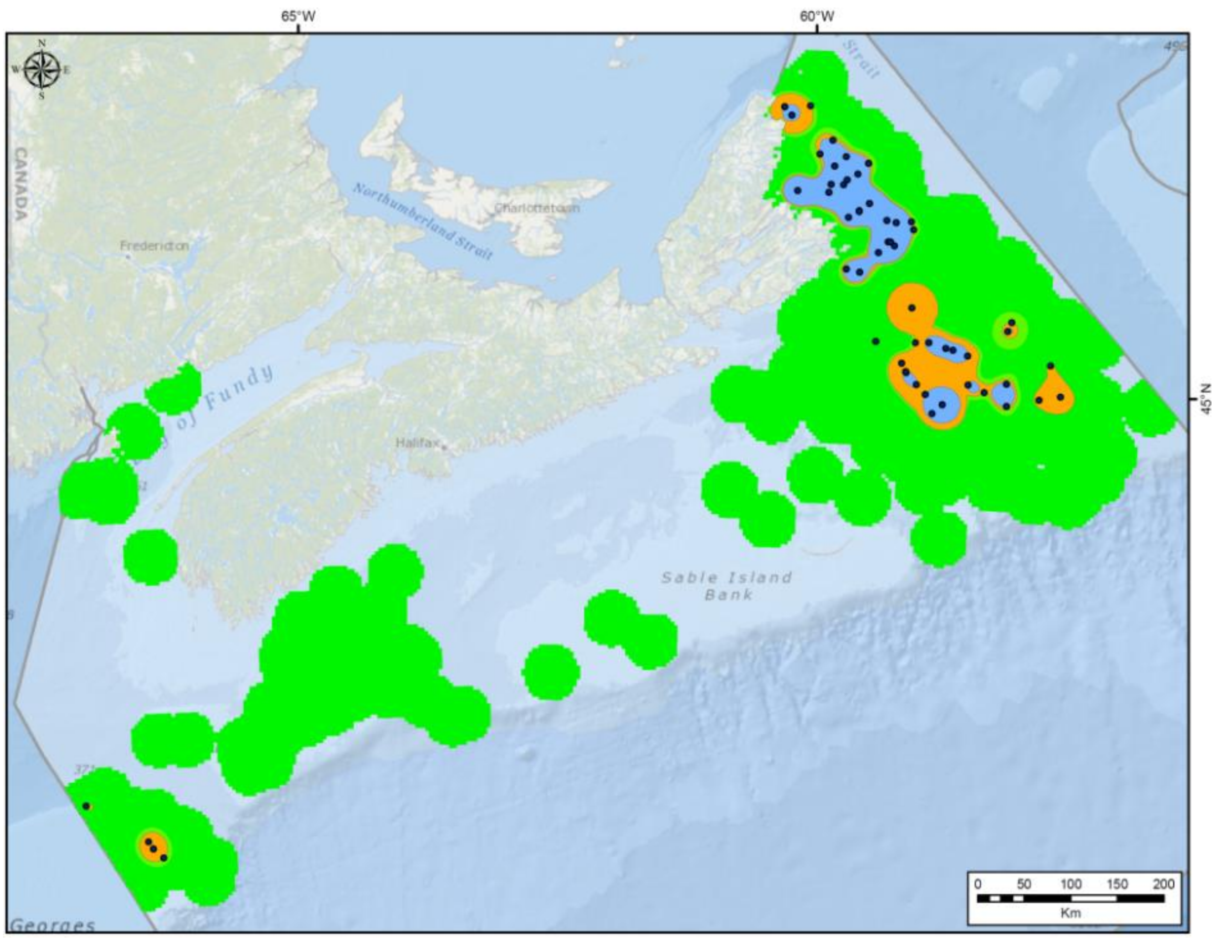

Figure 31. Polygons encompassing *Boltenia ovifera* catch weights greater than or equal to 1 kg (blue) and 0.5 kg (orange) overlain on the kernel density surface. Tow locations associated with biomass values above these weight thresholds are also shown. The 1 kg threshold was used to delineate the habitat.

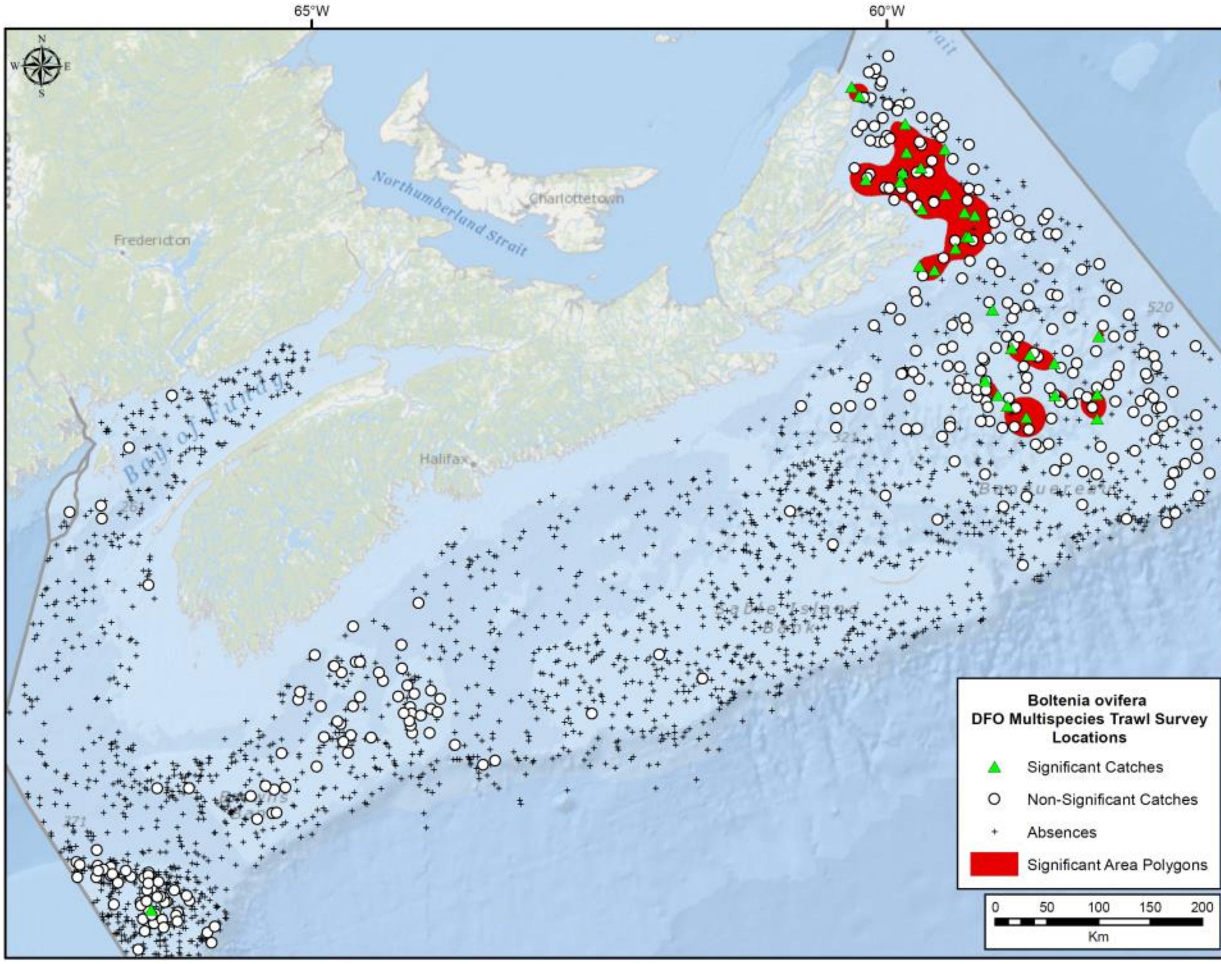

Figure 32. Location of polygons identifying significant *Boltenia ovifera* concentrations in the Maritimes Region. Also shown are the significant and non-significant catches of *B. ovifera* and absence data (nulls) from the DFO multispecies trawl surveys.

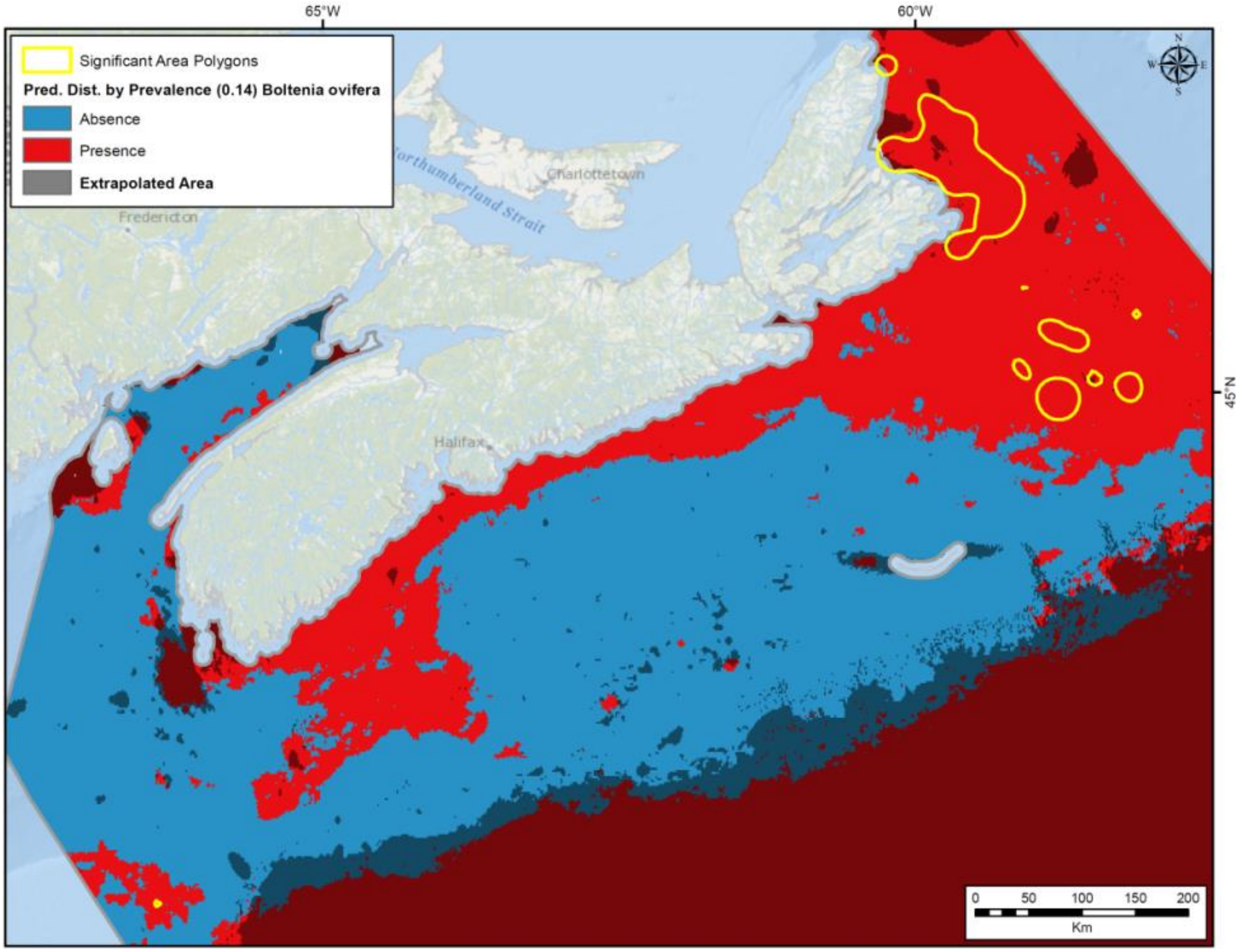

Figure 33. KDE polygons denoting significant concentrations of *Boltenia ovifera* overlain on the predicted distribution by prevalence surface of *B. ovifera* generated from the random forest model.

Sand Dollar Beds

Data Sources and Distribution

Sand dollars live as epifauna and shallow infauna, often completely buried in the upper sediment layer. Sand dollar bioturbation is considered the second most important modifier of the surficial sediments of Sable Island Bank after current activity (Stanley and James, 1971). A combined *in situ* camera and sampling survey on Sable Island Bank revealed that the sand dollar *Echinarachnius parma* can reach densities of 180 individuals per m^{-2} (Stanley and James, 1971). Figure 34 shows the distribution of available sand dollar records in the Maritimes Region. There is a clear east-west pattern in the distribution of sand dollars on the Scotian Shelf, with the majority of presences occurring on the shallow banks in the east, with some records in Bay of Fundy and on Georges Bank.

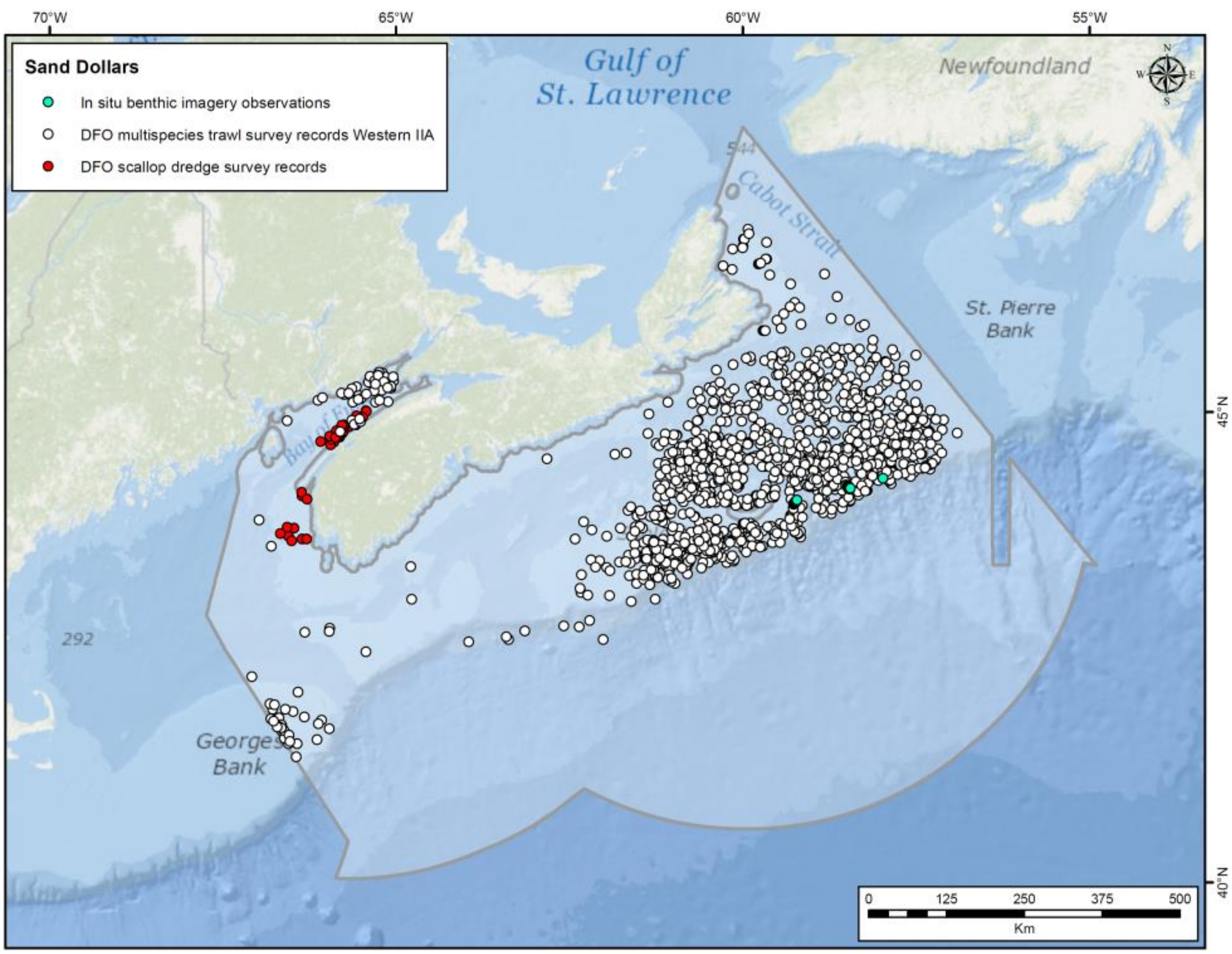

Figure 34. Available sand dollar presence data in the Maritimes Region from scientific surveys and DFO research vessel surveys.

In the DFO multispecies surveys, sand dollars are recorded under two taxa: the Order Clypeasteroida, and species *Echinarachnius parma*. The latter belongs to the Order Clypeasteroida and represents the only species of sand dollar recorded in the Maritimes Region. Thus, all specimens identified to Order Clypeasteroida are likely *E. parma*. Given the low number of sand dollar records from the DFO multispecies survey in the Bay of Fundy and southwest Nova Scotia area where sand dollars are known to occur, this dataset was augmented with data from the DFO scallop stock assessment surveys conducted off Digby (Scallop Production Area (SPA) 4) and the Brier Island, Lurcher Shoal, and St. Mary's Bay area (SPA 3) to ensure that the environmental niche of the species was captured (see Figure 35). The combined dataset consists of 1360 presences and 3619 absences (see Table 11 for data separated by survey type and year) after the data were reduced to a single record per cell. *In situ* benthic survey data were not included in this model due to its limited spatial extent and spatial overlap with the DFO multispecies trawl survey records.

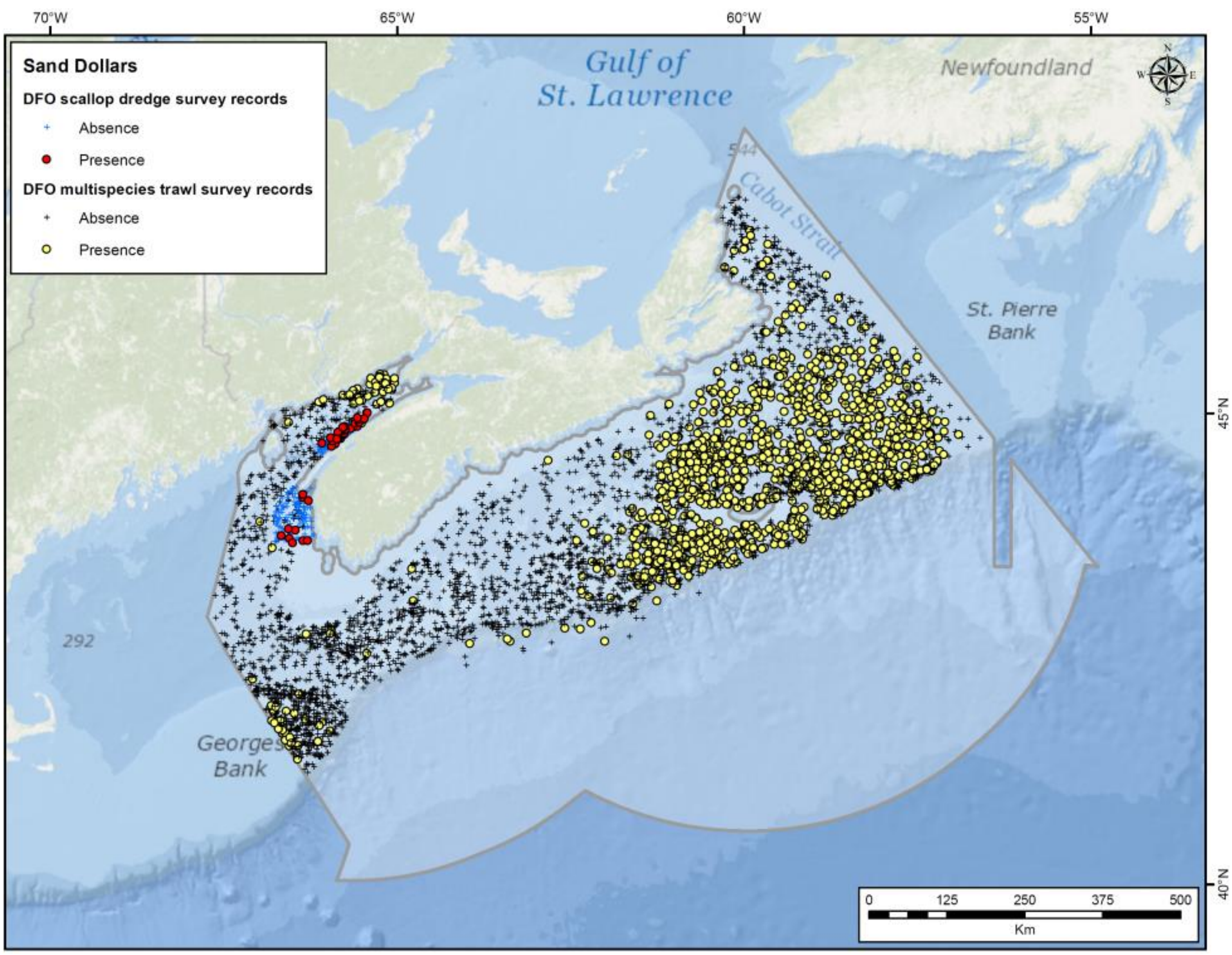

Figure 35. Presence and absence records of sand dollar catch recorded from DFO multispecies trawl surveys conducted between 1999 and 2015 and scallop dredge stock assessment surveys conducted in 1997 and 2007 in the Maritimes Region.

Given the relatively high prevalence of this taxon (0.27) we chose to run two models following the protocol outlined in Beazley et al. (2016a): one model with a balanced number of presences and absences (termed ‘Model 1’), and another model using all presence and absence records and a threshold equal to species prevalence (termed ‘Model 2’). The results of each model are shown below.

Table 11. Number of presence and absence records of sand dollar catch in the Maritimes Region from DFO scallop dredge surveys in 1997 and 2007, and DFO multispecies trawl surveys from 2000 to 2015. N/A indicates that the presence data from the survey in this year was outside the study extent, but absences are considered valid and used in the model.

Survey	**Year**	**Total number of presences**	**Total number of absences**
DFO scallop dredge survey	1997	15	121
	2007	26	83
DFO multispecies trawl survey	1999	12	89
	2000	113	217
	2001	67	225
	2002	119	205
	2003	121	197
	2004	21	87
	2005	168	340
	2006	114	277
	2007	78	213
	2008	73	207
	2009	75	185
	2010	104	225
	2011	70	233
	2012	58	198
	2013	78	260
	2014	48	224
	2015	N/A	33
	Total:	**1360**	**3619**

Model 1 – Balanced Species Prevalence

Accuracy measures for the random forest model on balanced species prevalence (1360 presences and 1360 absences; Model 1) are presented in Table 12. The highest mean AUC of 0.885 was associated with Model Run 6 and is therefore considered the optimal model for the prediction of the sand dollar response data. The sensitivity and specificity measures of this model were 0.833 and 0.807, respectively. The confusion matrix of the optimal model is also presented in Table 12. Class error was lower for the presence class than the absence class.

The presence probability prediction surface of sand dollars is presented in Figure 36. High predictions of presence probability occurred on Banquereau, Misaine, Middle, Sable Island and Western Banks. High presence probability was also predicted for the Bay of Fundy and in a small patch on Georges Bank. These areas corresponded well with the spatial distribution of presence records (see Figure 37), and areas of high presence probability do not appear to be grossly extrapolated beyond presence locations. Areas of extrapolation are also shown in Figure 37. All deep water beyond the Scotian Shelf was considered extrapolated area. Smaller areas of extrapolation also occurred off southwestern Nova Scotia and the northeast tip of Cape Breton where presence probability was moderate. In Bay of Fundy, an area predicted with high sand dollar presence probability was also considered an area of model extrapolation. Figure 38 shows the actual data observations (1360 presences and 1360 absences) used in Model 1. There appeared to be little or no spatial bias in the presence and absence records used in the model.

Bottom temperature and salinity variables ranked high in this model, with Bottom Temperature Average Minimum and Bottom Temperature Mean holding the top two positions (Figure 39). Depth was the third-ranking variable. Prior to spatial interpolation, Bottom Temperature Average Minimum displayed a slightly bimodal distribution (Beazley et al., in prep.). Examination of the Q-Q plot revealed a spatial pattern to data points over- and under-predicted by a normal distribution, with over-predicted points located in the Gulf of Maine, Emerald Basin, on Banquereau and Misaine Banks, and in the deepest portion of the study extent, and under-predicted points located in Bay of Fundy, off southwestern Nova Scotia, the Laurentian Channel and the area just beyond the shelf break. Bottom Temperature Average Minimum was followed more distantly in terms of importance by Bottom Temperature Mean, Depth, and the other variables in the Model. Partial dependence of sand dollars on the top six environmental variables is shown in Figure 40. Generally, presence probability was highest at lower bottom temperature values. Bottom temperature values in this range coincided mainly with those data points over-predicted by a normal distribution on the banks of the eastern Scotian Shelf. However, the fit between predicted and observed values of the kriging model was good for values in this range. Along the gradient in Depth, presence probability was highest at the shallowest depths (< 200 m), then decreased, and then increased and plateaued at 1000 m.

Table 12. Accuracy measures for all 10 model repetitions of 10-fold across validation of a random forest model of sand dollar presence-absence data collected within the Maritimes Region. The confusion matrix is shown for the model with the highest AUC value (Model Run 6) which is considered the optimal model for predicting the presence probability of sand dollars in the region.

Model Run	AUC	Sensitivity	Specificity
1	0.887	0.845	0.789
2	0.873	0.838	0.769
3	0.887	0.827	0.795
4	0.883	0.835	0.777
5	0.891	0.849	0.800
6	**0.894**	**0.833**	**0.807**
7	0.885	0.838	0.785
8	0.882	0.839	0.779
9	0.881	0.828	0.771
10	0.888	0.850	0.782
Mean	**0.885**	**0.838**	**0.785**
SD	**0.006**	**0.008**	**0.013**

Confusion matrix of model with highest AUC:

Observations	Predictions		Total n	Class error
	Absence	**Presence**		
Absence	1097	263	1360	0.193
Presence	227	1133	1360	0.167

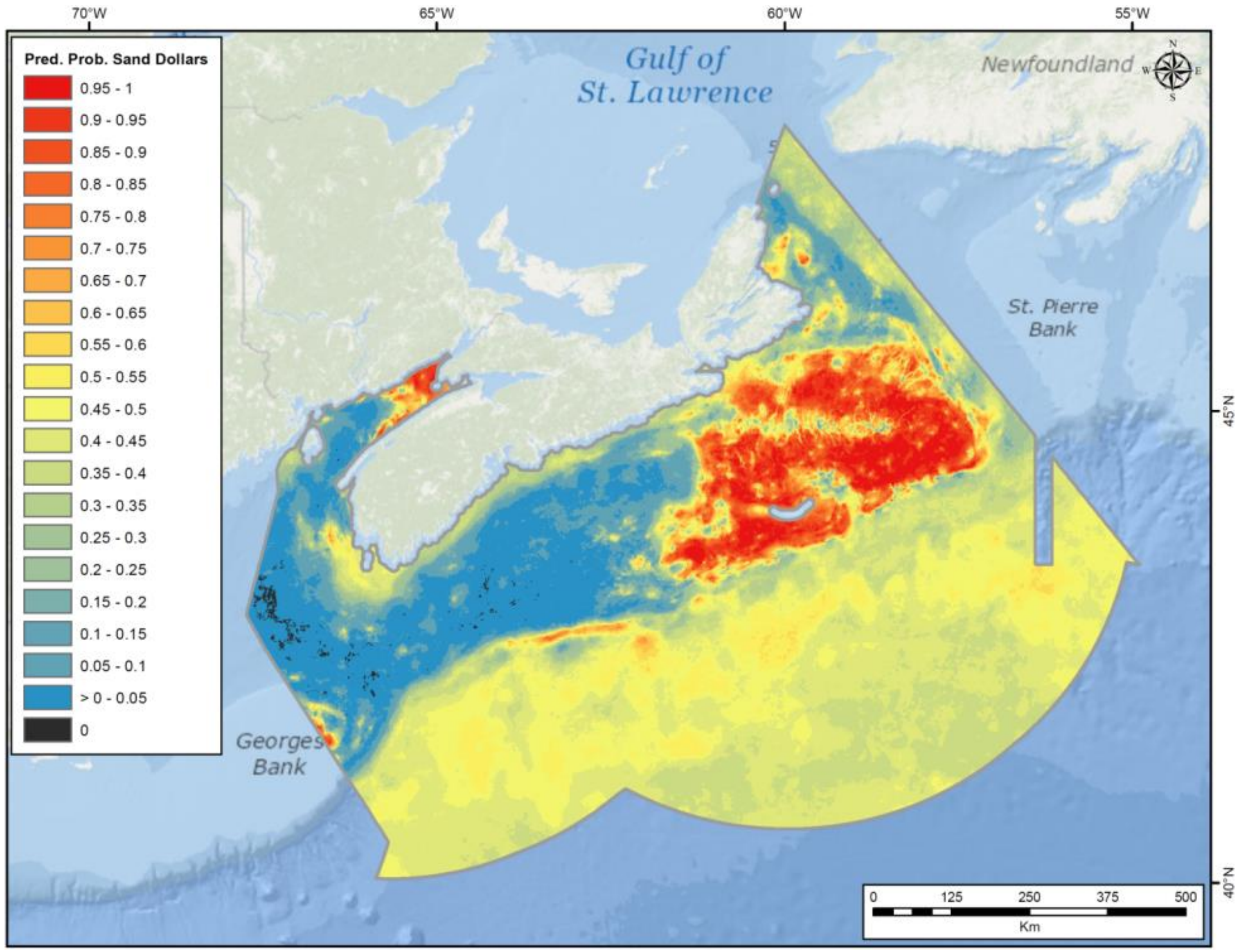

Figure 36. Predictions of presence probability (Pres. Prob.) from the optimal random forest model of sand dollar presence and absence data collected from DFO multispecies trawl surveys and DFO scallop dredge conducted in the Maritimes Region between 1997 and 2015.

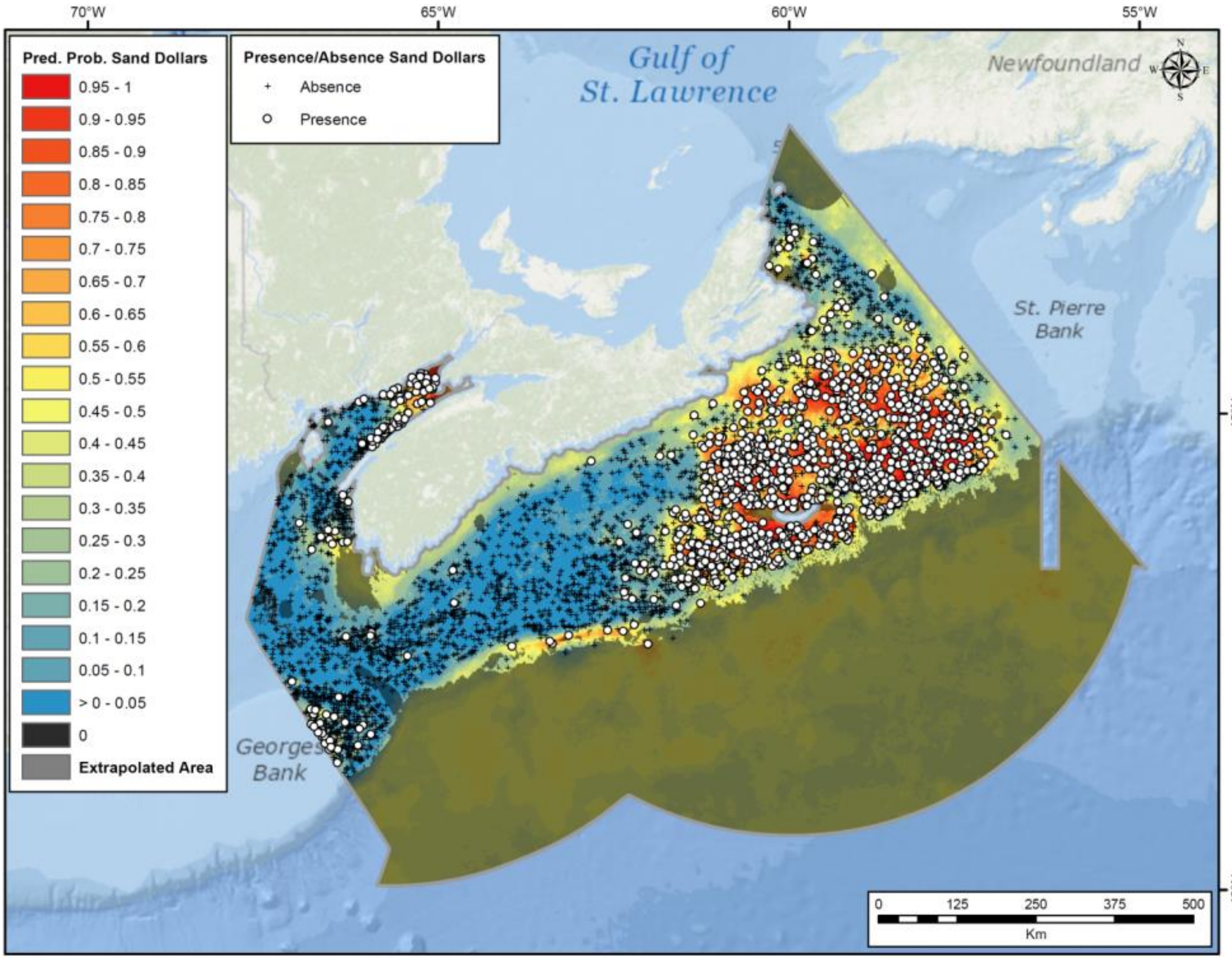

Figure 37. Presence and absence observations and predictions of presence probability (Pres. Prob.) from the optimal random forest model overlain with all sand dollar presence and absence data collected from DFO multispecies trawl surveys and DFO scallop dredge surveys conducted in the Maritimes Region between 1997 and 2015. Also shown are the areas of model extrapolation.

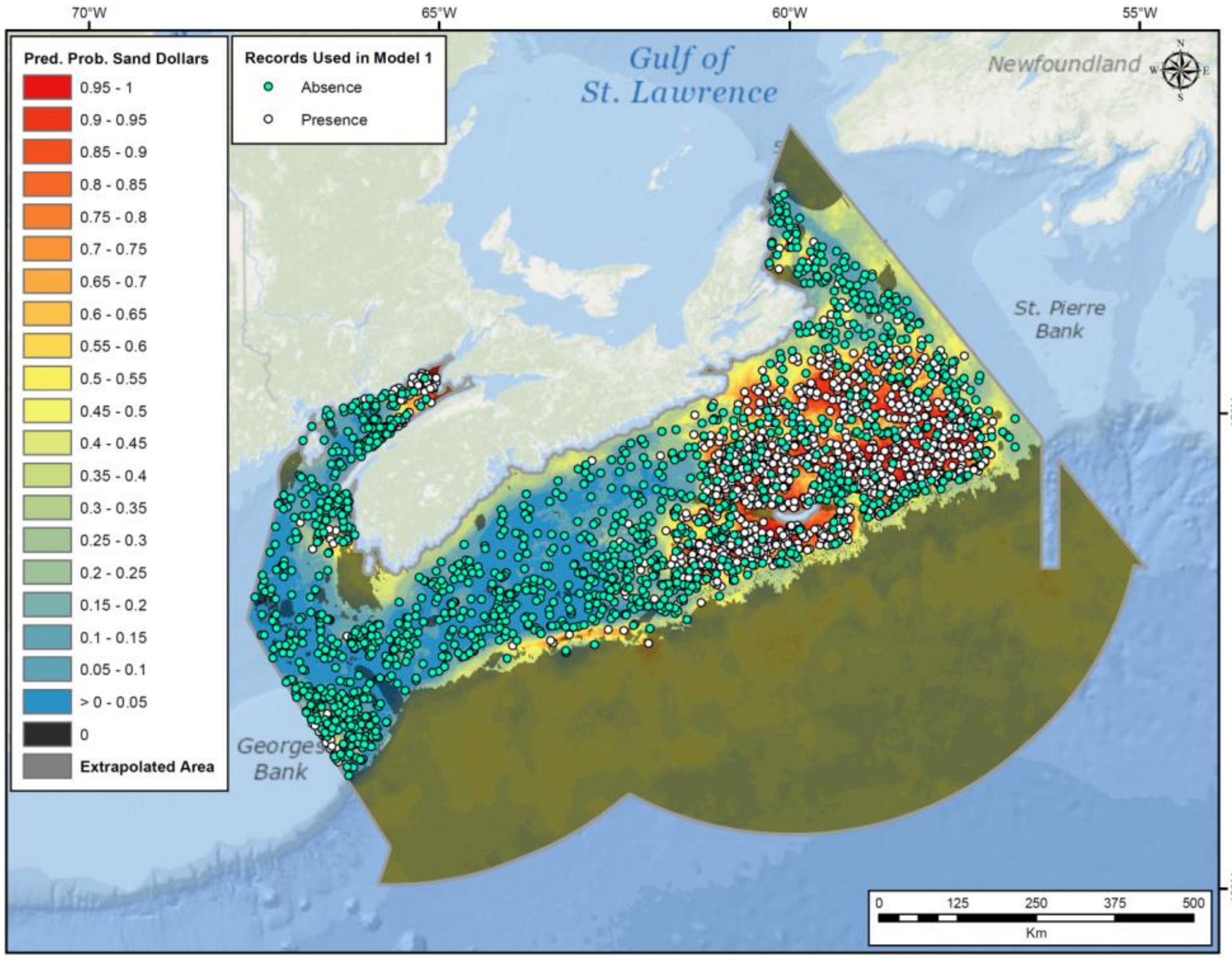

Figure 38. Map of the 2720 records (1360 presences and 1360 absences) of sand dollars used in the optimal random forest Model 1. Also shown in the predicted presence probability (Pres. Prob.) of sand dollars and areas of model extrapolation.

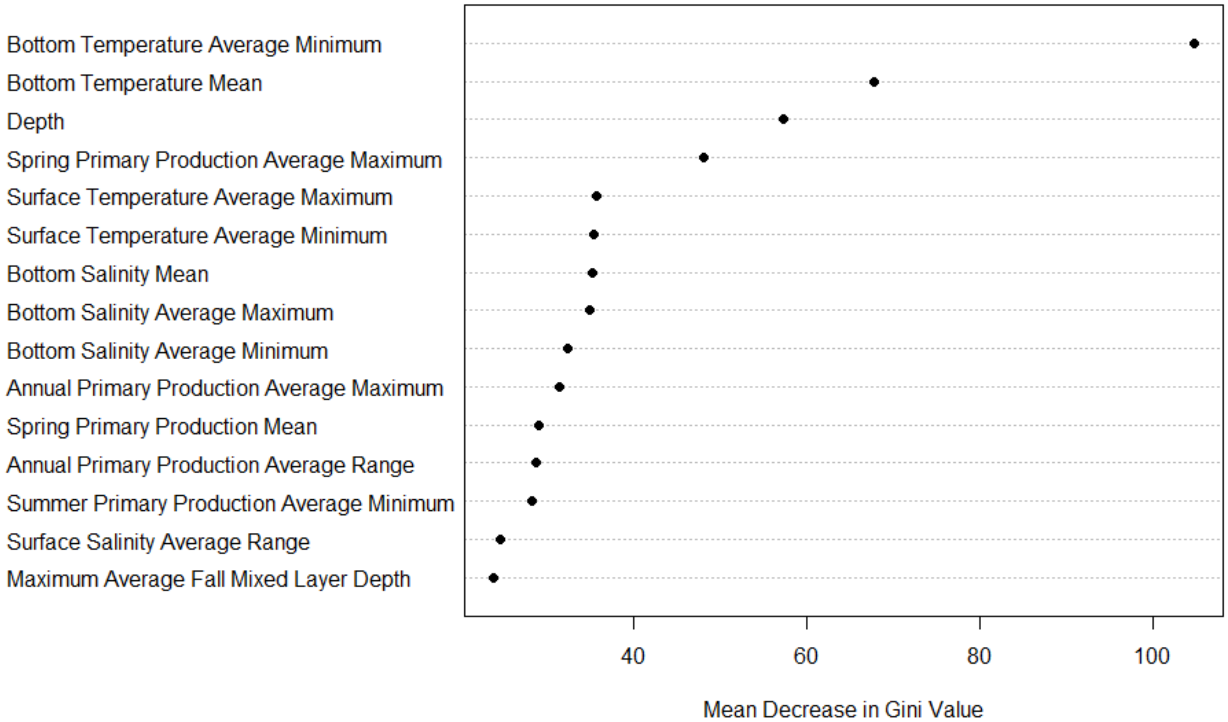

Figure 39. Importance of the top 15 predictor variables measured as the Mean Decrease in Gini value of the optimal random forest model predicting sand dollar presence and absence data within the Maritimes Region. The higher the Mean Decrease in Gini value the more important the variable is for predicting the response data.

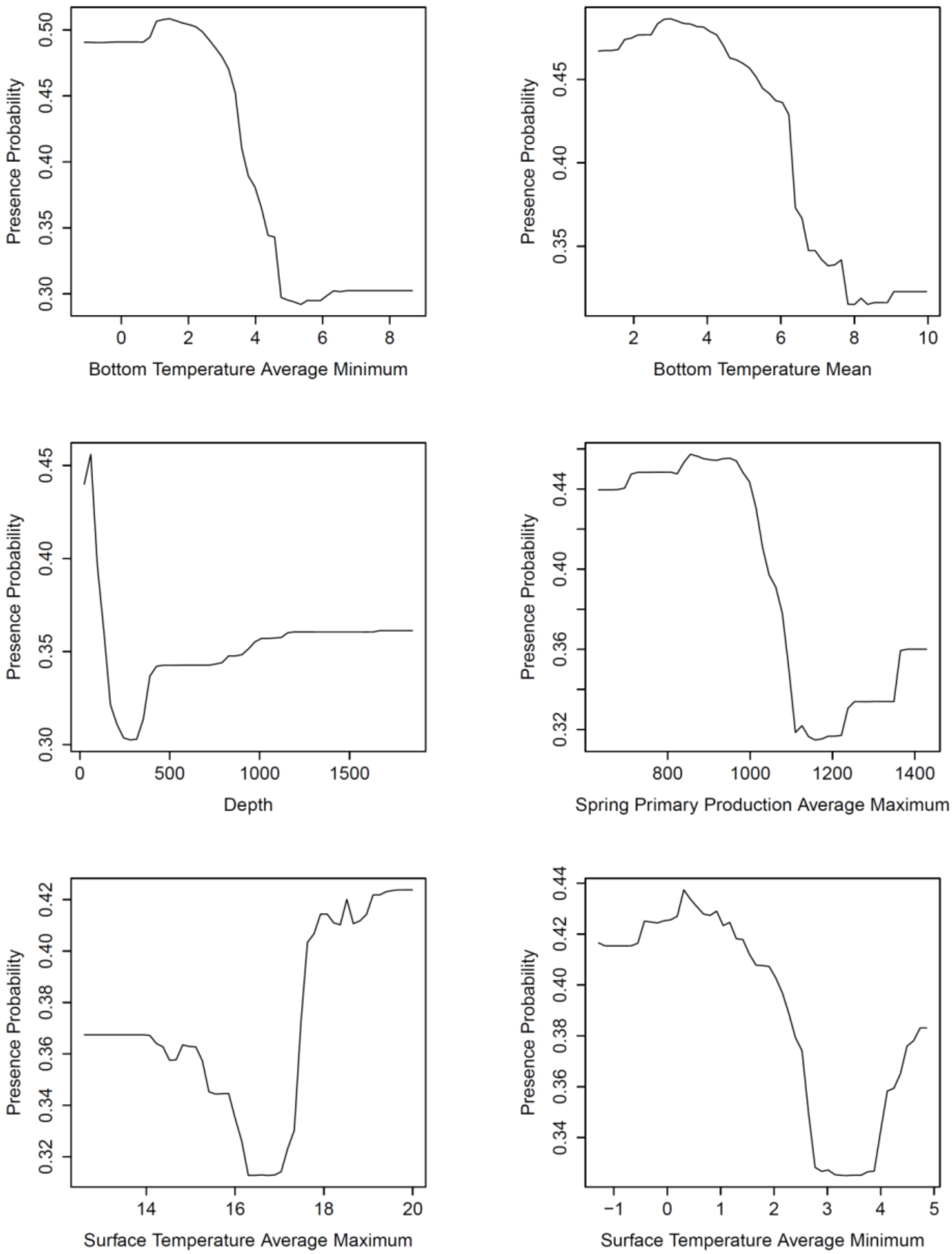

Figure 40. Partial dependence plots of the top six predictors from the optimal random forest model of sand dollar presence and absence data collected within the Maritimes Region, ordered left to right from the top. Presence probability is shown on the *y*-axis.

Model 2 – Unbalanced Data and Threshold Equal to Species Prevalence

Table 13 shows the accuracy measures for the random forest model on all sand dollar presence and absence data (1360 presences and 3619 absences; Model 2) and a threshold equal to species prevalence (0.27). The average AUC of this model was 0.886, slightly lower than the optimal run of Model 1 (AUC = 0.894). Class error of both the presence and absences, and sensitivity and specificity were similar to that of Model 1.

The predicted presence probability surface of sand dollars generated from Model 2 is shown in Figure 41. In this model the same areas were predicted with high presence probability as Model 1, but with slightly less intensity. Overall, the area predicted with low or zero presence probability expanded in Model 2, particularly on the central and western Scotian Shelf and Gulf of Maine. The model does not appear to predict areas of presence far beyond the location of presence records (see Figure 42), likely due to the inclusion of all absence records in the model. There is little difference in the areas considered extrapolated by Models 1 and 2, although the area of extrapolation near the shelf break is slightly reduced in Model 2. As in Model 1, the area in Bay of Fundy predicted with high sand dollar presence probability is also considered an area of model extrapolation.

Figure 43 depicts the classification of sand dollar presence probability into discrete classes of presence and absence based on the prevalence threshold of 0.27. In this map, all presence probability values generated from Model 2 that were greater than 0.27 were classified as presence, while values less than 0.27 were classed as absence. The majority of the eastern Scotian Shelf and some portions of the shelf break were predicted to be suitable habitat (i.e. presence) for sand dollars. The area southwest of Nova Scotia that is excluded from the trawl surveys due to hard bottom was also classified as presence of sand dollars.

The order of importance of the top environmental predictor variables was similar to that of Model 1, with Bottom Temperature Average Minimum, Depth, and Bottom Temperature Mean as the top variables (Figure 44). The partial dependence plots of the top six variables are shown in Figure 45. Like Model 1, presence probability was highest at Bottom Temperature Average Minimum values less than 2˚C and at the shallowest depths.

Table 13. Accuracy measures and confusion matrix from 10-fold cross validation of a random forest model of presence and absence of sand dollar within the Maritimes Region. Observ. = Observations; Sensit. = Sensitivity, Specif. = Specificity.

Model Fold	AUC	Observ.	Predictions		Total n	Class error	Sensit.	Specif.
1	0.916		**Absence**	**Presence**				
2	0.883	**Absence**	2834	785	3619	0.300	0.848	0.783
3	0.881	**Presence**	207	1153	1360	0.152		
4	0.892							
5	0.916							
6	0.866							
7	0.878							
8	0.883							
9	0.889							
10	0.860							
Mean	**0.886**							
SD	**0.018**							

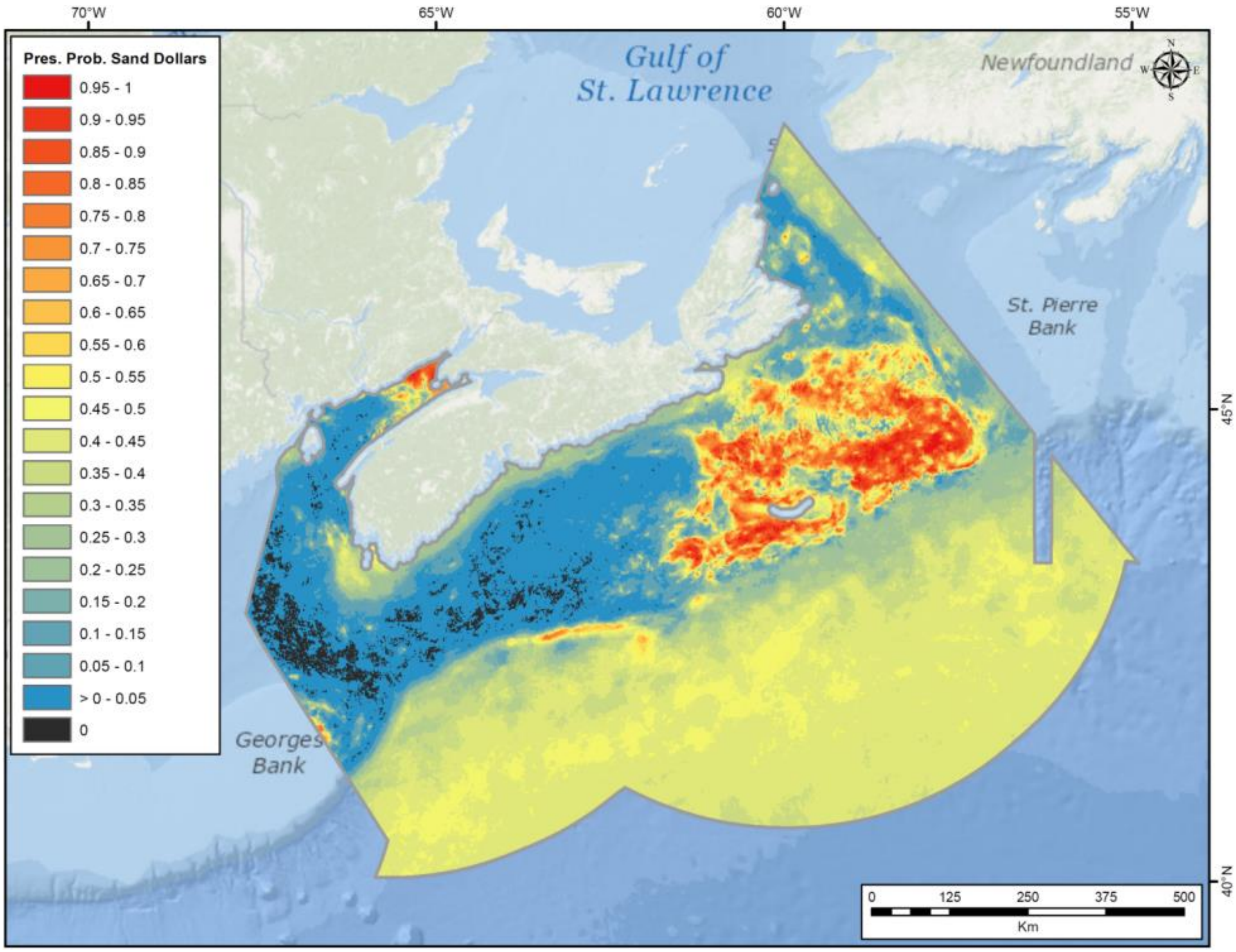

Figure 41. Predictions of presence probability (Pres. Prob.) of sand dollars based on a random forest model on unbalanced presence and absence sand dollar catch data collected from DFO collected from DFO multispecies trawl surveys and DFO scallop dredge surveys conducted in the Maritimes Region between 1997 and 2015.

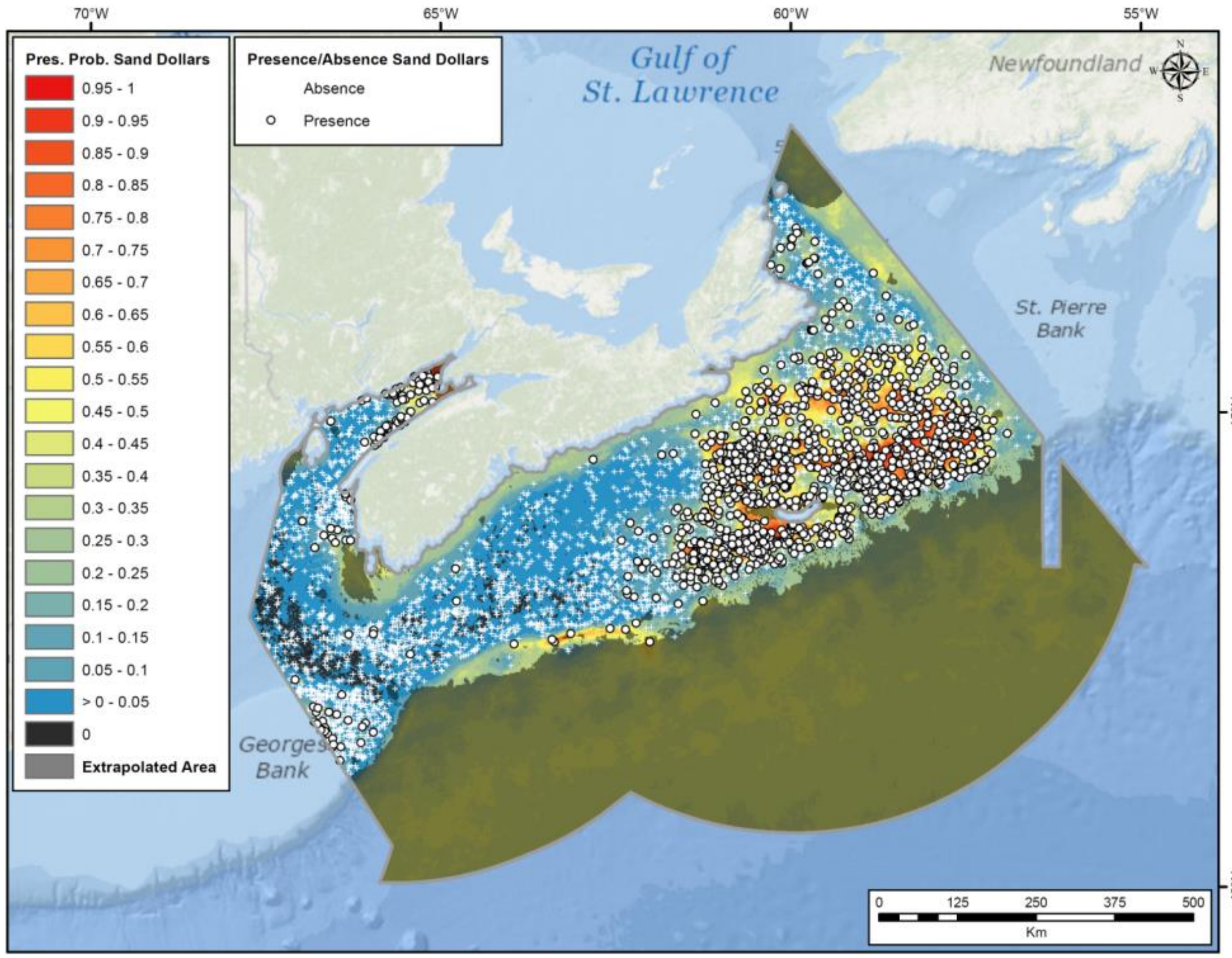

Figure 42. Presence and absence observations and predictions of presence probability (Pres. Prob.) of sand dollars based on a random forest model on unbalanced presence and absence sand dollar catch data collected from DFO multispecies trawl surveys and DFO scallop dredge surveys conducted in the Maritimes Region between 1997 and 2015. Also shown are the areas of model extrapolation.

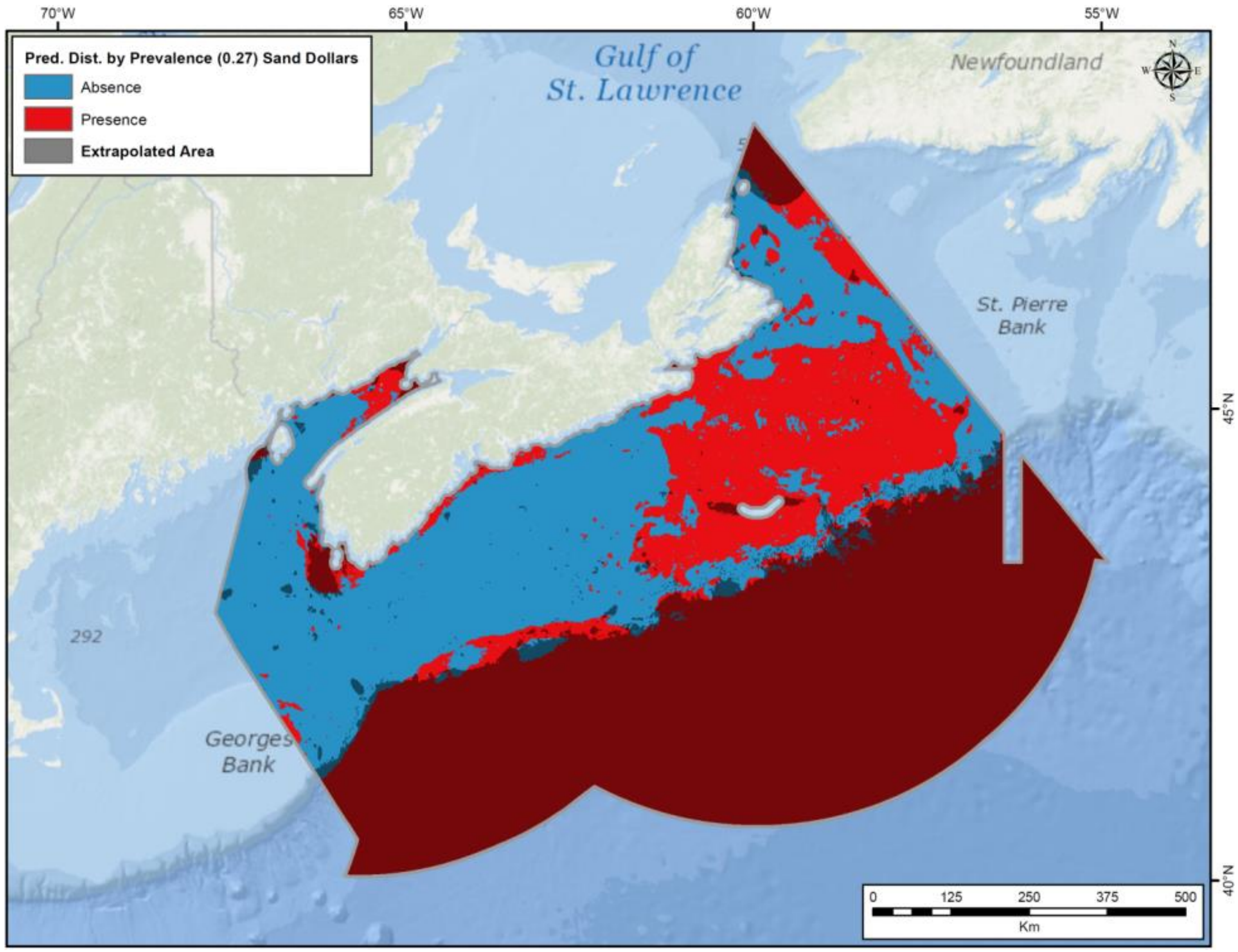

Figure 43. Predicted distribution (Pred. Dist.) of sand dollars in the Maritimes Region based on the prevalence threshold of 0.27 of sand dollar presence and absence data used in Model 2. Also shown are the areas of model extrapolation (grey polygon may appear red or blue when overlain on the prevalence surface).

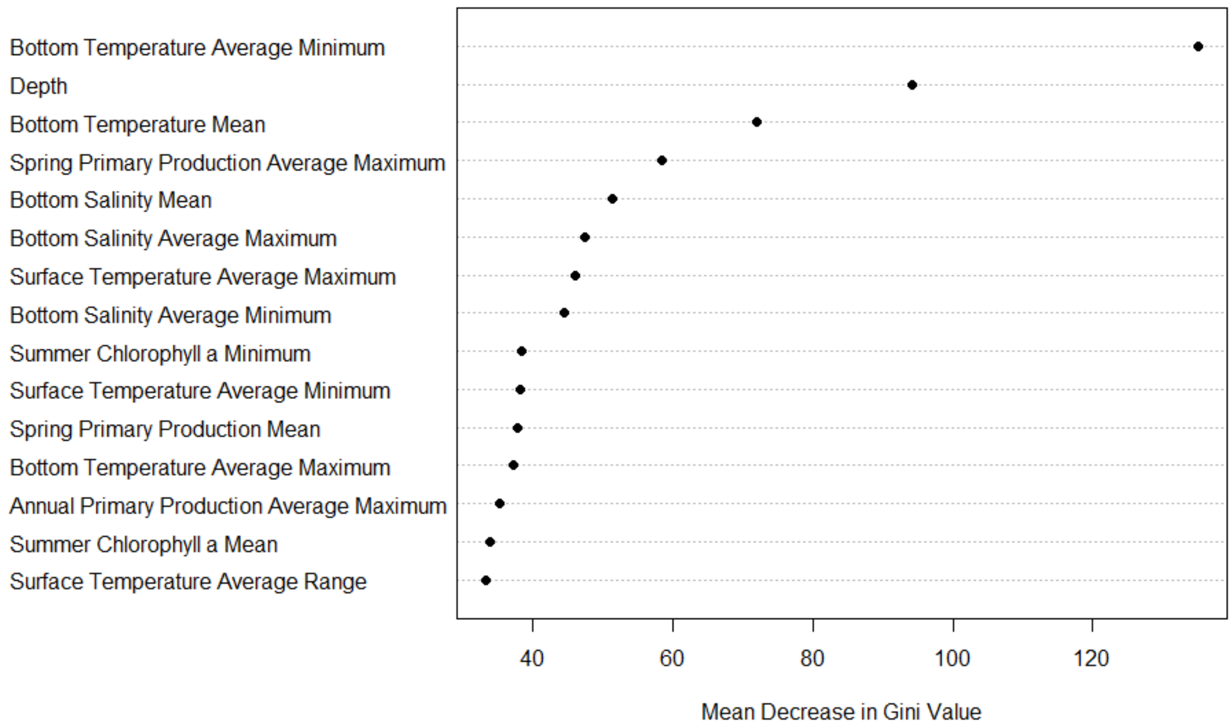

Figure 44. Importance of the top 15 predictor variables measures as the Mean Decrease in Gini value of the random forest model on unbalanced sand dollar presence and absence data in the Maritimes Region. The higher the Mean Decrease in Gini value the more important the variable is for predicting the response data.

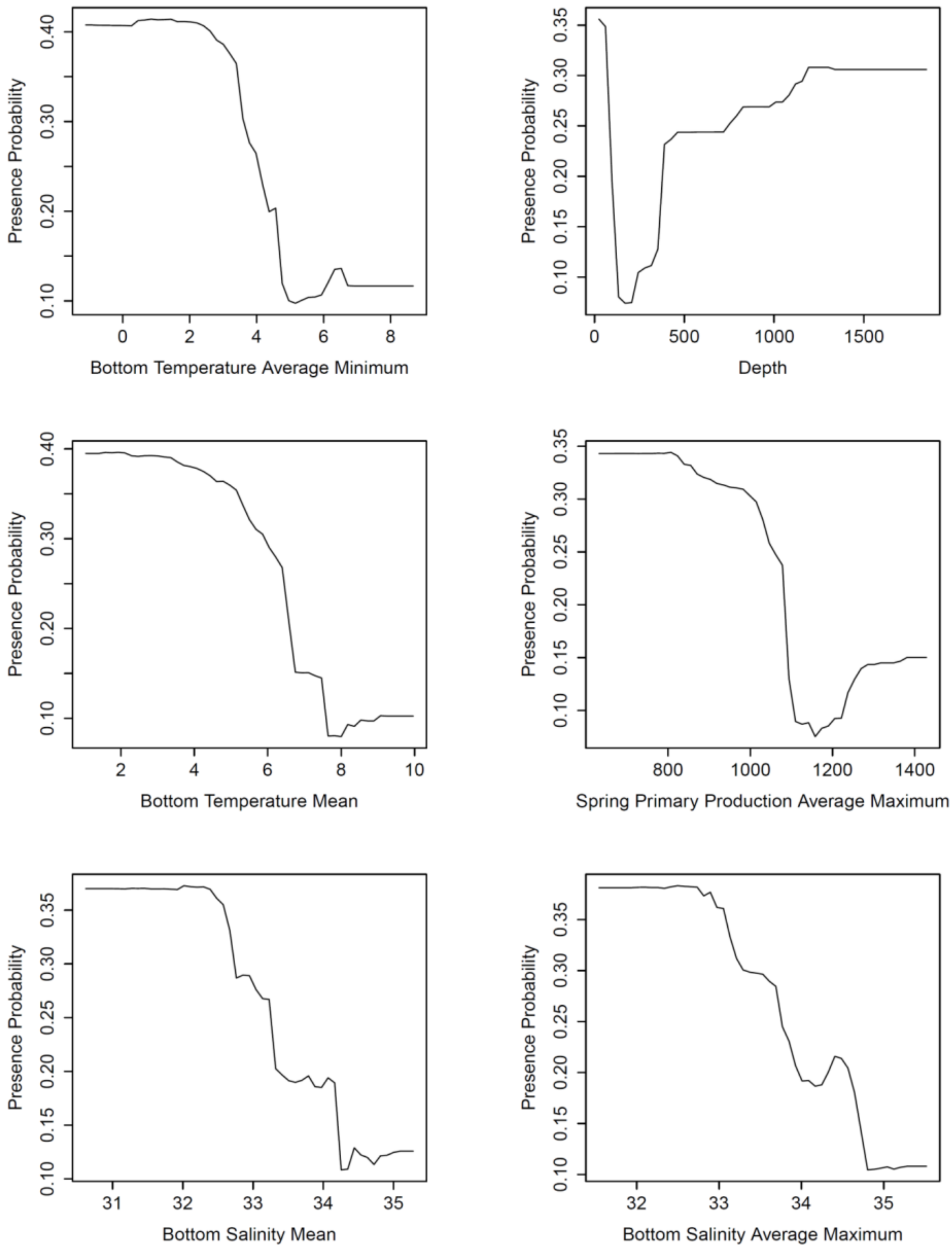

Figure 45. Partial dependence plots of the top six predictors from the random forest model of sand dollar unbalanced presence and absence data collected within the Maritimes Region, ordered left to right from the top. Presence probability is shown on the *y*-axis.

Model Selection

Model 2, which was generated using all the available sand dollar data from the DFO multispecies and scallop stock assessment surveys and a threshold equivalent to species prevalence (0.27), was chosen as the best predictor of sand dollar distribution in the Maritimes Region. The presence probability surfaces and accuracy measures were very similar between Models 1 and 2. However, we felt that the inclusion of all the absence data in Model 2 gave a more accurate depiction of the distribution of sand dollars throughout the region.

Validation of Selected Model Using Independent Data

Additional sand dollar records from DFO *in situ* benthic imagery surveys conducted in 1999 and 2008 were used for validation of random forest Model 2 (see Figure 46). The 1999 records were collected on Sable Island Bank east of Sable Island using a 'TowCam' camera system, while the 2008 records were collected along the southern edge of Banquereau Bank and at the head of the Gully submarine canyon using a video camera system called 'Campod'. The spatial extent of these records was limited, with records occurring in three isolated patches on Banquereau Bank where the random forest model predicted suitable habitat for sand dollars. Eighteen records remained after they were filtered to a single record per raster cell. These were located near the southern extent of a large area predicted as suitable habitat by the model. Of these 18 records, 2 (11%) were predicted as absence based on prevalence (0.27). These occurred in small pockets of unsuitable habitat within the larger area of suitable habitat and should therefore not be over-interpreted (Kenchington et al., 2016a).

The extrapolated area off southwest Nova Scotia that is predicted as suitable habitat for sand dollars by the model (see Figure 43) has a variety of substrates including large boulders or bedrock, glacial till, and soft sediments (Brown et al., 2012). Dense sand dollar beds have been found on sandy substrate here (in Scallop Fishing Area 29 on German Bank specifically; Brown et al., 2012)) which helps further validate the model results in this unsampled area.

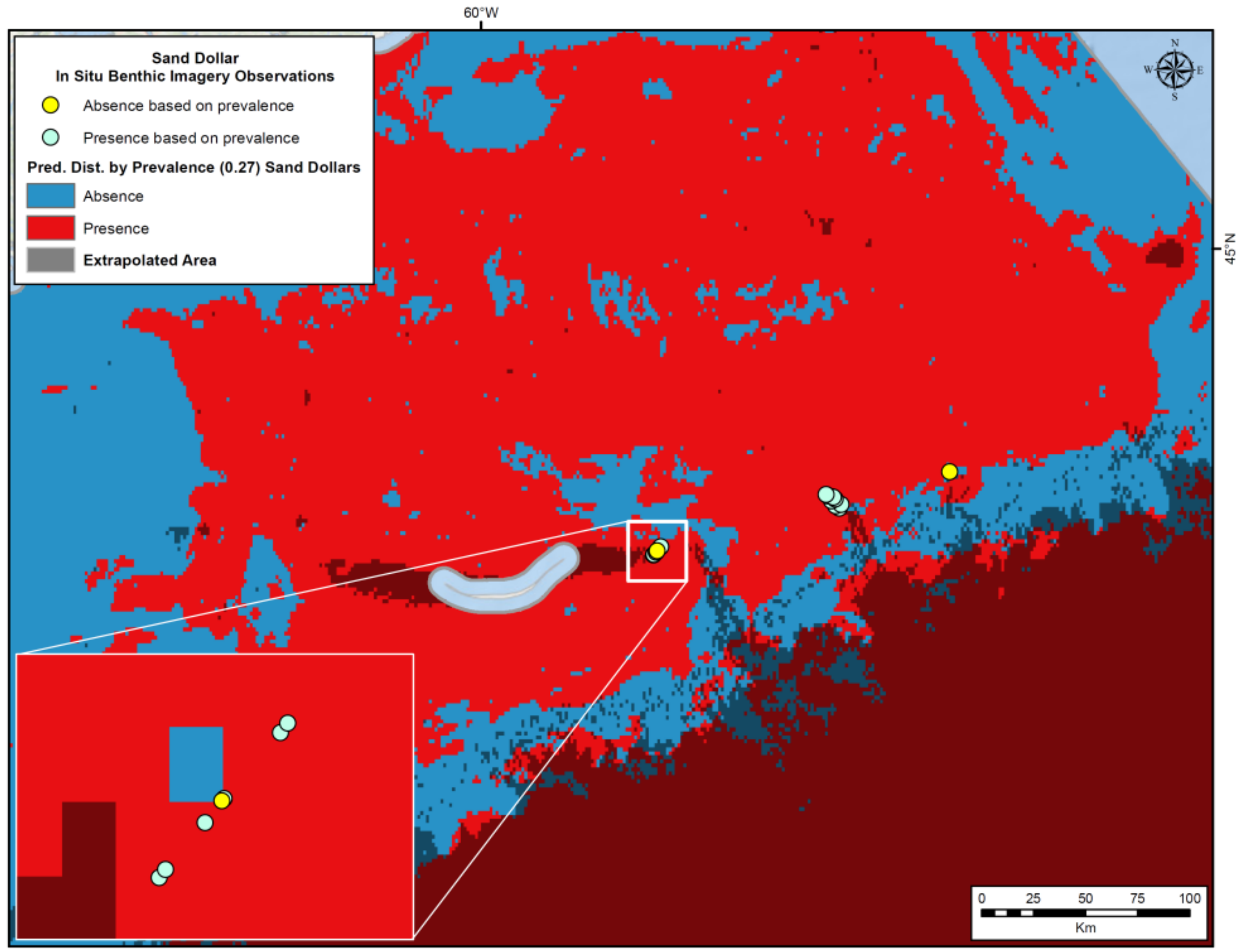

Figure 46. Validation of the prevalence surface from Model 2 using sand dollar records from DFO *in situ* benthic imagery surveys conducted in 1999 and 2008.

Prediction of Sand Dollar Biomass Using Random Forest

The accuracy measures of the regression random forest model on mean sand dollar biomass per grid cell from DFO multispecies trawl surveys are presented in Table 14. These data likely capture individuals on the surface and not those in the sediments, and consequently are underestimates of true biomass. The highest R^2 value was 0.215, while the average was 0.094 ± 0.071 SD, indicating poor model performance. The average Normalized Root-Mean-Square Error (NRMSE) was 0.027 ± 0.012 SD. The average percent variance explained was 2.05% ± 1.43 SD.

Figures 47 and 48 show the predicted biomass surface of sand dollars. The majority of the spatial extent was predicted to have a low (> 0 – 0.41 kg) sand dollar biomass. Small pockets of high biomass were predicted to occur on the banks of the eastern Scotian Shelf. These areas coincided with the highest biomass records (see Figure 48). Overall the biomass of sand dollars was under-predicted by the model, with a maximum predicted value of 14.9 kg, approximately half of the empirical maximum biomass in the raw data (30.85 kg).

Depth and Slope (two non-interpolated variables) were the top predictors of the sand dollar biomass data (see Figure 49). The partial dependence plots shown that predicted biomass was highest at the shallowest depths and on slopes greater than 2˚ (Figure 50).

Table 14. Accuracy measures from 10-fold cross validation of a random forest model of average sand dollar biomass (kg) per grid cell recorded from DFO multispecies trawl surveys in the Maritimes Region. RMSE = Root-Mean-Square Error; NRMSE = Normalized Root-Mean-Square Error.

Model Fold	**R^2**	**RMSE**	**NRMSE**	**Percent (%) variance explained**
1	0.098	0.642	0.021	1.09
2	0.133	0.758	0.025	0.94
3	0.215	0.463	0.015	0.06
4	0.091	0.955	0.031	1.20
5	0.027	1.492	0.048	3.37
6	0.042	0.669	0.022	3.31
7	0.048	0.801	0.026	2.83
8	0.028	1.404	0.045	3.71
9	0.211	0.419	0.014	0.43
10	0.049	0.826	0.027	3.58
Mean	**0.094**	**0.843**	**0.027**	**2.05**
SD	**0.071**	**0.358**	**0.012**	**1.43**

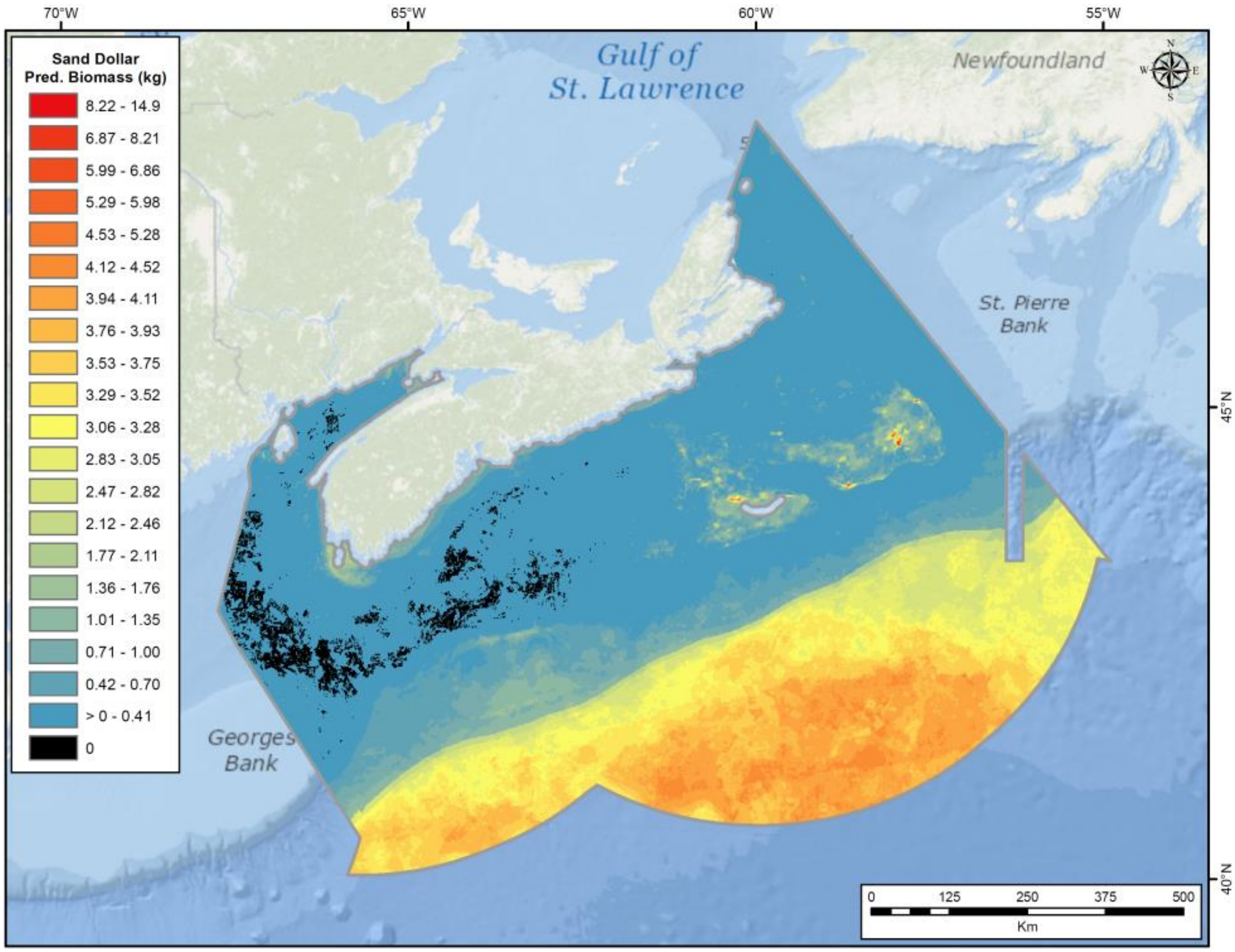

Figure 47. Predictions of biomass (kg) of sand dollars from the random forest model on catch data recorded in DFO multispecies trawl surveys conducted in the Maritimes Region between 1999 and 2015.

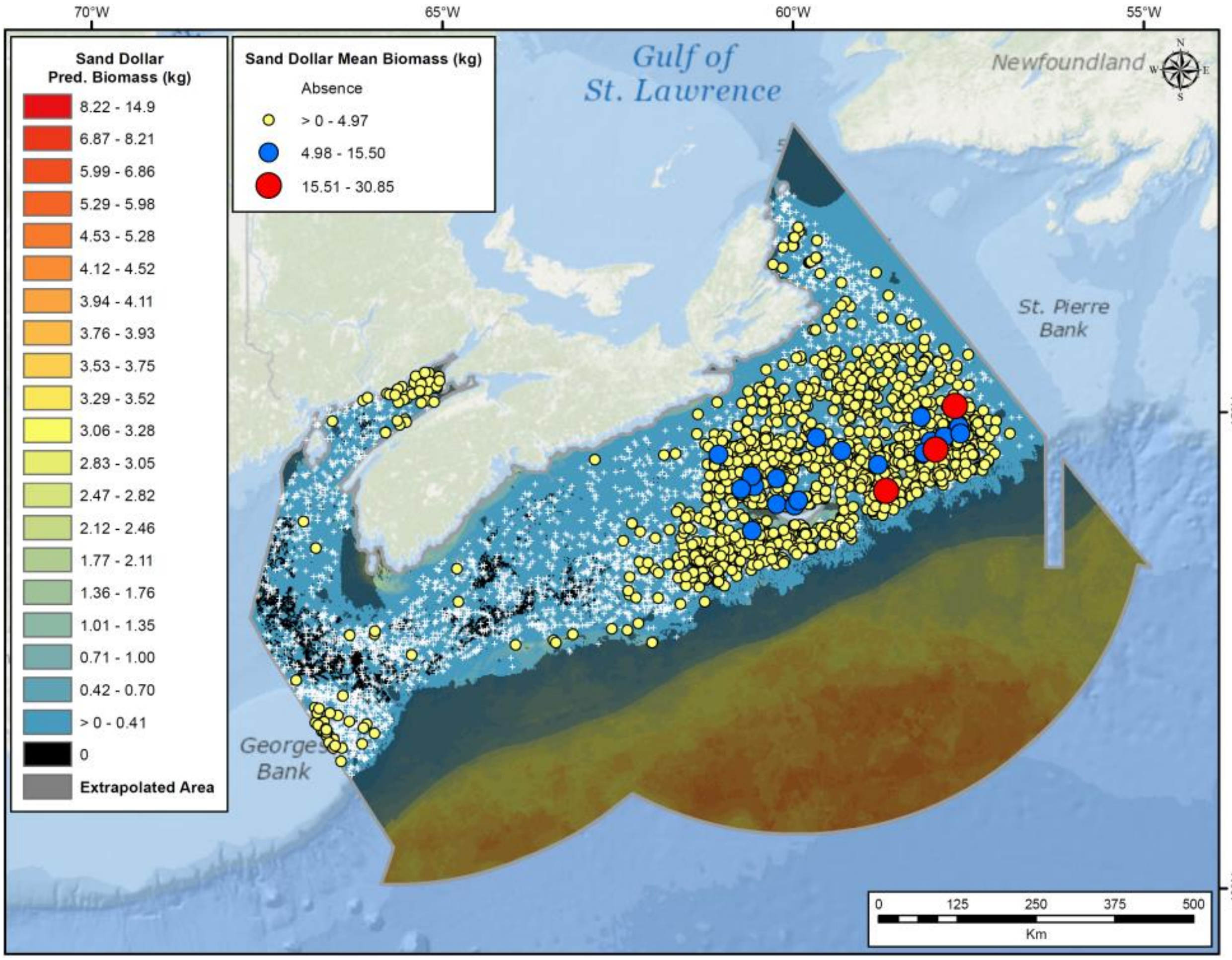

Figure 48. Predictions of biomass (kg) of sand dollars from the random forest model on catch data recorded in DFO multispecies trawl surveys conducted in the Maritimes Region between 1999 and 2015. Also shown are the mean biomass values per grid cell and areas of model extrapolation.

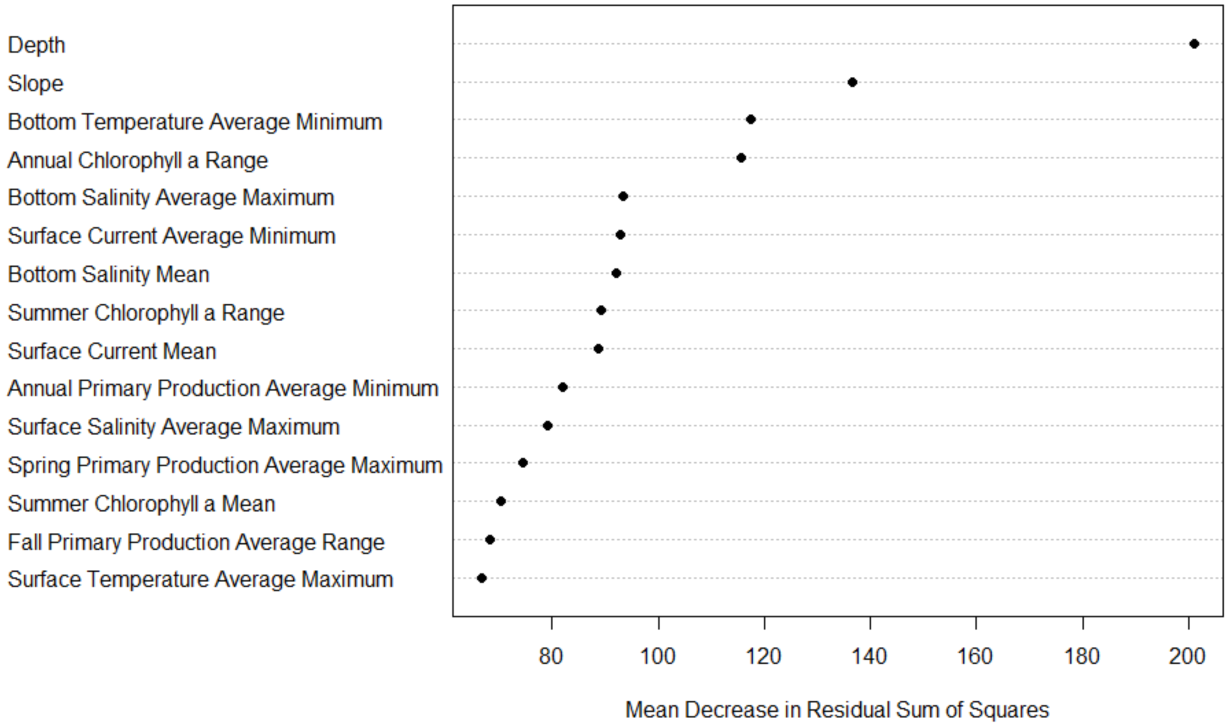

Figure 49. Importance of the top 15 predictor variables measured as the Mean Decrease in Residual Sum of Squares of the regression random forest model on sand dollar mean biomass per grid cell. The higher the Mean Decrease in Residual Sum of Squares the more importance the variable is for predicting the response data.

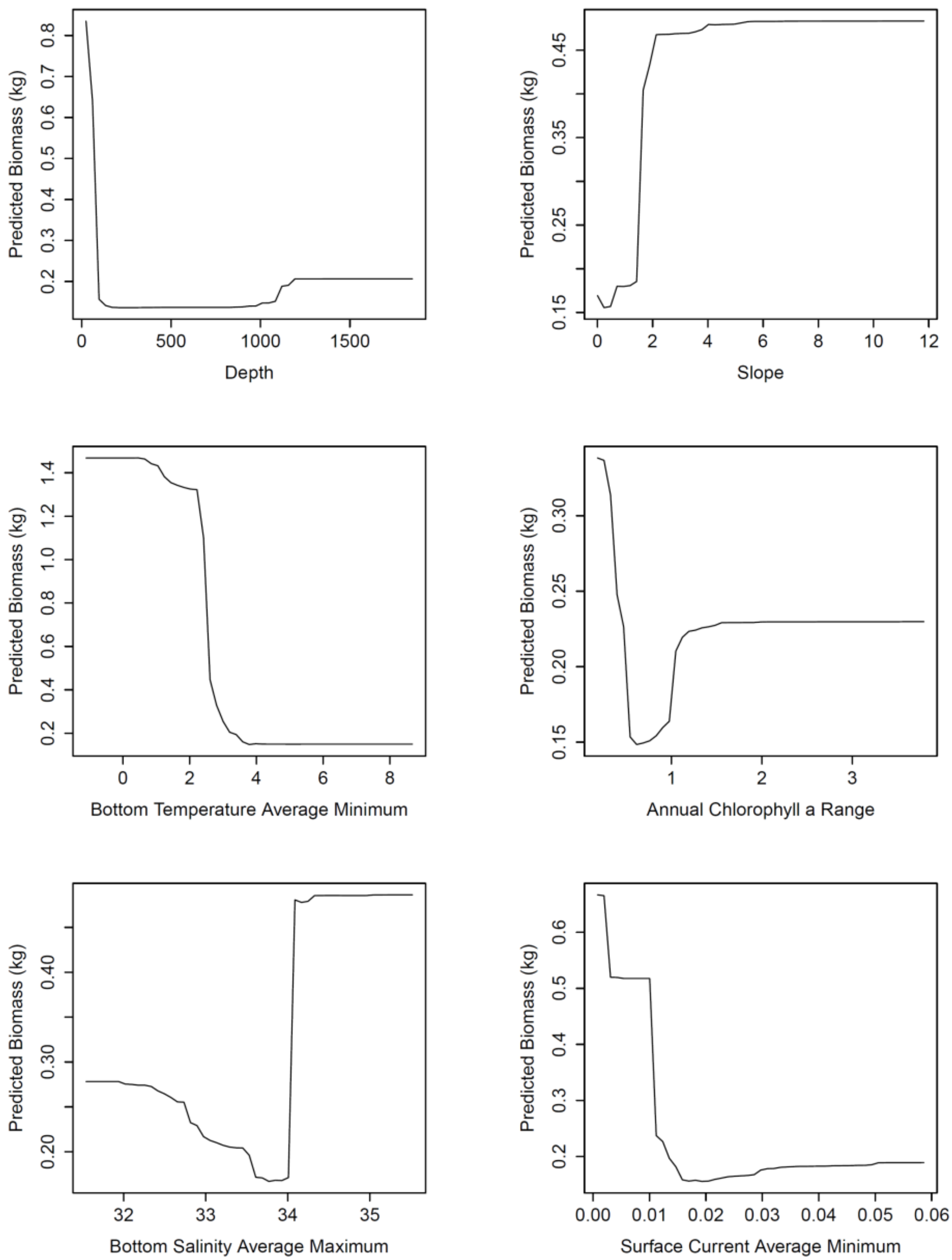

Figure 50. Partial dependence plots of the top six predictors from the random forest model of sand dollar biomass collected within the Maritimes Region, ordered left to right from the top. Predicted biomass (kg) is shown on the *y*-axis.

Kernel Density Analysis of Sand Dollar Catch Data

Kernel density analysis of the sand dollar biomass data showed concentrations of sand dollars on the eastern Scotian Shelf, with smaller concentrations in the Bay of Fundy and on Georges Bank (Figure 51). As for the RF regression models with biomass, the data are not representative of true biomass due to the infaunal burrowing of the species, however the approach can identify relative epifaunal biomass. The largest concentrations were observed on Banquereau Bank. The KDE search radius for this group was 20.2 km. Polygons encompassing each of 13 catch thresholds showed that the greatest increase in area between successive weights occurred between 10 kg and 5 kg, where the area increased by 1837% (Table 15, Figure 52). However, this threshold was only supported by 3 additional points and was discounted as it is indicative of the initial steps of the analyses (Kenchington et al., 2016b) where the habitats are being delineated. The next threshold reviewed was the 5 kg threshold (Figure 52). This showed a 77% increase in area in going from 5 to 3 kg, however, there were only 4 data points supporting this increase (Table 15) and it was due to new areas being defined around those points. The selected threshold used to define the sand dollar habitat was 2 kg. At this threshold the polygon area increased by 68% in going to 1 kg and was supported by 7 data points (Table 15, Figure 53). Given that catchability is likely to be low as the sand dollars bury, this is also a precautionary threshold. These sand dollar beds occupy 20,137 km^2 of seabed. Figure 54 shows the final KDE significant area polygons, the location of significant and non-significant catches, and null records of sand dollars from the DFO multispecies trawl survey.

There was good spatial congruence between the location of the KDE polygons and areas predicted as suitable habitat of sand dollars from random forest Model 2 (Figure 55). Much of the large area on the eastern Scotian Shelf predicted as sand dollar ‘presence’ was encompassed by KDE polygons. A large area of suitable habitat on Misaine Bank was not covered by a KDE polygon. Although numerous, the biomass of these catches was smaller (see Figure 48). Following the experience in the CSAS process (Kenchington et al., 2016a) these polygons would not be candidates for trimming based on the SDM results. The KDE polygons were not compared to the RF biomass results for this species due to the poor performance of that model.

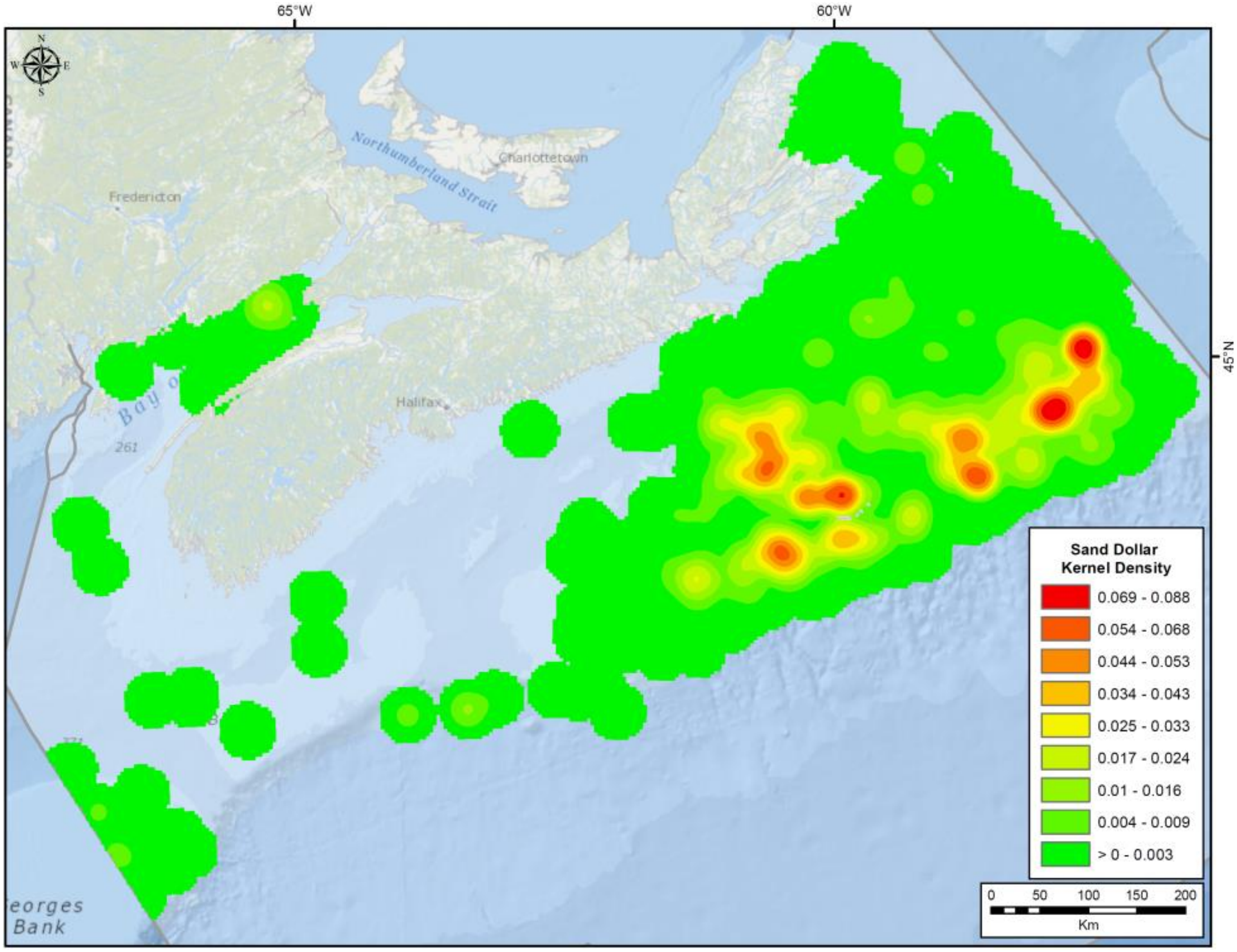

Figure 51. Kernel density analysis of sand dollar catches in the Maritimes Region.

Table 15. Summary characteristics for 13 sand dollar catch weight thresholds. The threshold corresponding to the habitat delineation is indicated by shading.

Number of Points in Polygon	Area (km^2)	Catch Threshold (kg)	Percent Change in Area (%)
4	84	15	431.22
8	446	10	1836.88
11	8640	5	77.08
15	15300	3	31.61
18	20137	2	68.41
25	33911	1	9.81
32	37239	0.75	21.12
39	45105	0.5	23.52
53	55714	0.25	25.86
58	70124	0.1	12.21
68	78684	0.05	27.85
85	100601	0.01	3.84
145	104466	0.001	

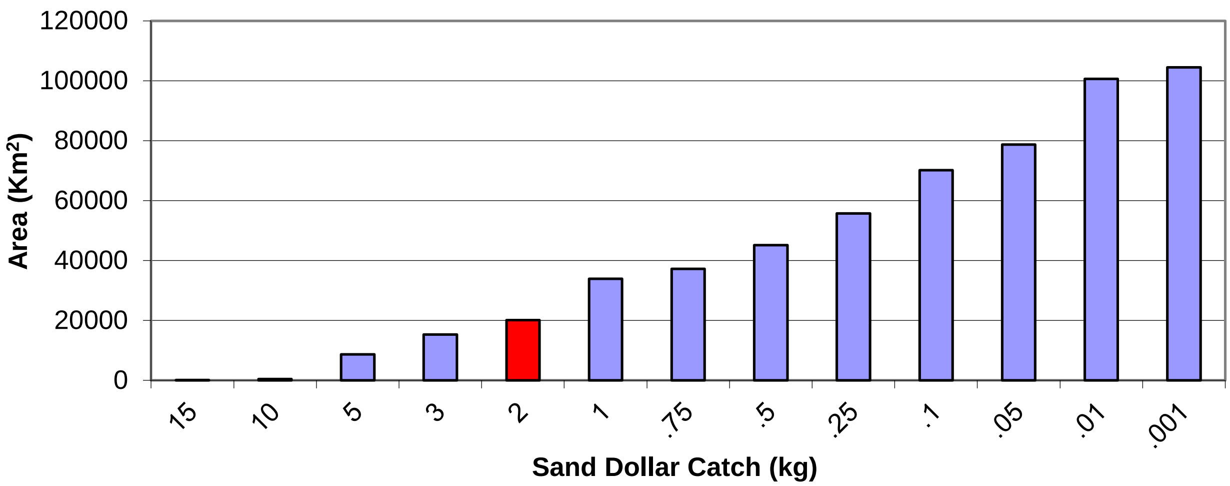

Figure 52. Area occupied by polygons encompassing catches of sand dollars greater than or equal to each of 13 catch thresholds (15 kg to 0.001 kg). The red bar indicates the catch threshold where the percent area increased the greatest between adjacent bins after the initial delineation of the sand dollar beds.

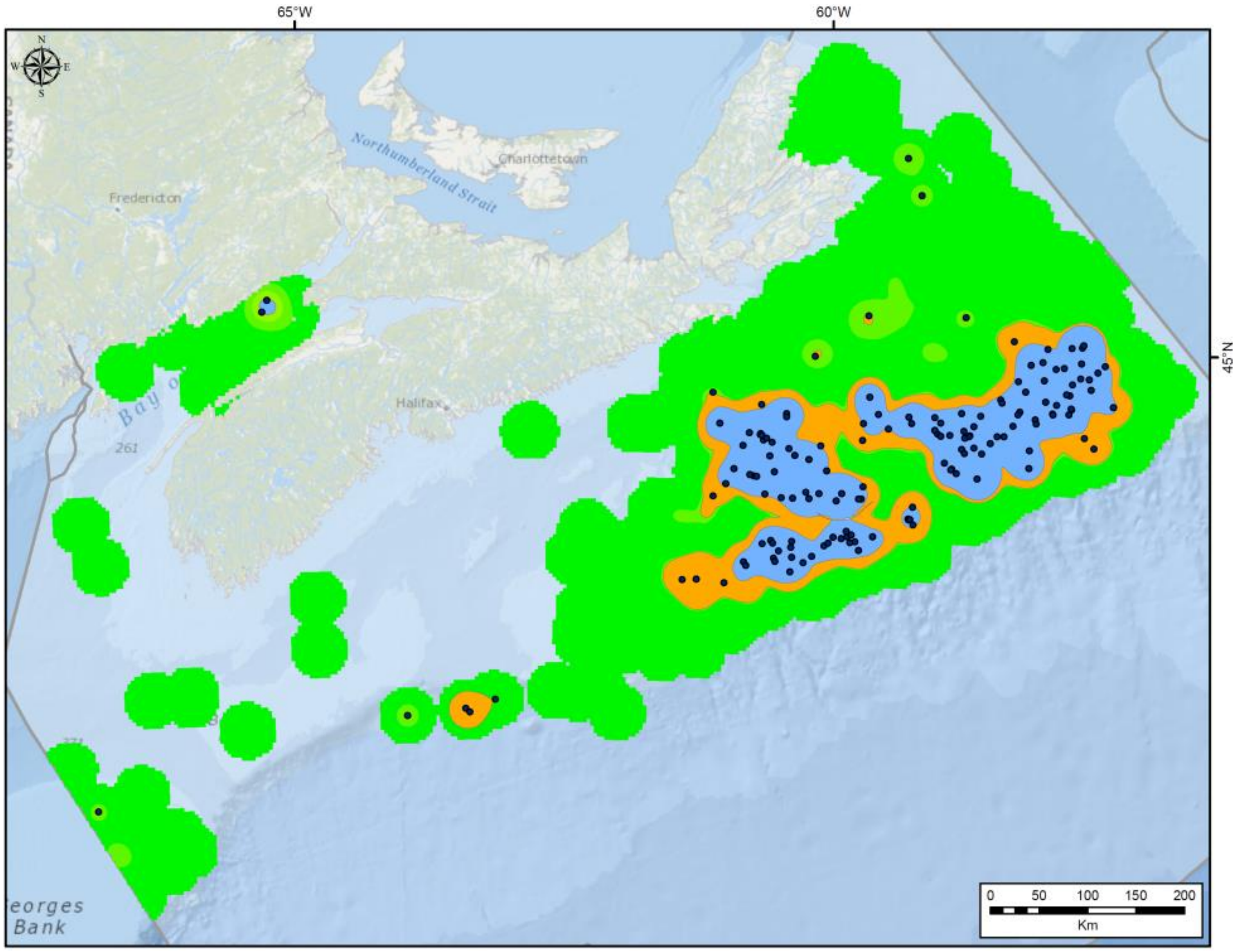

Figure 53. Polygons encompassing sand dollar catch weights greater than or equal to the 2 kg (blue) and 1 kg (orange) overlain on the kernel density surface. Tow locations associated with biomass values above these weight thresholds are also shown. The 2 kg threshold was used to delineate the habitat.

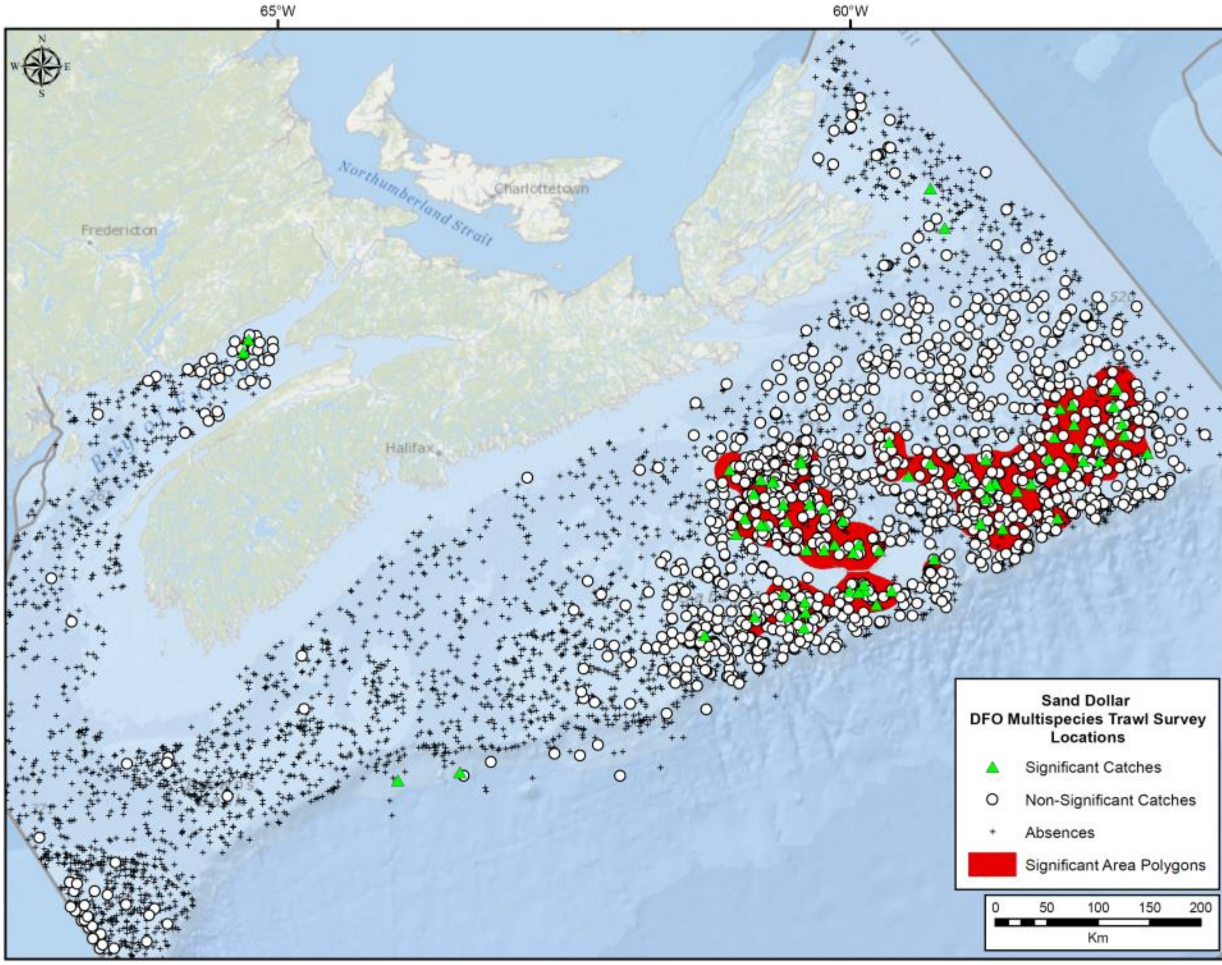

Figure 54. Location of polygons identifying significant sand dollar concentrations in the Maritimes Region. Also shown are the significant and non-significant catches of sand dollars and absence data (nulls) from the DFO multispecies trawl surveys.

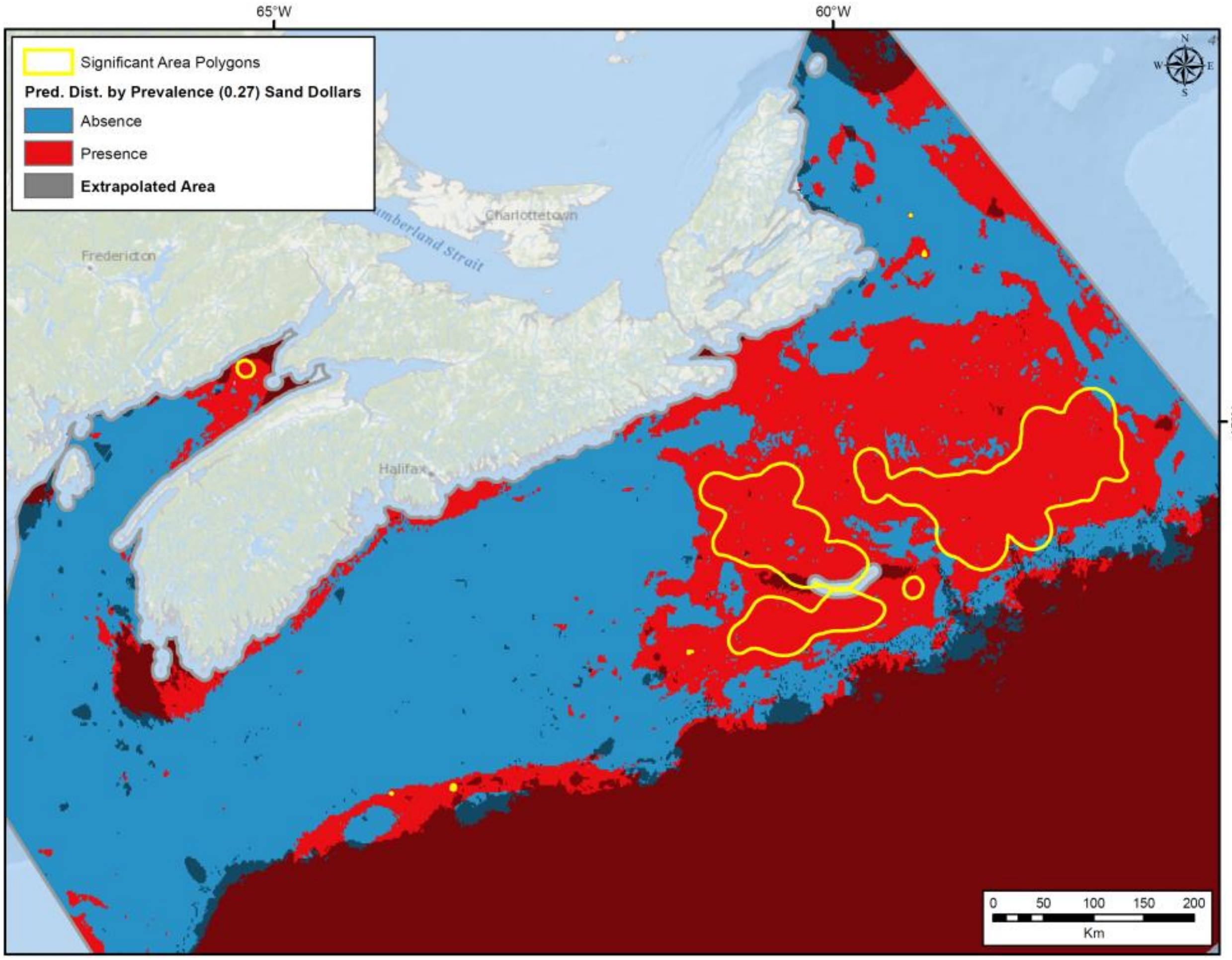

Figure 55. KDE polygons denoting significant concentrations of sand dollars overlain on the predicted distribution by prevalence of sand dollars generated from random forest Model 2.

Soft Coral Gardens

Data Sources and Distribution

Figure 56 shows the distribution of available soft coral records in the Maritimes Region. Aside from the *in situ* benthic imagery records, DFO multispecies trawl survey records using Western IIA gear (white circles) accounted for the majority of soft coral records in the region. These records were distributed mainly on the eastern Scotian Shelf and in the mouth of the Gulf of Maine. The *in situ* benthic imagery records were concentrated in the canyons of the Scotian Shelf.

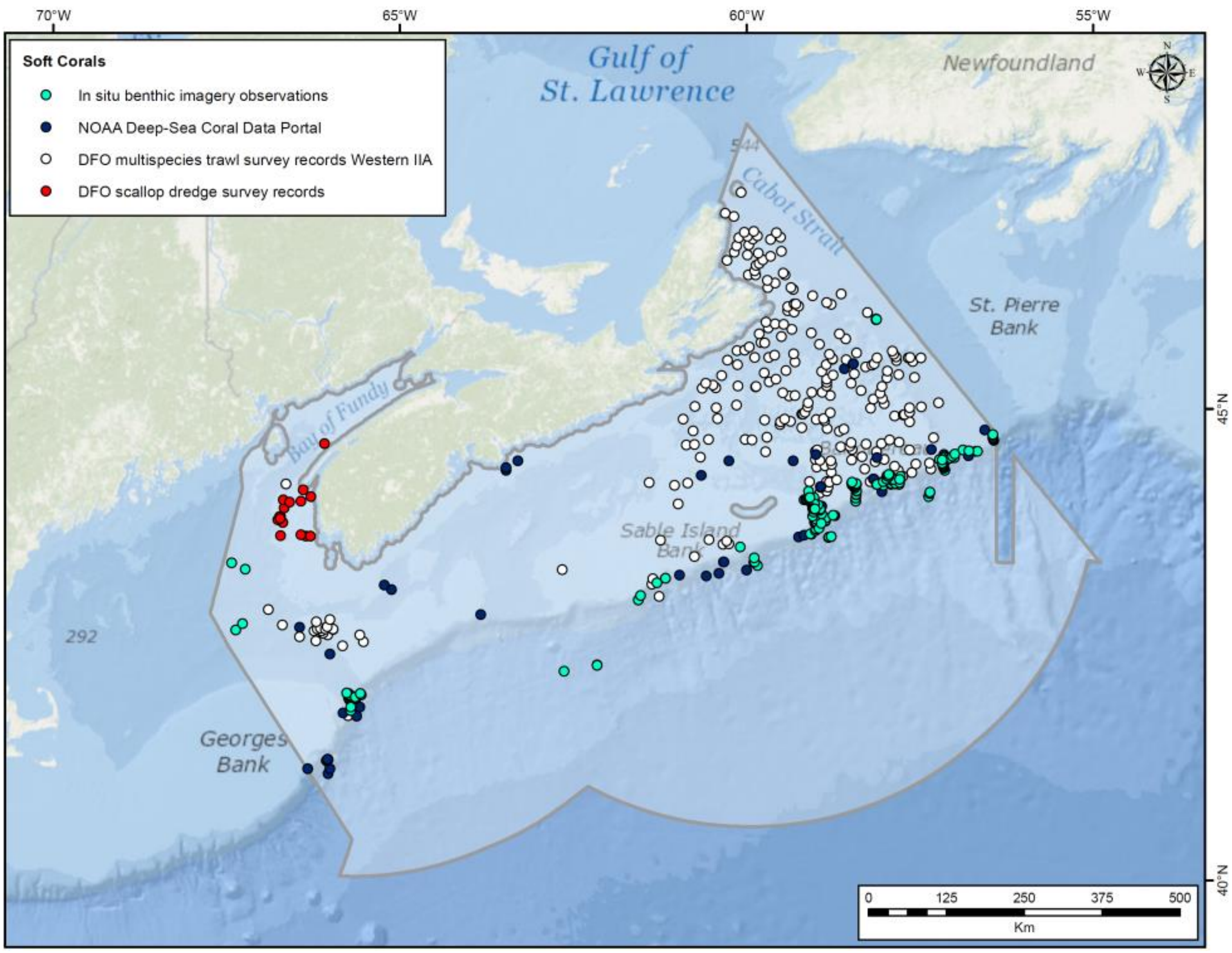

Figure 56. Available soft coral presence data in the Maritimes Region from scientific missions, the NOAA Deep-Sea Coral Data Portal, and DFO multispecies and scallop dredge surveys.

Soft corals collected in the DFO multispecies trawl surveys are identified to one of three genera/groups: *Gersemia rubiformis*, *Anthomastus grandiflorus*, or 'unidentified soft coral' (see

Table 2). The unidentified soft coral group could be comprised of a number of different soft coral species (e.g. *Duva florida*, *Drifa glomerata*) that are known to occur in the region but are difficult to distinguish without thorough examination (Cogswell et al., 2009). Queries into the recording of biomass data for this group while at sea has raised some doubts as to whether the corals are consistently removed from their rocks before weighing. Consequently, their biomass data may be unreliable. Initial random forest models of soft corals were run using only catch data from DFO multispecies trawl surveys to avoid mixing gears with different catchabilities (Figure 57). This dataset consisted of 238 presence and 2348 absence records (prevalence = 0.09) collected over a period of 12 years from 2003 to 2014 (Table 16).

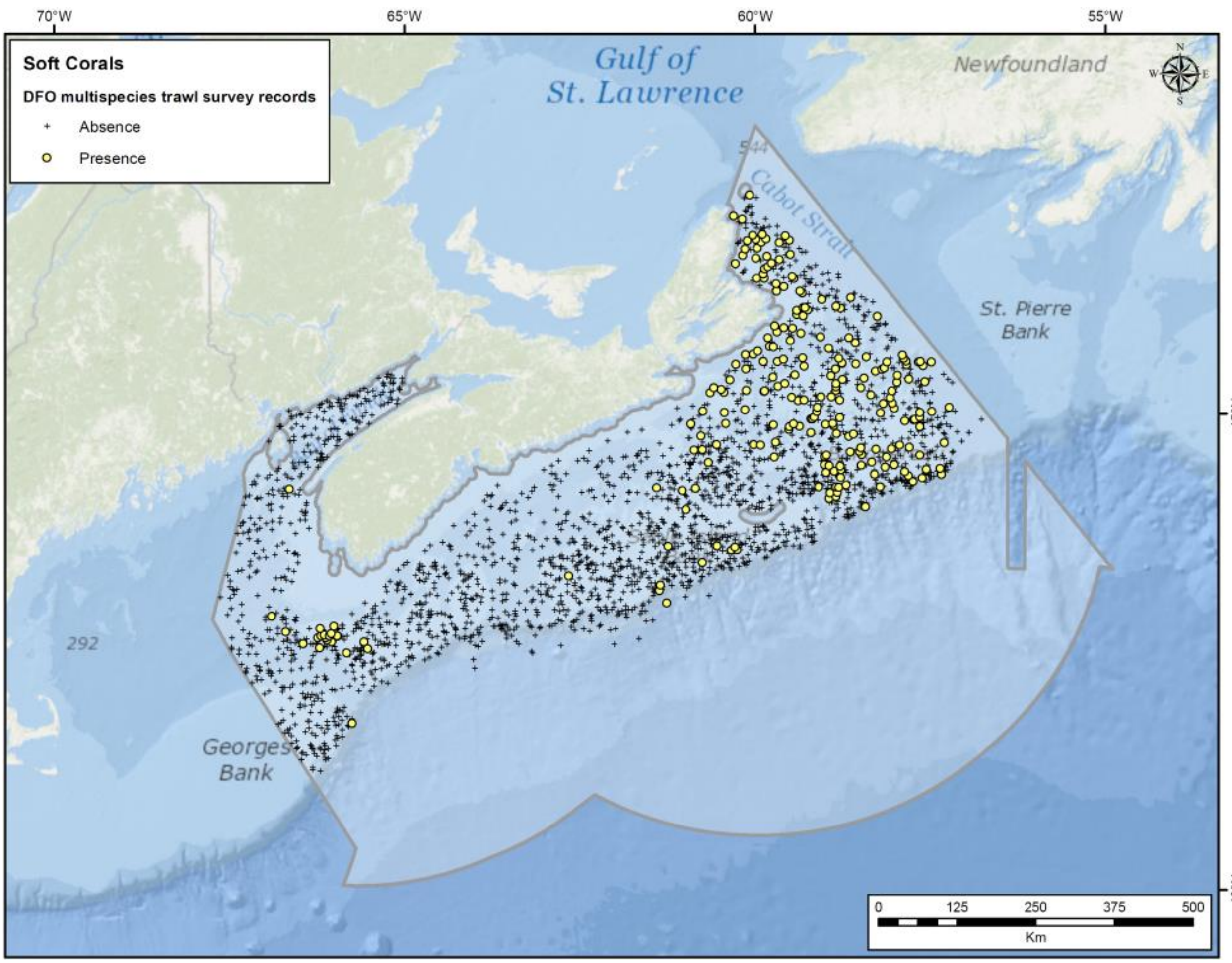

Figure 57. Presence and absence records of soft corals recorded in DFO multispecies trawl surveys conducted between 2003 and 2014 in the Maritimes Region.

Table 16. Number of presence and absence records of soft coral catch recorded from DFO multispecies trawls surveys conducted between 2003 and 2014 in the Maritimes Region.

Year	Total number of presences	Total number of absences
2003	7	68
2005	36	73
2006	40	317
2007	28	230
2008	27	199
2009	18	226
2010	26	310
2011	22	235
2012	9	257
2013	9	207
2014	16	226
Total:	**238**	**2348**

Random Forest Model Results

The random forest model on the presence-absence soft coral data had a very good performance rating, with an average AUC of 0.830 ± 0.044 SD (Table 17). Class error was higher for the absence class than the presence class (0.252 versus 0.185). Class error for both the presence and absence classes was low, with the absence records being incorrectly classified more often than the presence records. Sensitivity and specificity of this model were high (0.815 and 0.748, respectively).

The predicted presence probability surface of soft corals from the model is shown in Figure 58. The eastern Scotian Shelf was predicted to have the highest occurrence of soft corals, particularly on St. Anns Bank off Cape Breton. Most of the Scotian Slope, western Scotian Shelf, Gulf of Maine, and Bay of Fundy were predicted to have a low or zero occurrence of soft corals. Figure 59 shows the predicted surface overlain with the presence-absence observations and areas of model extrapolation. This model did not appear to over-predict areas of suitable habitat outside the range distribution of the training data, and there was little interpolation of high presence probability between data observations. Areas of extrapolation did not overlap with areas of high presence probability. Predicted distribution by prevalence (Figure 60) showed areas of suitable habitat being located along the coast, on the shallow banks of the eastern Scotian

Shelf, and along the eastern Scotian Slope. The area of suitable habitat off southwest Nova Scotia that is also considered an area of extrapolation may support soft corals due to the presence of hard bottom, but they have not been recorded there (Brown et al., 2012).

Bottom and surface temperature variables were the most important for the classification of the soft coral presence-absence data (see Figure 61). Bottom Temperature Average Minimum was the top variable, followed by Bottom Temperature Mean and Surface Temperature Mean. Slope and Depth were the 5th and 6th most important variables. Bottom Temperature Average Minimum showed a right-skewed distribution prior to modelling (Beazley et al., in prep.). Examination of the Q-Q plot revealed a spatial pattern to data points over- and under-predicted by a normal distribution, with over-predicted points located in the Gulf of Maine, Emerald Basin, and on Banquereau and Misaine Banks, and in the deepest regions of the study extent, and under-predicted points located in Bay of Fundy, off southwestern Nova Scotia, Laurentian Channel, and in deep water beyond the shelf break. Partial dependence of the soft coral presence and absence data on the top 6 predictor variables is shown in Figure 62. Presence probability was highest at Bottom Temperature Average Minimum values less than 4˚C. Values in this range were predicted well by the interpolation model. Presence probability increased in a step-like fashion along the gradient in Slope. Presence probability was highest at the shallowest depths, then decreased between ~100 and 300 m, and then increased in a step-like fashion and levelled off at 1500 m. This pattern is likely indicative of the distribution of the different species within the soft coral group, with *Gersemia rubiformis* abundant in shallower waters less than 100 m depth, and *Anthomastus grandiflorus* and the other species more prominent at deeper depths (Kenchington et al., 2015).

Table 17. Accuracy measures and confusion matrix from 10-fold cross validation of a random forest model of presence and absence of soft corals within the Maritimes Region. Observ. = Observations; Sensit. = Sensitivity, Specif. = Specificity.

Model Fold	AUC	Observ.	Predictions		Total n	Class error	Sensit.	Specif.
1	0.886		**Absence**	**Presence**				
2	0.833	**Absence**	1757	591	2348	0.252	0.815	0.748
3	0.807	**Presence**	44	194	238	0.185		
4	0.804							
5	0.827							
6	0.924							
7	0.838							
8	0.801							
9	0.790							
10	0.788							
Mean	**0.830**							
SD	**0.044**							

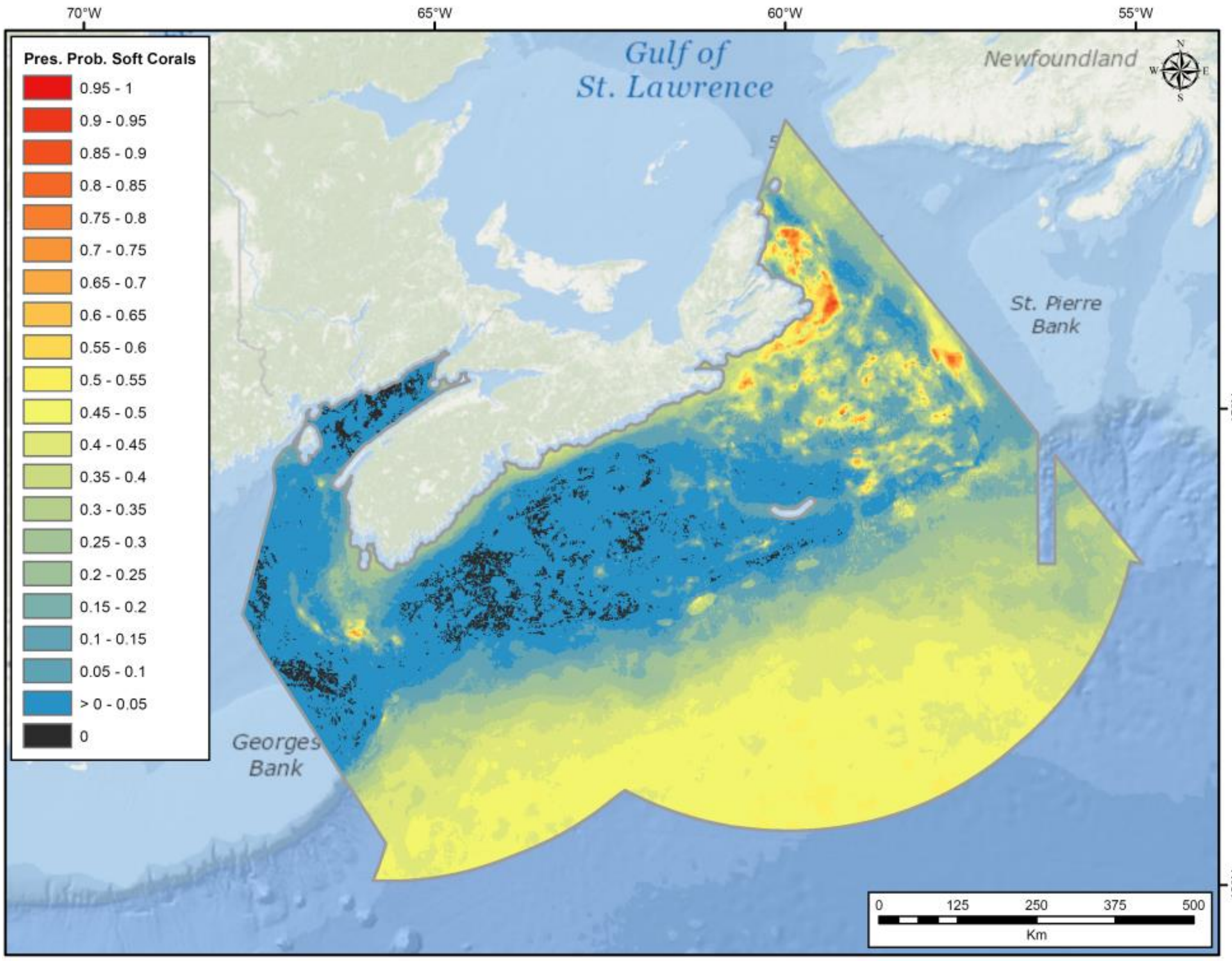

Figure 58. Predictions of presence probability (Pres. Prob.) of soft corals based on a random forest model on presence and absence soft coral catch data collected from DFO multispecies trawl surveys conducted within the Maritimes Region between 2003 and 2014.

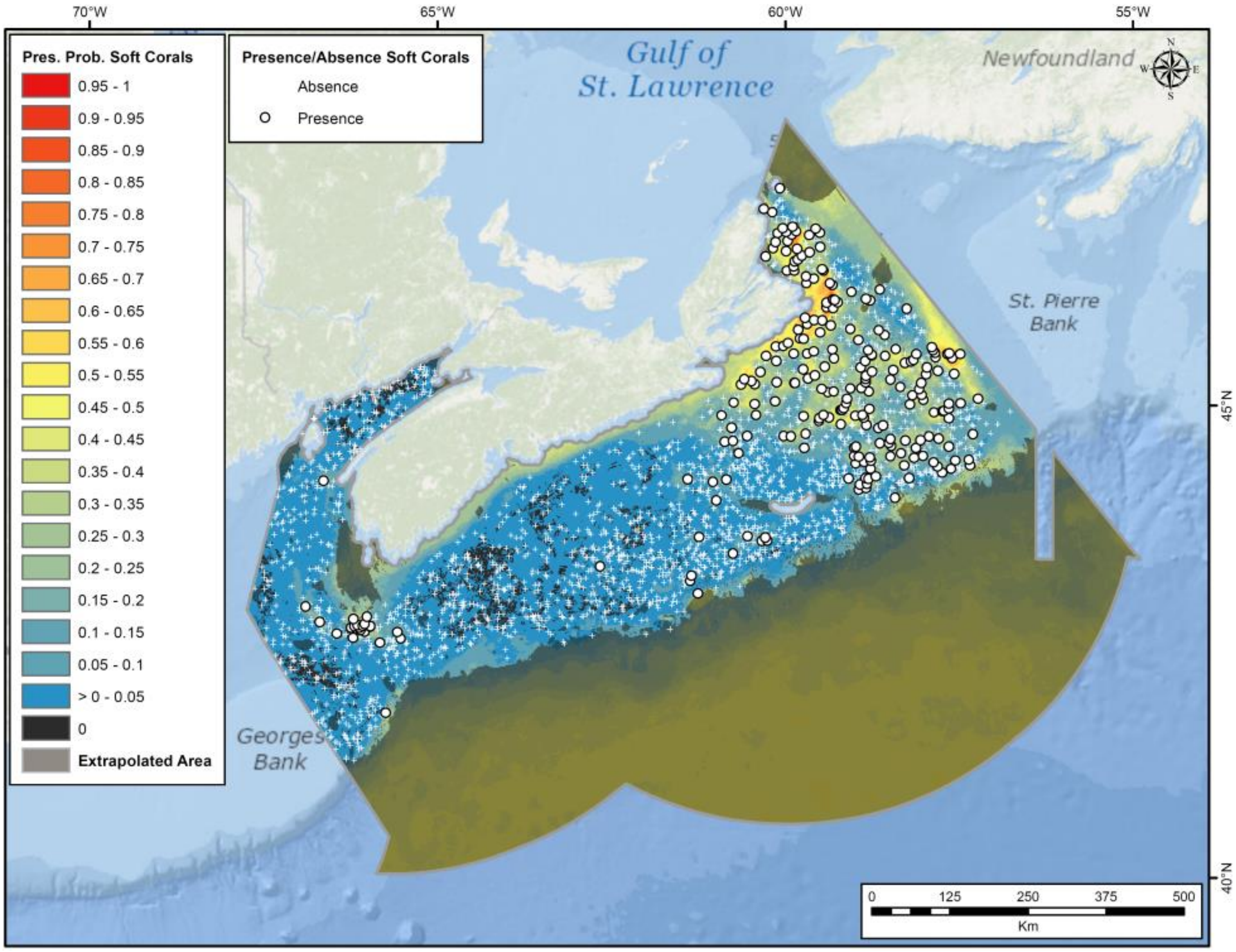

Figure 59. Presence and absence observations and predictions of presence probability (Pres. Prob.) of soft corals based on a random forest model on presence and absence soft coral catch data collected from DFO multispecies trawl surveys conducted within the Maritimes Region between 2003 and 2014. Also shown are the areas of model extrapolation.

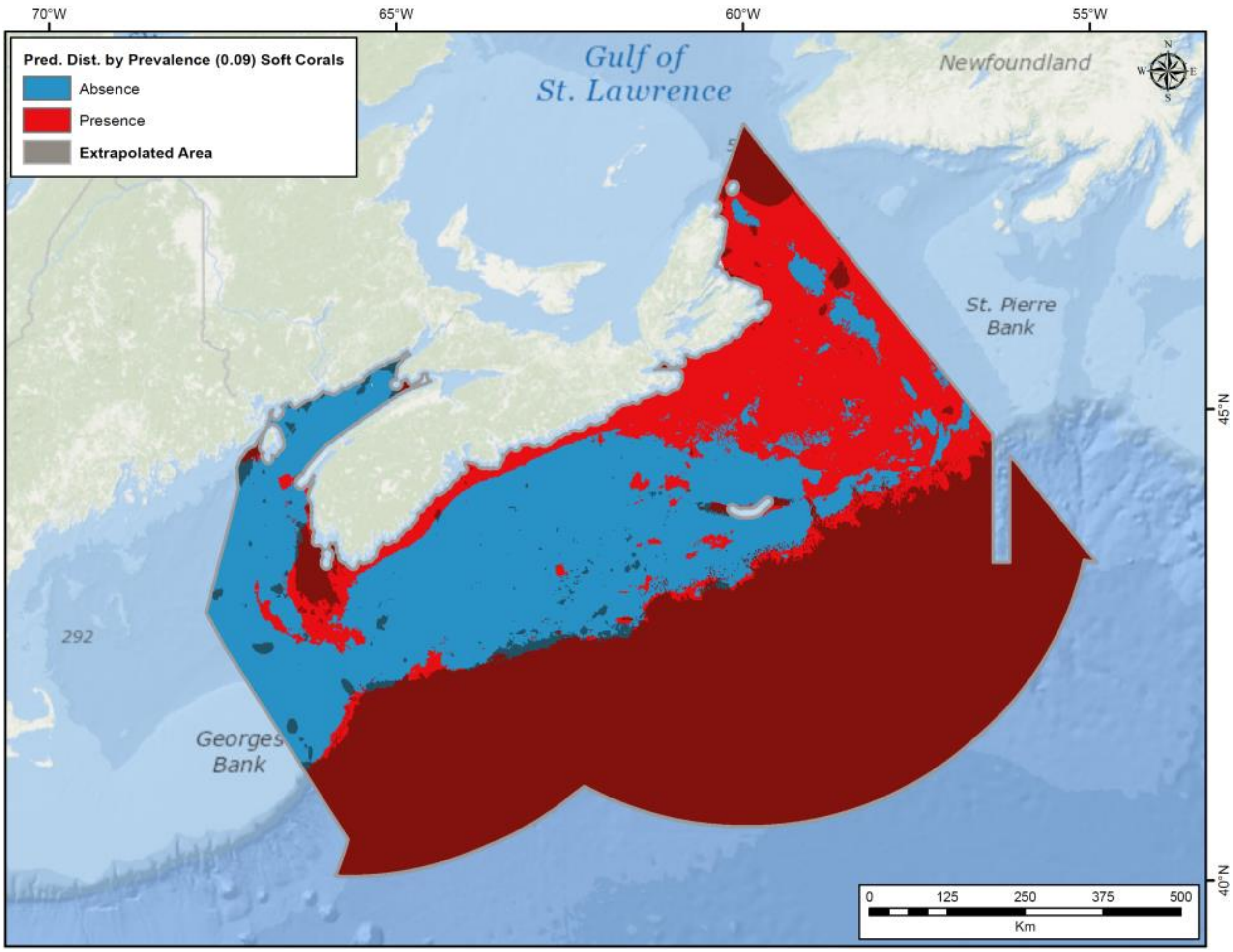

Figure 60. Predicted distribution (Pred. Dist.) of soft corals in the Maritimes Region based on the prevalence threshold of 0.09 of the soft coral presence and absence data used in the random forest model. Also shown are the areas of model extrapolation (grey polygon may appear red or blue when overlain on the prevalence surface).

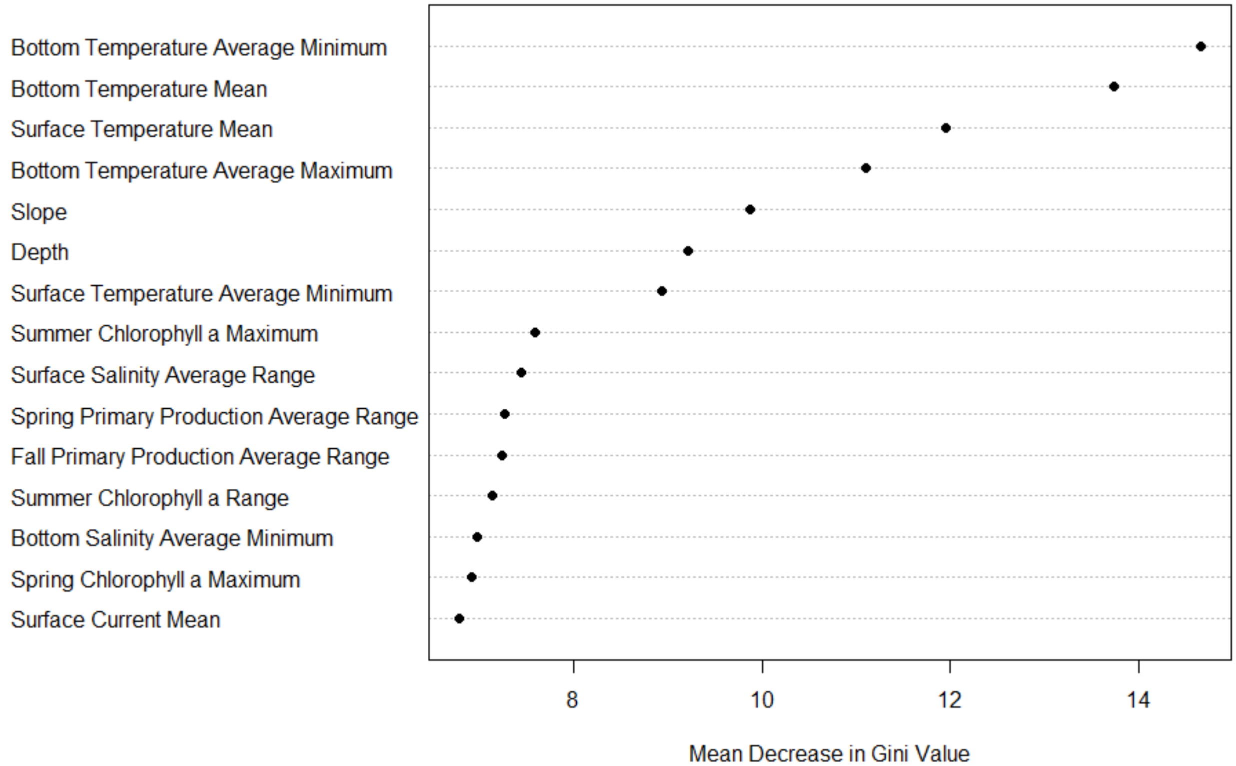

Figure 61. Importance of the top 15 predictor variables measured as the Mean Decrease in Gini value of the random forest model on soft coral presence and absence data within the Maritimes Region. The higher the Mean Gini value the more important the variable is for predicting the response data.

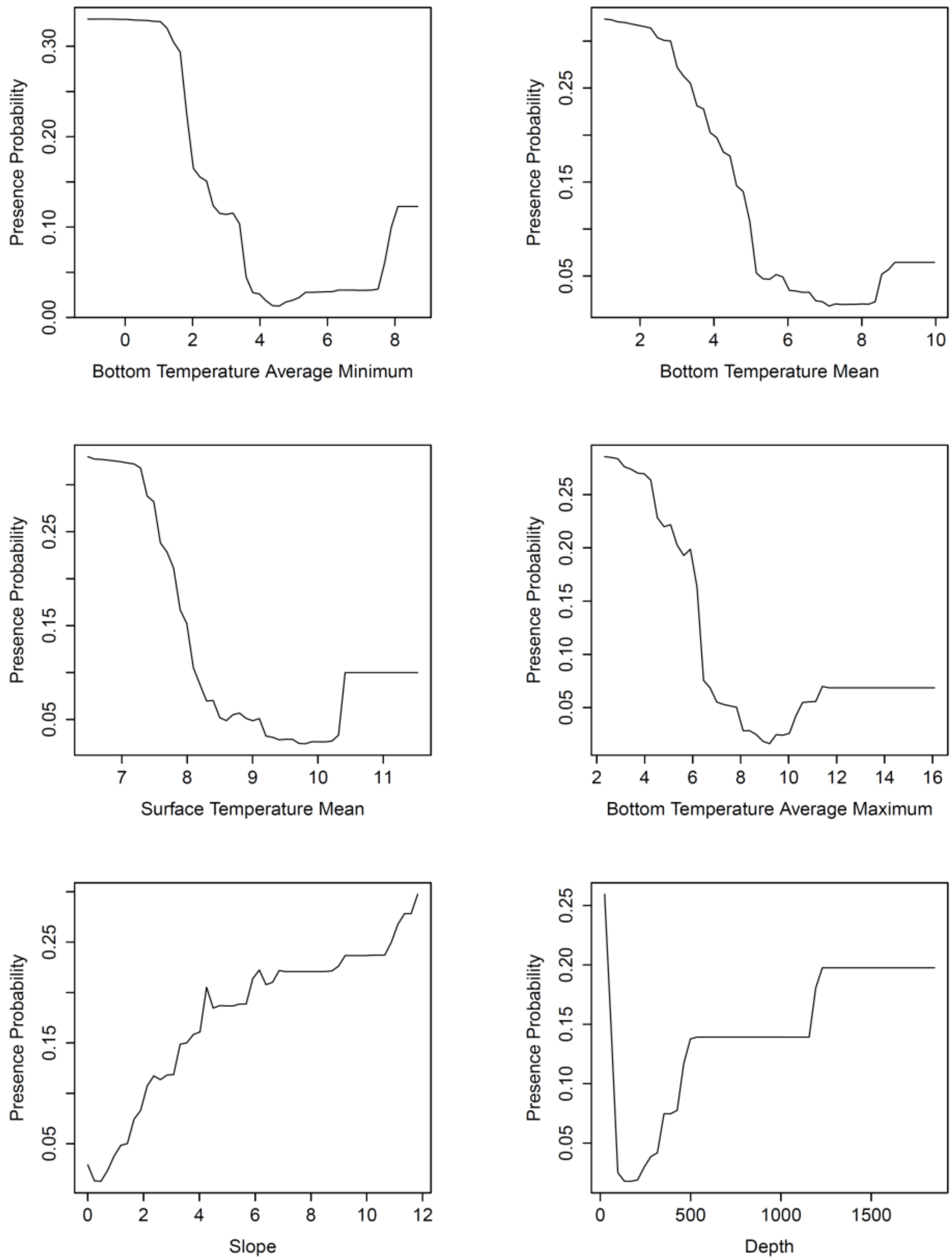

Figure 62. Partial dependence plots of the top six predictors from the random forest model of soft coral presence and absence data collected within the Maritimes Region, ordered left to right from the top. Presence probability is shown on the *y*-axis.

Addition of *In Situ* Benthic Imagery and DFO Scallop Stock Assessment Records

Given the relatively low number of soft coral presence records from the DFO multispecies surveys, and their uneven spatial distribution and under-representation along the slope and on the western Scotian Shelf where they are known to occur (see Figure 56), the DFO multispecies trawl survey data were augmented with additional *in situ* benthic imagery records and DFO scallop stock assessment survey records collected in the Maritimes Region between 1965 and 2014 (see Tables 18 and 19). This data consisted of a total of 177 *in situ* benthic imagery records collected between 1965 and 2014, and 16 presences and 232 absences collected in 1997 and 2007 from the DFO stock assessment survey. By including these data, new taxa were introduced to the model, which renders this model not directly comparable to the former model unless these taxa are captured within the 'soft coral unidentified' taxon from the multispecies trawl surveys. When added to the DFO multispecies trawl survey presence and absence records, the final dataset consisted of 430 presence and 2578 absences after filtering the data so that only one presence or absence record occurs in a single cell. This combined dataset was remodelled using an unbalanced design and a threshold equal to species prevalence (0.14).

Table 18. Taxonomic composition of the soft coral taxa from DFO *in situ* benthic camera and video surveys conducted in the Maritimes Region between 1965 and 2014, and the DFO scallop dredge surveys from 1997 and 2007. These records were added to the DFO multispecies trawl survey records and remodelled.

Survey	**Taxon**	**Number of presences**
DFO *in situ* benthic imagery surveys	*Anthomastus grandiflorus*	40
	Anthomastus sp.	4
	Anthomastus spp.	6
	Clavularia spp.	19
	Duva florida	1
	Heteropolypus sol	6
	Nephtheidae spp.	101
DFO scallop dredge survey	*Gersemia rubiformis*	16P/232A

Table 19. Number of soft coral *in situ* benthic imagery and DFO scallop dredge survey records collected between 1965 and 2014 in the Maritimes Region. Absences from the DFO scallop stock assessment survey were included. Note that there were no soft coral presence records from the 2007 scallop survey.

Year	Gear	Total number of presences
1965	NRCan Drop Camera	1
1997	Scallop Dredge	16P/125A
1997	Campod	3
1999	Campod	4
2000	Campod	3
2000	NRCan Drop Camera	4
2001	ROPOS	3
2001	Campod	16
2002	Campod	6
2003	NRCan Drop Camera	1
2003	Campod	11
2004	NRCan Drop Camera	1
2005	Campod	19
2006	ROPOS	8
2006	DSIS	3
2007	Scallop Dredge	107A
2007	ROPOS	29
2008	Campod	49
2008	NRCan Drop Camera	6
2011	Campod	8
2014	Towed Camera	2

The accuracy measures for this model are shown in Table 20. The average AUC computed from 10-fold cross validation was 0.878 ± 0.035 SD, which was slightly higher than the former model (AUC = 0.830). Class error, and sensitivity and specificity of this model were comparable to that of the former model.

The additional presence records expanded the area of high soft coral presence probability in the region, particularly along the eastern slope and in the deep-water canyons (Figure 63). The Gully submarine canyon east of Sable Island and the Northeast Channel on the western Scotian Shelf showed much higher soft coral presence probability compared to the former model. These areas of higher probability corresponded well with the location of additional *in situ* benthic imagery observations. Interestingly, the scallop dredge records from Bay of Fundy did not result in a higher presence probability for that area, possibly due to the mix of presence and absence records in the area. The area of extrapolation along the slope off eastern Scotian Shelf is reduced with the addition of science survey records there (Figure 64). Figure 65 depicts the predicted distribution of soft corals based on a prevalence threshold of 0.14. In this map, all presence probability values greater than 0.14 were classified as presence, while values less than 0.14 were classed as absence. The eastern portion of the study extent and the majority of the Scotian Slope was predicted as suitable habitat for soft corals.

In contrast to the former model, Slope and Depth, two non-interpolated variables, were the top two predictors in this model (Figure 66). The high importance of these two variables in this model are likely from the addition of the science survey records which are concentrated in deeper waters along the slope. These variables were followed in importance by Bottom and Surface Temperature Mean and the other variables in the model. Partial dependence plots of the top 6 environmental variables are shown in Figure 67. Presence probability was highest at Slope values of 5˚ and greater. Like in the former model, presence probability was highest at the shallowest depths, decreased, then increased and plateaued. However, in this model, presence probability increased and stabilized more rapidly in the deeper depth range, likely due to the inclusion of more soft coral records from deeper depths.

Table 20. Accuracy measures and confusion matrix from 10-fold cross validation of a random forest model of presence and absence of soft corals within the Maritimes Region. Observ. = Observations; Sensit. = Sensitivity, Specif. = Specificity.

Model Fold	**AUC**	**Observ.**	**Predictions**		**Total n**	**Class error**	**Sensit.**	**Specif.**
1	0.873		**Absence**	**Presence**				
2	0.942	**Absence**	2009	569	2578	0.221	0.793	0.779
3	0.887	**Presence**	89	341	430	0.207		
4	0.890							
5	0.883							
6	0.886							
7	0.827							
8	0.879							
9	0.890							
10	0.816							
Mean	**0.878**							
SD	**0.035**							

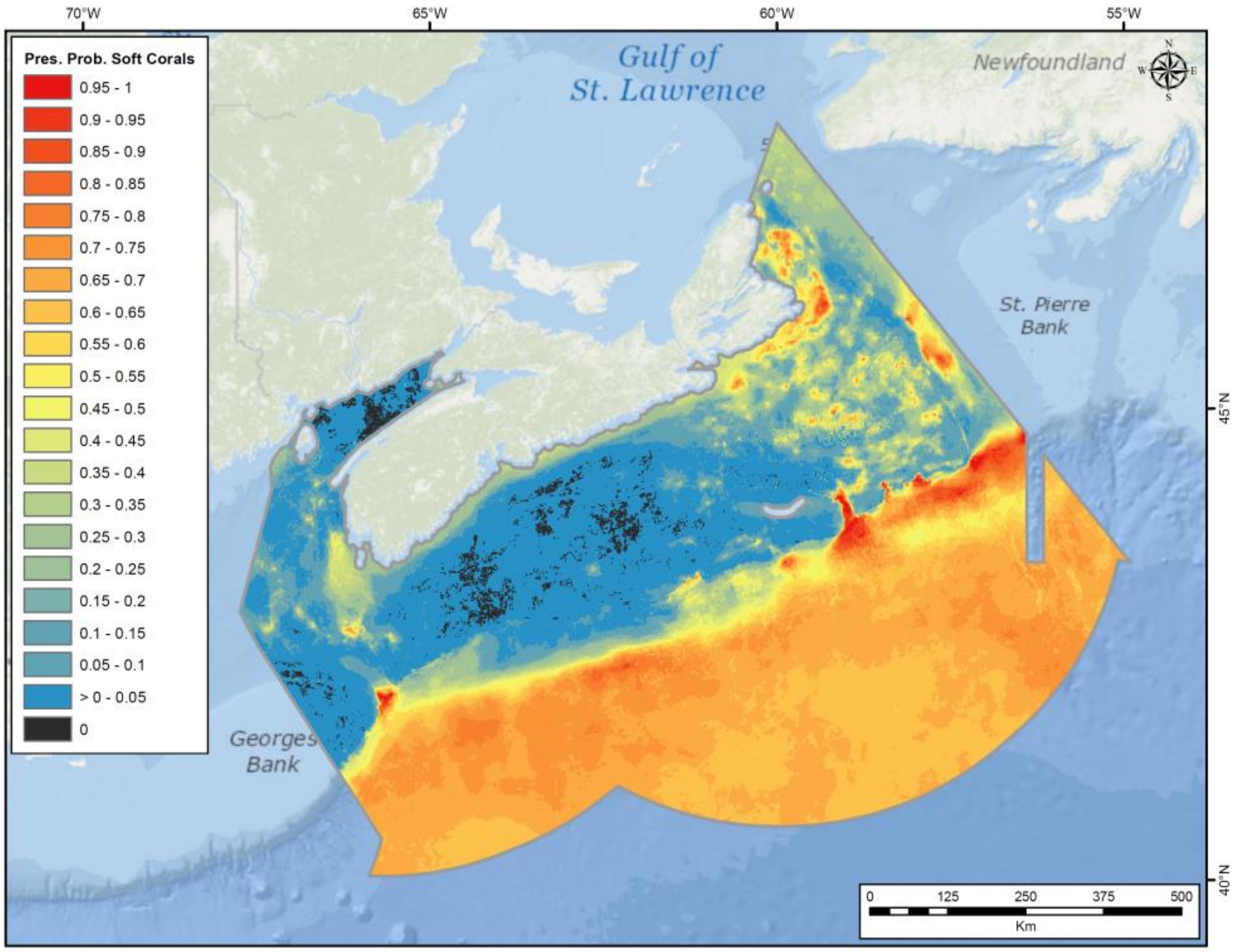

Figure 63. Predictions of presence probability (Pres. Prob.) of soft corals based on a random forest model on presence and absence soft coral catch data collected from DFO multispecies trawl surveys and DFO scallop dredge surveys, and *in situ* benthic imagery observations of soft corals collected from various surveys conducted in the Maritimes Region between 1965 and 2014.

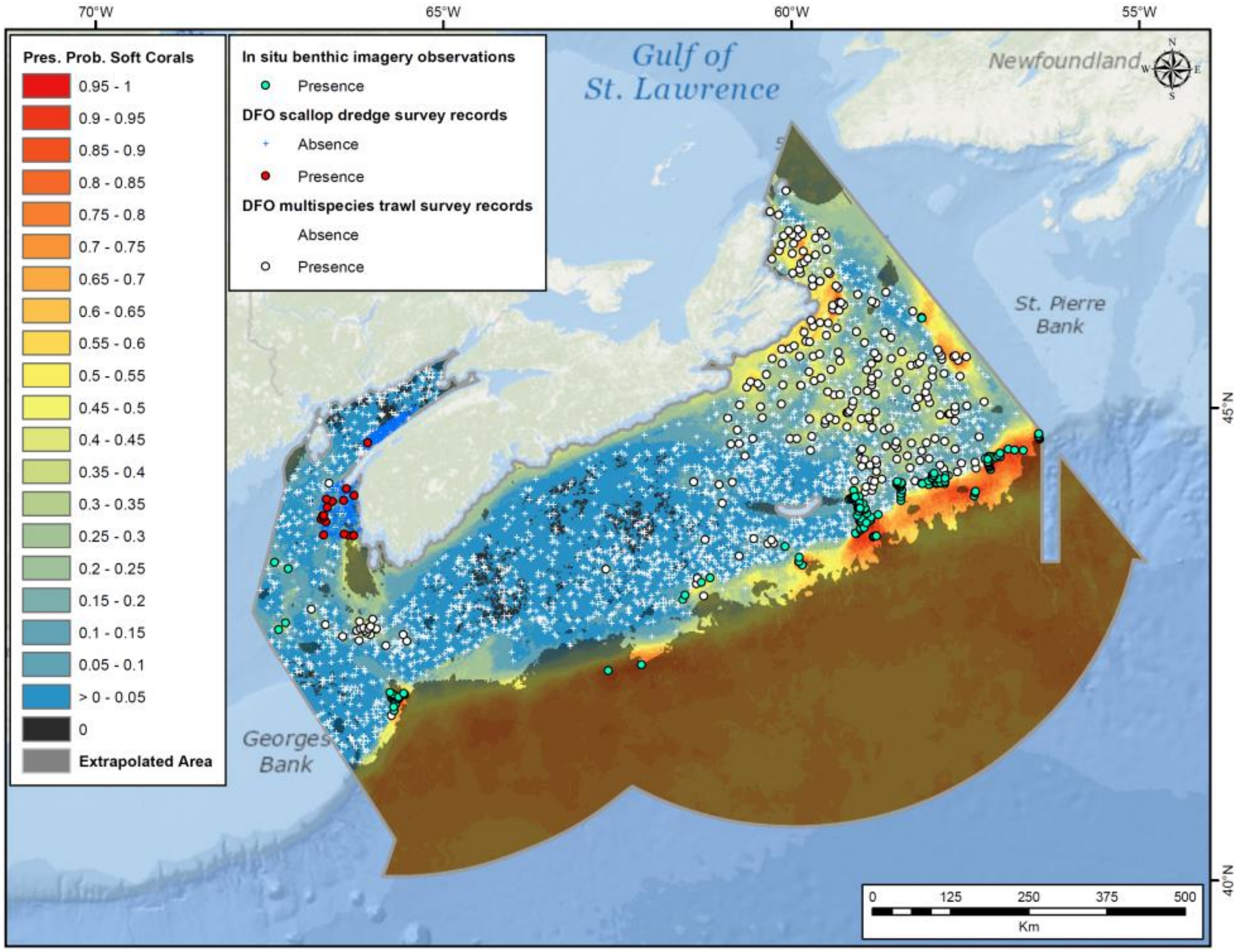

Figure 64. Presence and absence observations and predictions of presence probability (Pres. Prob.) of soft corals based on a random forest model on presence and absence soft coral catch data collected from DFO multispecies trawl surveys and DFO scallop dredge surveys, and *in situ* benthic imagery observations of soft corals collected from various surveys conducted in the Maritimes Region between 1965 and 2014. Also shown are areas of model extrapolation.

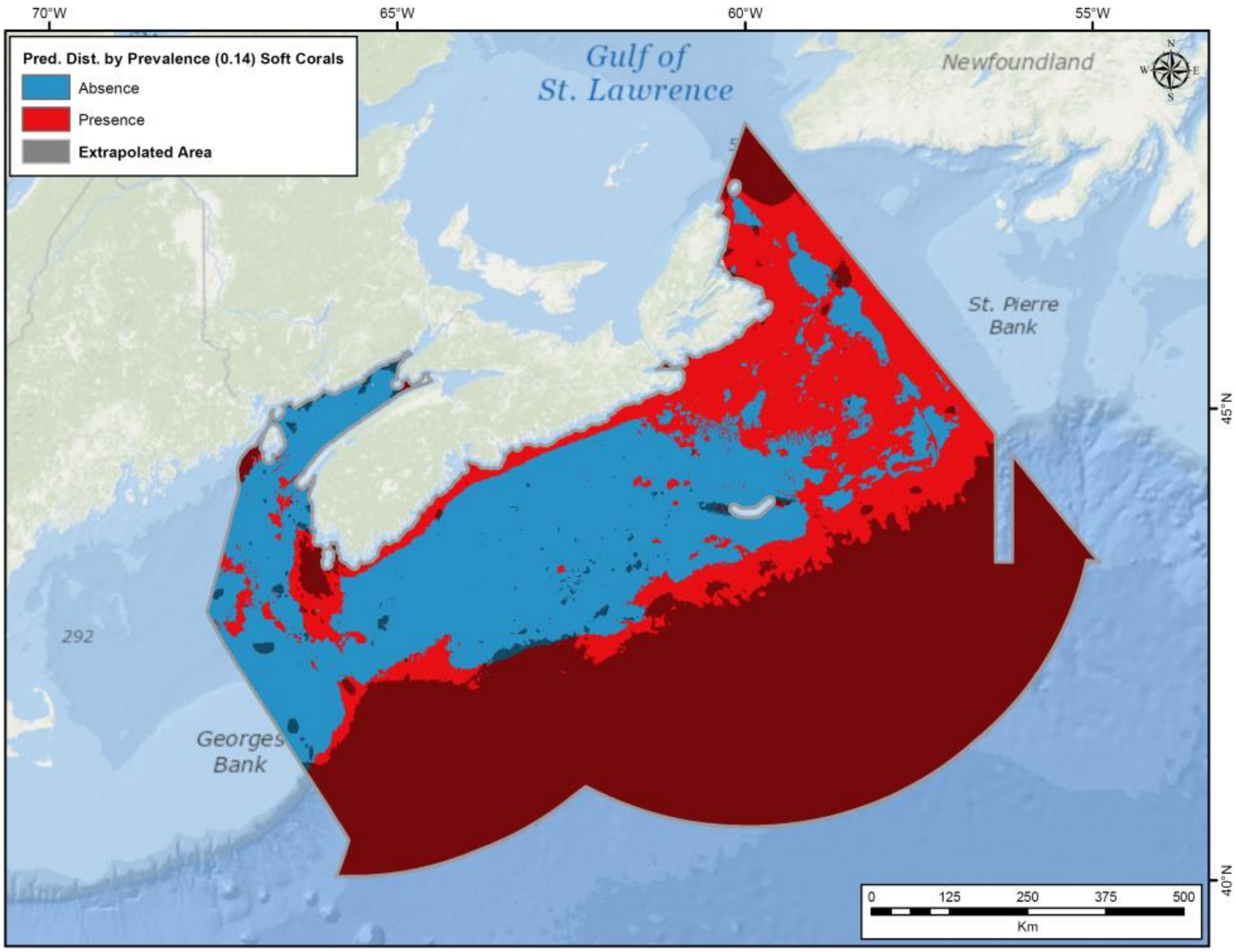

Figure 65. Predicted distribution (Pred. Dist.) of soft corals in the Maritimes Region based on the prevalence threshold (0.14) of soft coral presence and absence data used in the model. Also shown are the areas of model extrapolation (grey polygon may appear red or blue).

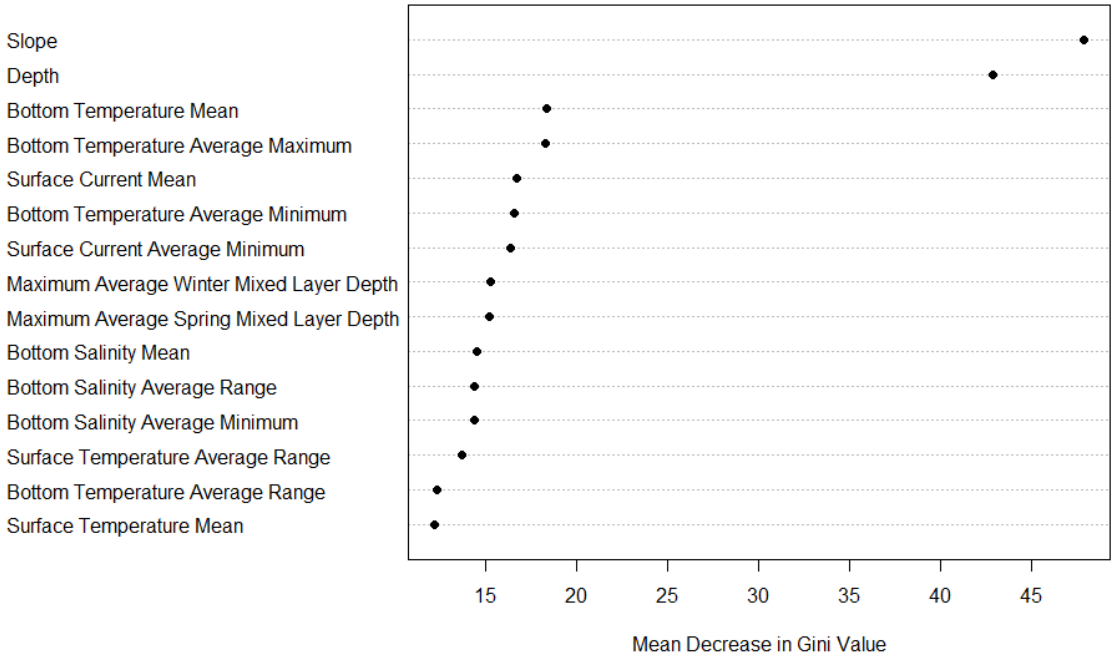

Figure 66. Importance of the top 15 predictor variables measured as the Mean Decrease in Gini value of the random forest model on soft coral presence and absence data from DFO multispecies trawl surveys, DFO scallop dredge surveys, and *in situ* benthic imagery observations collected from various surveys conducted in the Maritimes Region between 1965 and 2014. The higher the Mean Gini value the more important the variable is for predicting the response data.

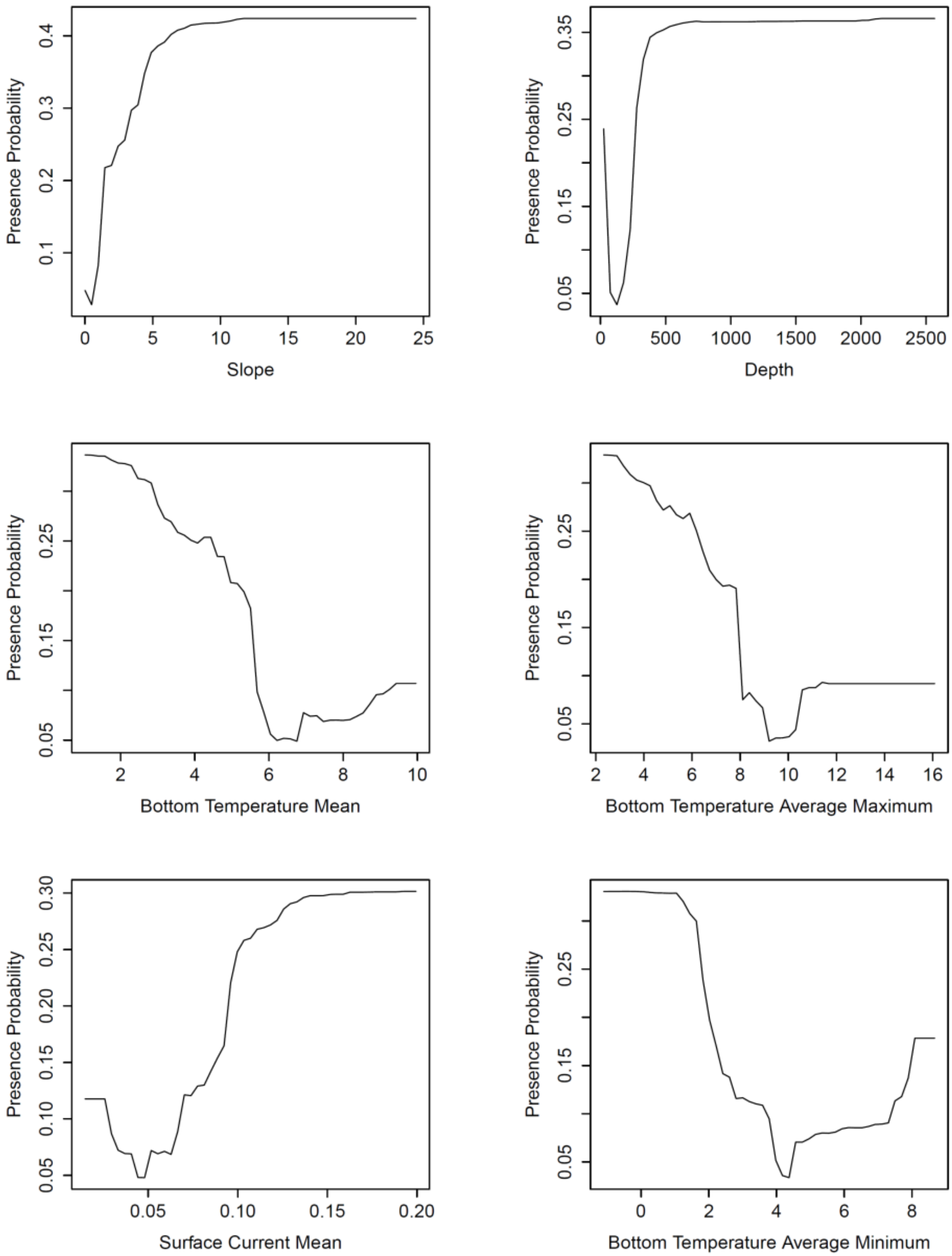

Figure 67. Partial dependence plots of the top six predictors from the random forest model of soft coral presence and absence observations collected from DFO multispecies trawl surveys, DFO scallop dredge surveys and *in situ* benthic imagery observations collected within the Maritimes Region between 1965 and 2014, ordered left to right from the top. Presence probability is shown on the *y*-axis.

Model Selection

The random forest model on the combined dataset consisting of records from the DFO multispecies trawl surveys, DFO *in situ* benthic camera/video surveys, and DFO scallop stock assessment surveys was chosen as the best predictor of the distribution of soft corals in the Maritimes Region. Although the model using only the DFO trawl survey records performed well, it did not extend its prediction of suitable habitat to areas where soft corals are known to occur. The second model including the records from Bay of Fundy and the slope is therefore a more accurate depiction of the distribution of this taxonomic group likely due to the increase in the environmental range represented in the model with the additional data.

Validation of Selected Model Using Independent Data

Additional soft coral records from the NOAA Deep-Sea Coral Data Portal were used for model validation of the selected model containing the additional scallop survey and science records (see Figure 68). Fifty-two NOAA soft coral records remained after they were filtered to a single record per raster cell. Of these 52 records, most (38, or 73%) fell into areas predicted as suitable habitat by the model. Many of these occurred in a narrow band along the slope or in the area of presence on Banquereau Bank that has many invaginations in the prevalence surface. Fourteen records (27%) were predicted as absence. Many of these occurred in the large area of unsuitable habitat on the central Scotian Shelf, but some occurred along the border of suitable habitat on the slope.

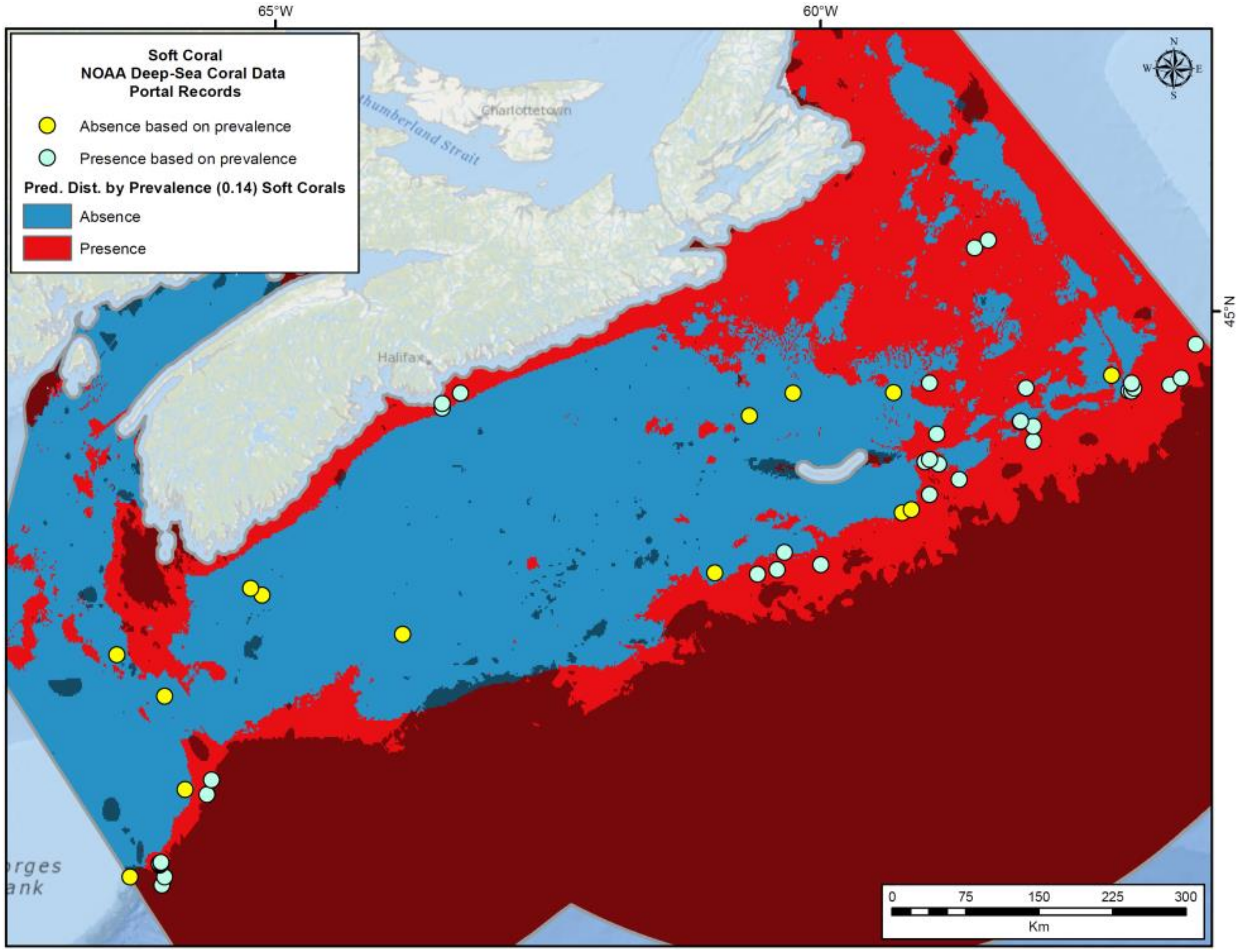

Figure 68. Validation of the prevalence surface from the selected random forest model using soft coral records from the NOAA Deep-Sea Coral Data Portal.

Prediction of Soft Coral Biomass Using Random Forest

Due to time constraints rocks are not always separated from the soft coral prior to recording their biomass on the DFO multispecies surveys, and so their biomass distribution may be more of a reflection of the size of the attachment substrate than of the corals themselves. Nevertheless, soft coral biomass distribution was predicted using random forest. The accuracy measures of the model are presented in Table 21. The highest R^2 value was 0.148, while the average was 0.021 ± 0.046 SD, indicating poor model performance. The average Normalized Root-Mean-Square Error was 0.014 ± 0.018 SD. The high standard deviation indicates high variability between model folds. The percent variance explained was -2.89 ± 0.82 SD.

Figures 69 and 70 show the predicted biomass surface of soft corals. Most of the study area was predicted to have a low biomass (> 0 – 0.12 kg) of soft corals, with the majority of the western Scotian Shelf predicted to have zero soft coral biomass. A small pocket of high biomass was predicted to occur on Banquereau Bank. This area corresponded to the highest mean catch of soft corals (see Figure 70). Several small pockets of moderate biomass were predicted to occur along the eastern Scotian Slope.

The top 15 most important environmental variables for predicting soft coral biomass are shown in Figure 71. Like the presence-absence random forest model on the DFO multispecies trawl records, Bottom Temperature Average Minimum was the most important variable in the biomass model. This was followed more distantly by Annual Chlorophyll *a* Minimum and two fall primary production variables. Generally, chlorophyll *a* and primary production variables ranked high in this model. The partial dependence of soft coral biomass on the top six most important predictor variables is shown in Figure 72. Predicted biomass was highest at the lowest Bottom Temperature Average Minimum (< 2˚C) and Annual Chlorophyll *a* Minimum (< 1 mg C m^{-2} day^{-1}) values.

Table 21. Accuracy measures from 10-fold cross validation of a random forest model of average soft coral biomass (kg) per grid cell recorded from DFO multispecies trawl surveys in the Maritimes Region. RMSE = Root-Mean-Square Error; NRMSE = Normalized Root-Mean-Square Error.

Model Fold	**R^2**	**RMSE**	**NRMSE**	**Percent (%) variance explained**
1	0.004	0.029	0.005	-3.34
2	0.002	0.031	0.005	-2.47
3	0.005	0.037	0.006	-2.26
4	0.148	0.016	0.003	-2.91
5	0.000	0.355	0.062	-4.58
6	0.004	0.031	0.005	-2.98
7	0.002	0.103	0.018	-2.27
8	0.001	0.039	0.007	-3.81
9	0.044	0.035	0.006	-2.26
10	0.001	0.102	0.018	-2.00
Mean	**0.021**	**0.078**	**0.014**	**-2.89**
SD	**0.046**	**0.102**	**0.018**	**0.82**

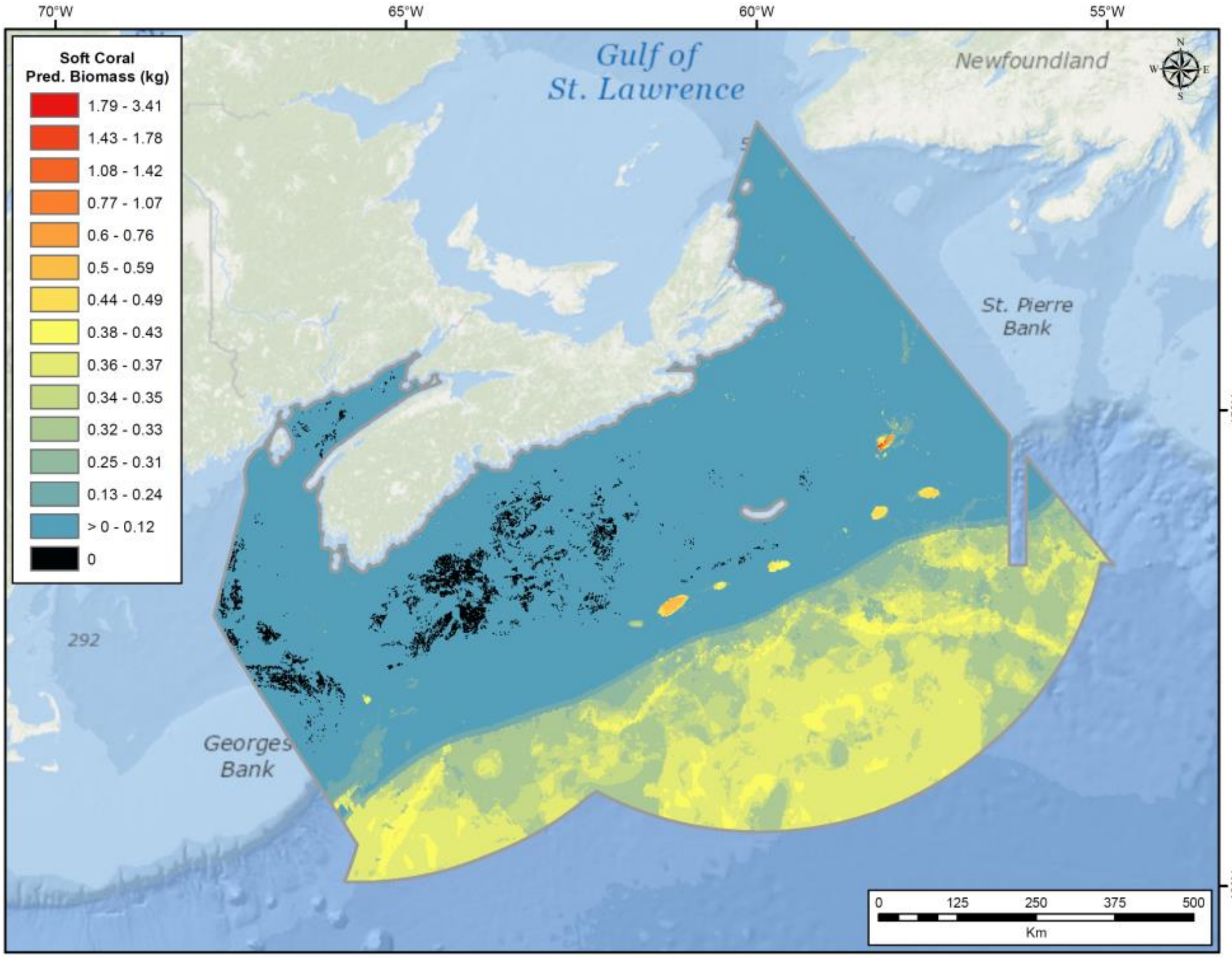

Figure 69. Predictions of biomass (kg) of soft corals from catch data recorded in DFO multispecies trawl surveys conducted in the Maritimes Region between 2003 and 2014.

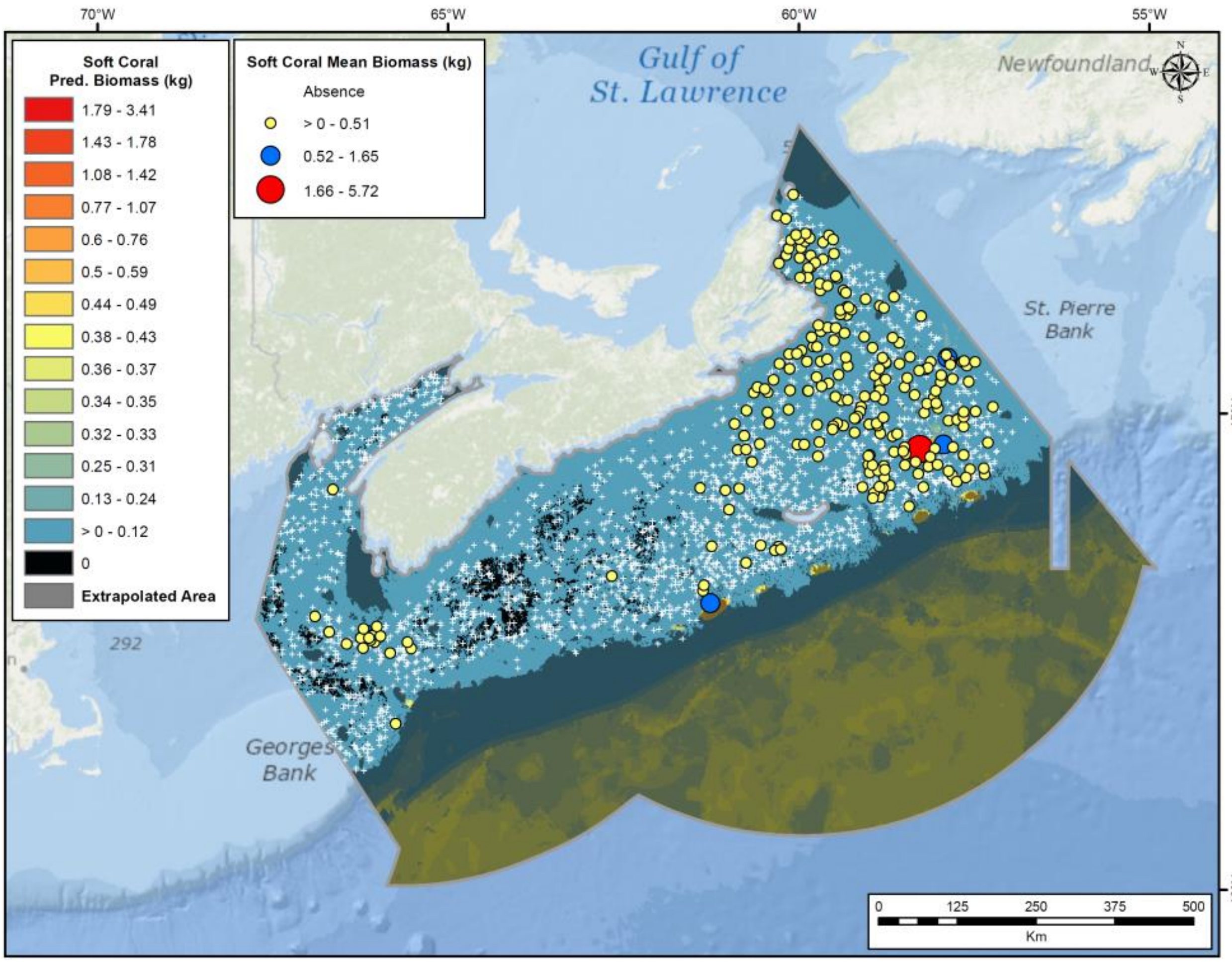

Figure 70. Predictions of biomass (kg) of soft corals from catch data recorded in DFO multispecies trawl surveys conducted in the Maritimes Region between 2003 and 2014. Also shown are the mean biomass values per grid cell and areas of model extrapolation.

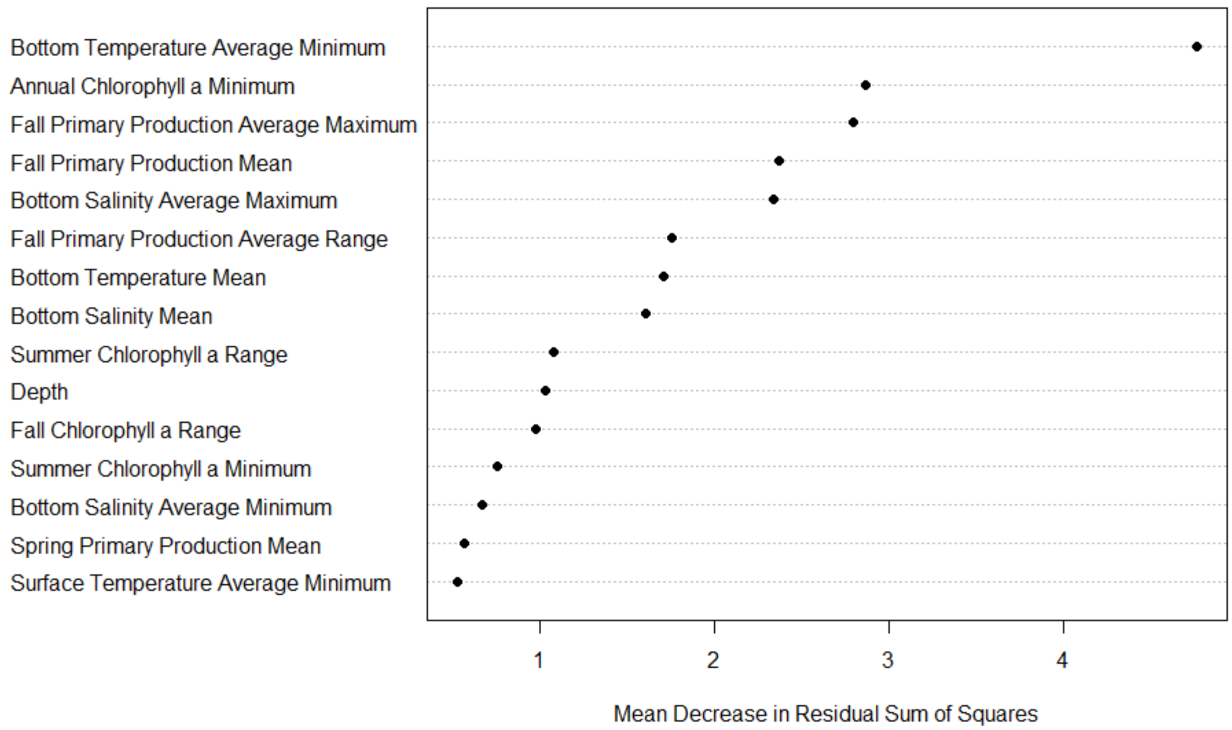

Figure 71. Importance of the top 15 predictor variables measured as the Mean Decrease in Residual Sum of Squares of the regression random forest model on soft coral biomass data from the Maritimes Region. The higher the Mean Decrease in Residual Sum of Squares, the more important the variable is for predicting the response data.

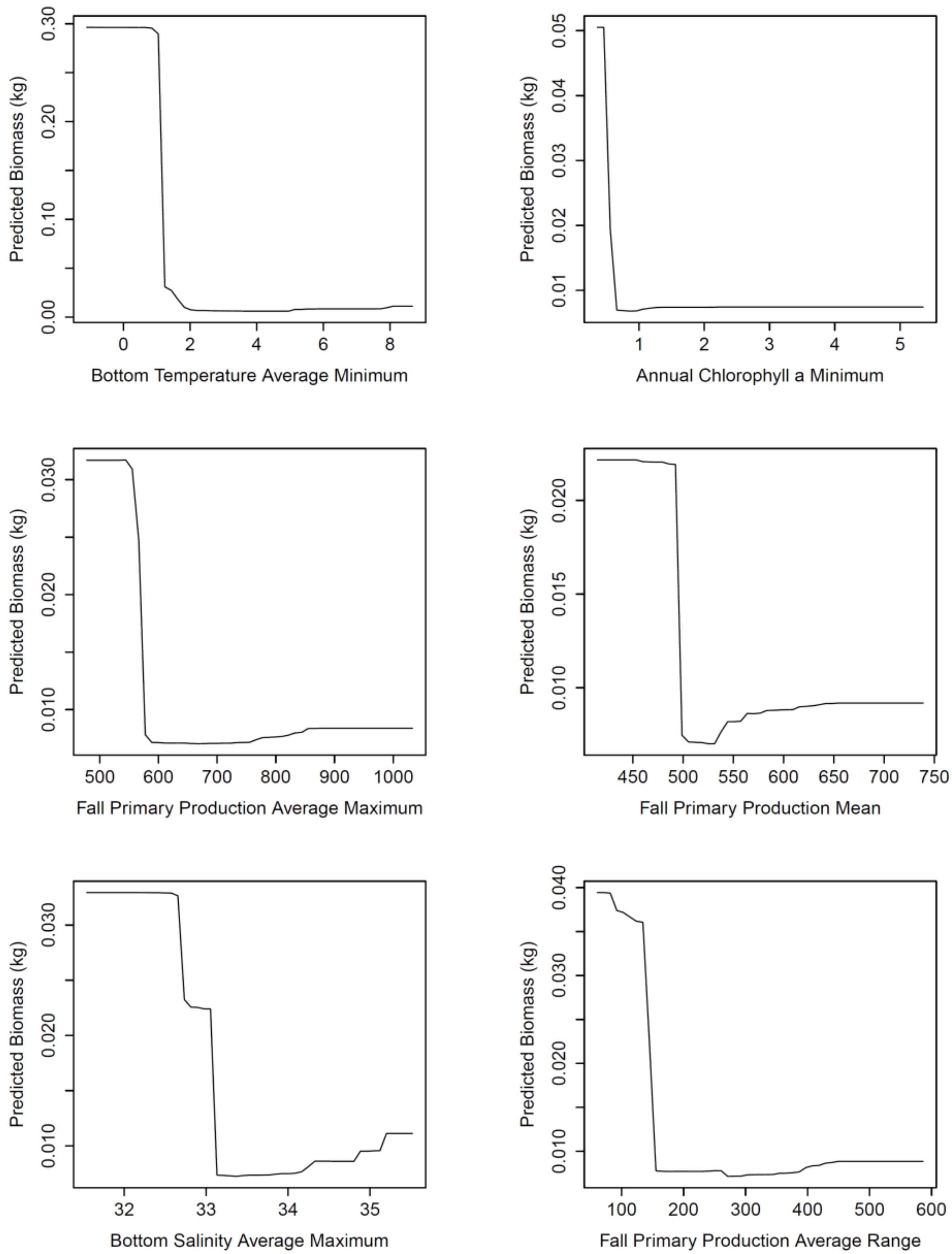

Figure 72. Partial dependence plots of the top six predictors from the random forest model of soft coral biomass data collected within the Maritimes Region, ordered left to right from the top. Predicted biomass (kg) is shown on the *y*-axis.

Kernel Density Analysis of Soft Coral Catch Data

Kernel density analysis of the soft coral biomass data showed concentrations on the eastern Scotian Shelf, with smaller concentrations in the Northeast Channel (Figure 73), but this may represent differences in catchability on different substrates given that this is not a true biomass and in some sets rocks may have been included in the weight (although only three catches are greater than 1 kg; Table A4, Appendix A). The KDE search radius for this group was 19.9 km. Polygons encompassing each of 13 catch thresholds showed that the greatest increase in area between successive weights, after the initial mapping out of the habitat, occurred between 0.05 kg and 0.025 kg, where the area increased by 166% (Table 22, Figure 74). There were 45 data points contributing to this expansion making it very robust (Table 22, Figure 75). Consequently 0.05 kg was considered as the threshold delineating soft coral gardens. These areas occupy 10,966 km^2 of seabed. Figure 76 shows the final KDE significant area polygons, the location of significant and non-significant catches, and null records of soft corals from the DFO multispecies trawl survey.

There was good spatial congruence between the location of the KDE polygons and areas predicted as suitable habitat by the random forest model (Figure 77). The large area of suitable habitat along the slope was not encompassed by KDE polygons. This was due to the lack of multispecies trawl survey records from the slope and the inclusion of targeted science survey records in the random forest model. There were small invaginations of unsuitable habitat within the KDE polygons that could be considered in a trimming process in the future (see Kenchington et al., 2016a).

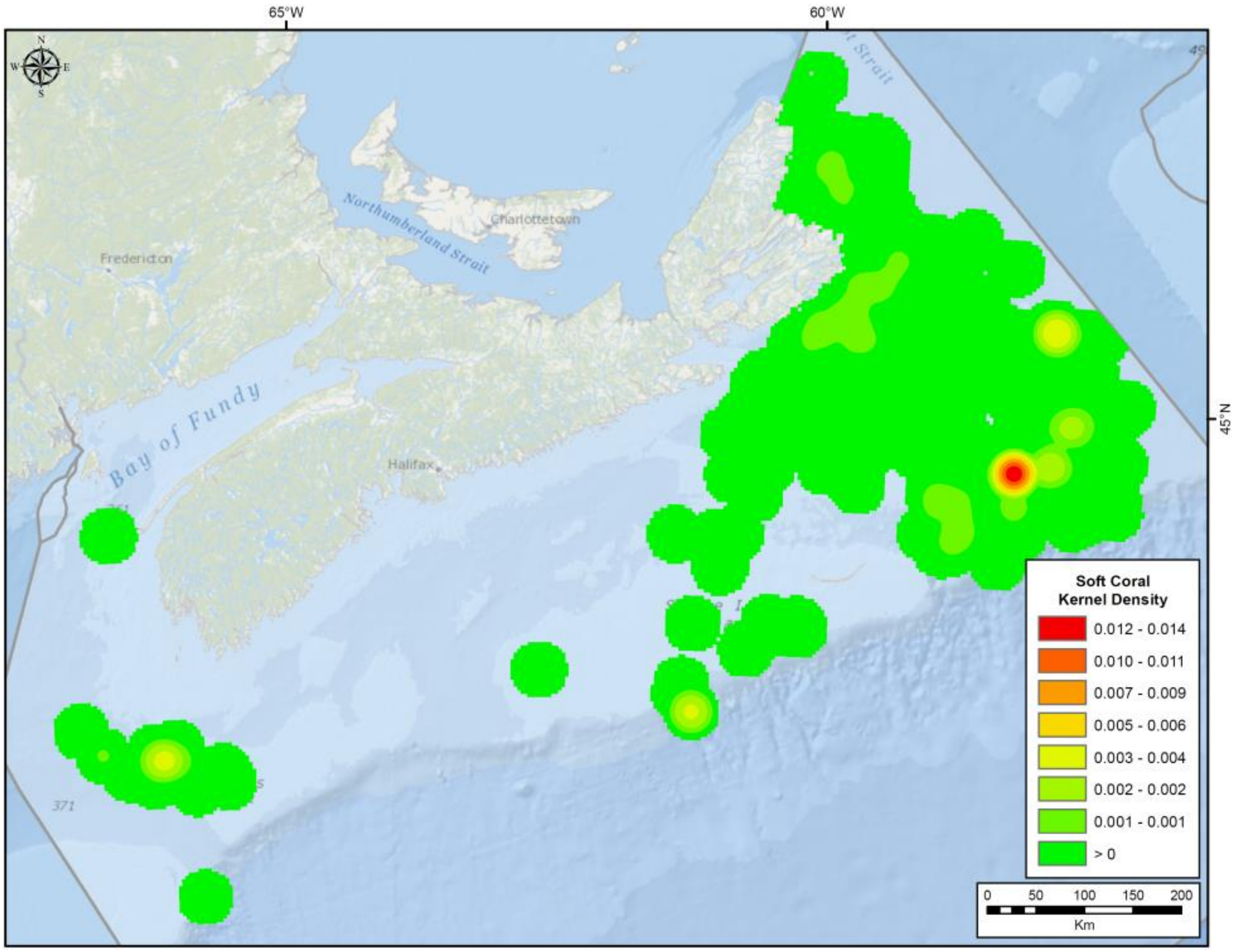

Figure 73. Kernel density analysis of soft coral catches in the Maritimes Region.

Table 22. Summary characteristics for thirteen soft coral catch weight thresholds. The threshold corresponding to the habitat delineation is indicated by shading.

Number of Points in Polygon	Area (km^2)	Catch Threshold (kg)	Percent Change in Area (%)
3	4	1	272.61
5	16	.75	59.68
6	25	.50	417.52
7	131	.25	4118.18
26	5533	.10	7.36
33	5940	.075	84.61
55	10966	.050	165.66
100	29131	.025	70.97
160	49806	.010	1.94
172	50771	.0075	13.61
196	57680	.0050	7.87
219	62222	.0025	0.00
236	62222	.0010	

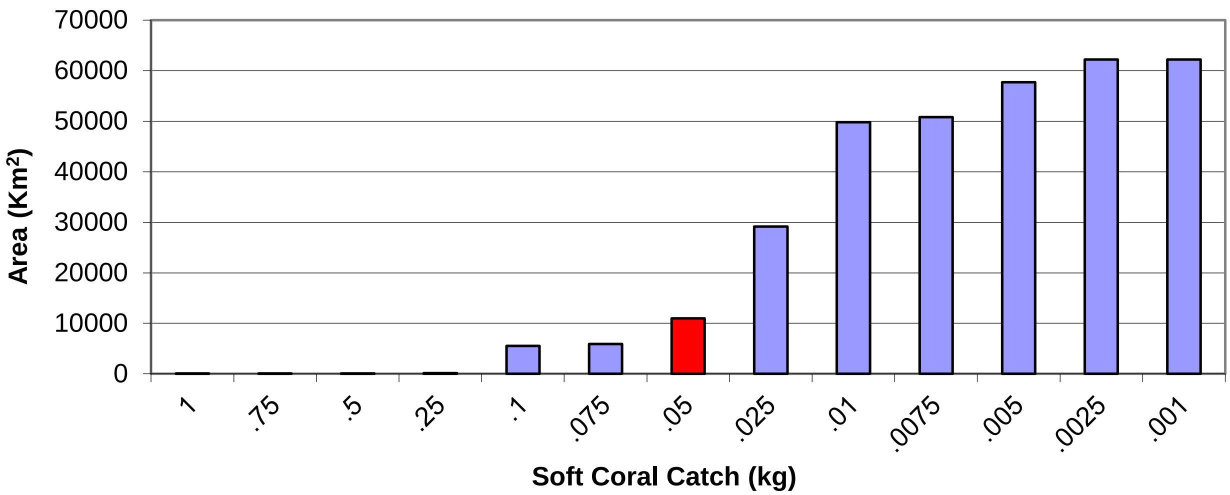

Figure 74. Area occupied by polygons encompassing catches of soft corals greater than or equal to each of 13 catch thresholds (1 kg to 0.001 kg). The red bar indicates the threshold separating out the soft coral gardens from the broader distribution.

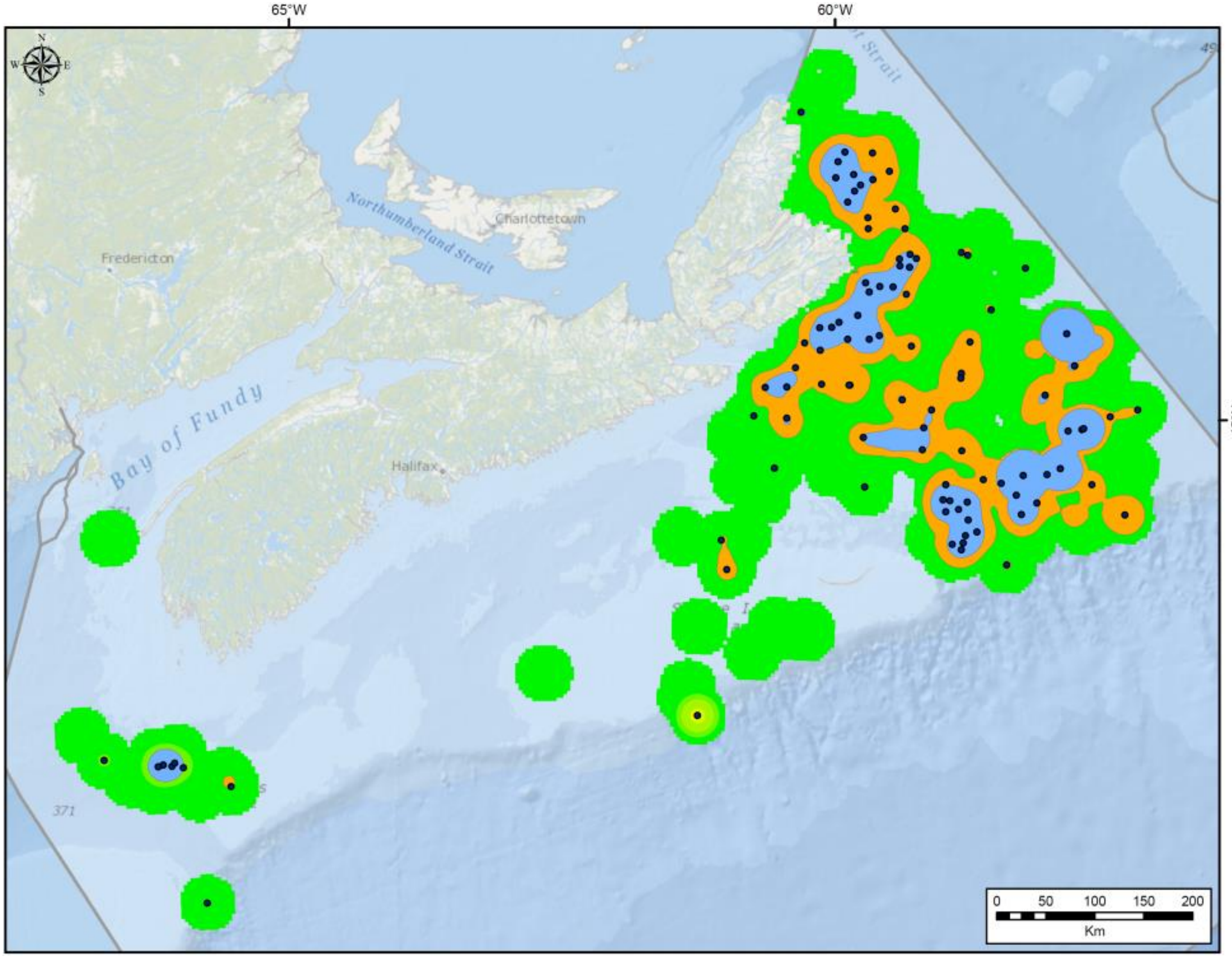

Figure 75. Polygons encompassing soft coral catch weights greater than or equal to 0.05 kg (blue) and 0.025 kg (orange) overlain on the kernel density surface. Tow locations associated with biomass values above these weight thresholds are also shown. The 0.05 kg threshold was used to delineate the habitat.

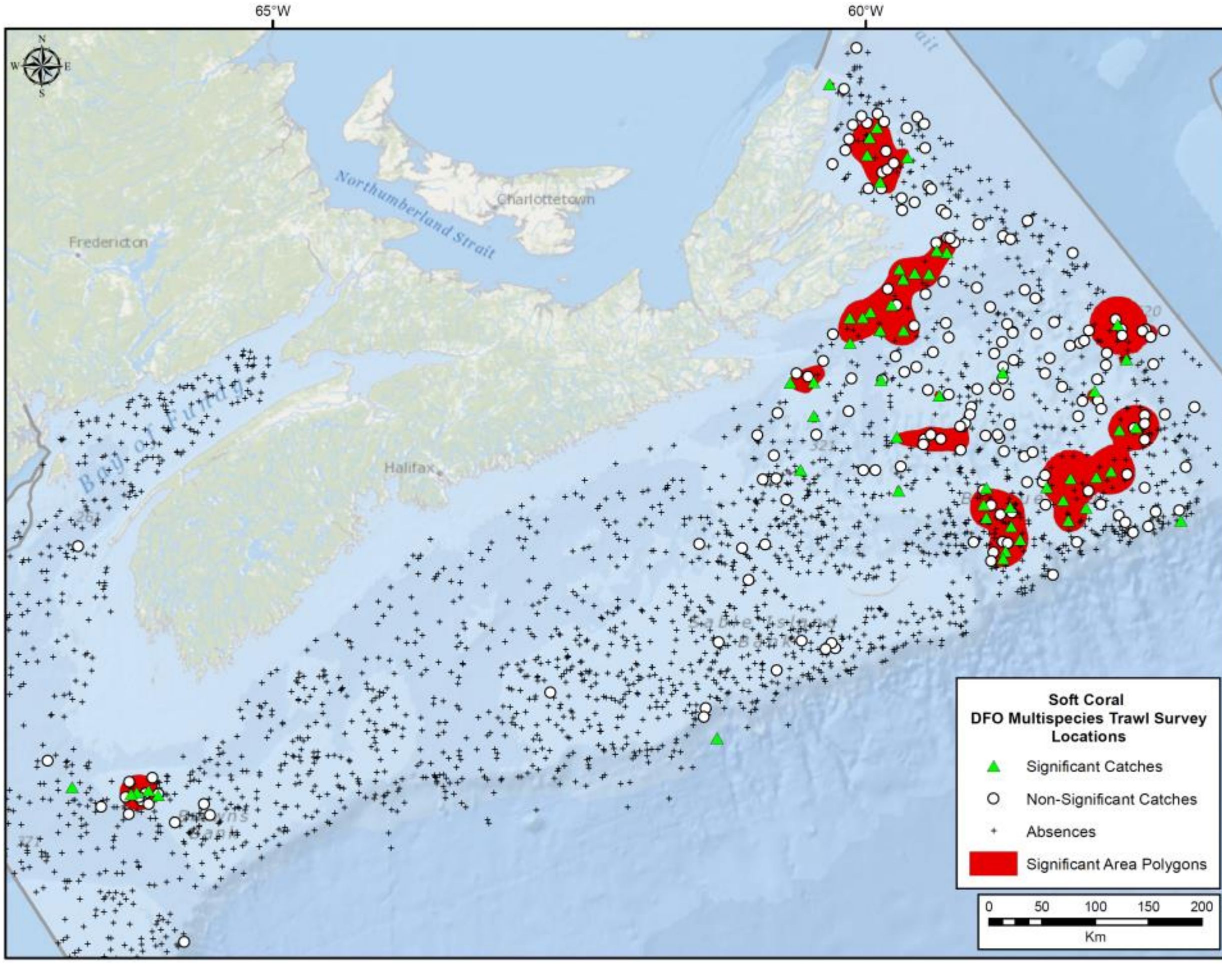

Figure 76. Location of polygons identifying significant soft coral concentrations in the Maritimes Region. Also shown are the significant and non-significant catches of soft coral and absence data (nulls) from the DFO multispecies trawl surveys.

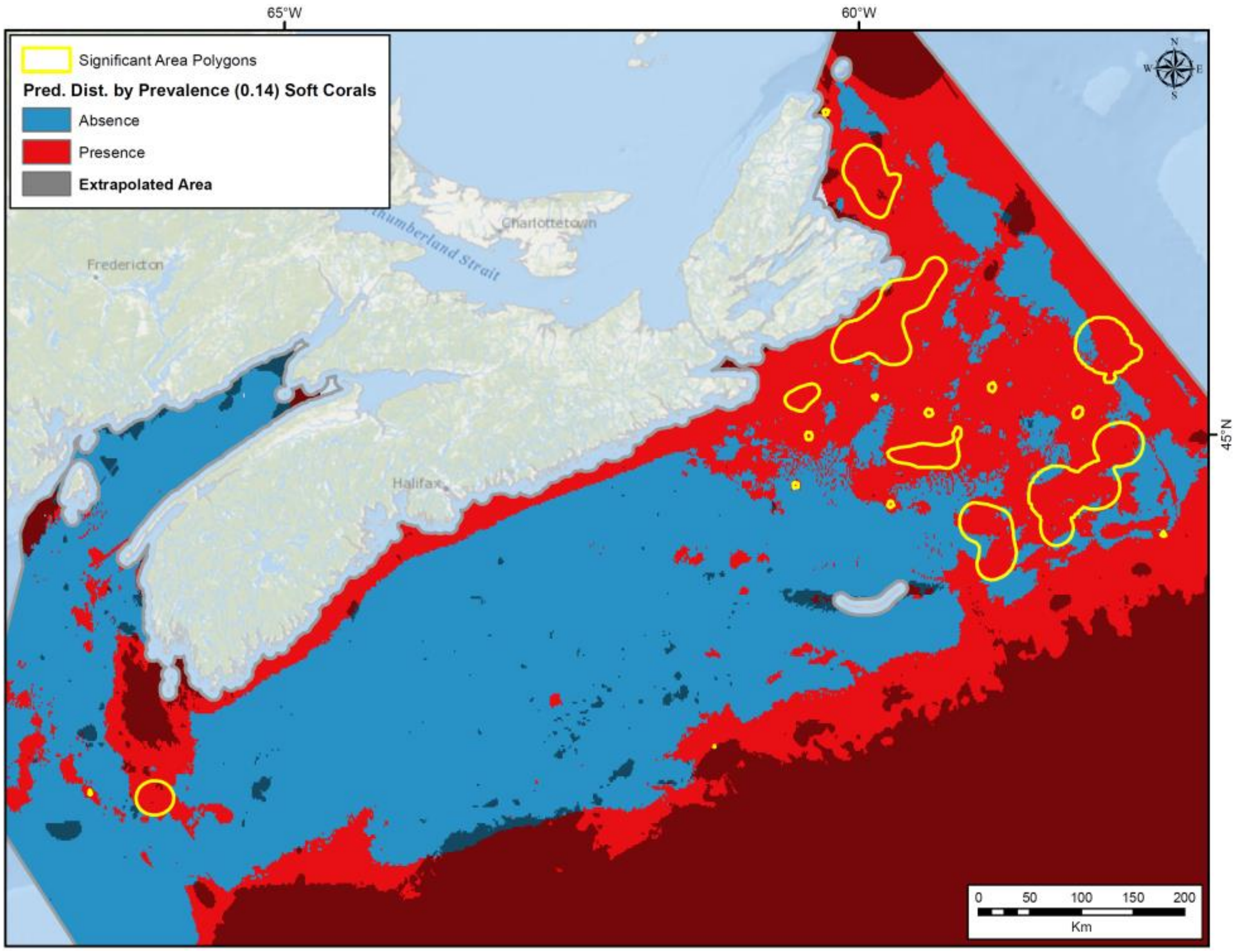

Figure 77. KDE polygons denoting significant concentrations of soft corals overlain on the predicted distribution by prevalence of soft corals generated from the selected random forest model.

Flabellum Cup Corals

Data Sources and Distribution

The functional role that stony cup corals play in the deep ocean ecosystem remains largely unknown. Stony cup corals can reach very high local densities along the Scotian Slope (up to 1150 individuals 100 m^{-2}; Buhl-Mortensen et al., 2007; Cogswell et al., 2009). However, they are not considered to form biogenic habitat (Etnoyer and Morgan, 2003) and therefore do not meet the EBSA criteria outlined in Kenchington (2014). Although DFO *in situ* surveys in the region have revealed a high diversity of cup coral species across different genera (Cogswell et al., 2009), only *Flabellum* cup corals are recorded in the DFO multispecies stock assessment surveys and are modelled in this report. Three *Flabellum* cup coral species have been definitively identified in the Maritimes Region: *Flabellum alabastrum*, *F. angulare*, and *F. macandrewi* (Gordon and Kenchington, 2007; Cogswell et al., 2009). The three species are difficult to distinguish from one another and as a result are often identified only to the genus level and grouped into a single taxon representing more than one species (denoted by 'spp.'; Buhl-Mortensen et al., 2007; Cogswell et al., 2009). In the NAFO region the overall depth distribution of all three species of *Flabellum* is 180 to 3200 m, with *F. macandrewi* occurring between 180 and 350 m, *F. alabastrum* between 200 and 2000 m, and *F. angulare* between 2200 and 3200 m depth (Kenchington et al., 2015). Figure 78 shows the distribution of known records of *Flabellum* in the Maritimes Region. In the DFO multispecies trawl surveys, *Flabellum* cup corals are identified to two taxa: *F. alabastrum* and *Flabellum* sp. Of the 53 records from the DFO multispecies trawl survey, only 3 were identified as *F. alabastrum* (see Table 2). It is likely that many of the records identified as *Flabellum* sp. are *F. alabastrum*. Therefore, we chose to combine the *Flabellum* sp. and *F. alabastrum* records into a single *Flabellum* spp. group for modelling, where the ‘spp.’ denotes that multiple species may be present. The DFO multispecies trawls survey records are concentrated along the Scotian Slope and in the southern Laurentian Channel (Figures 78 and 79).

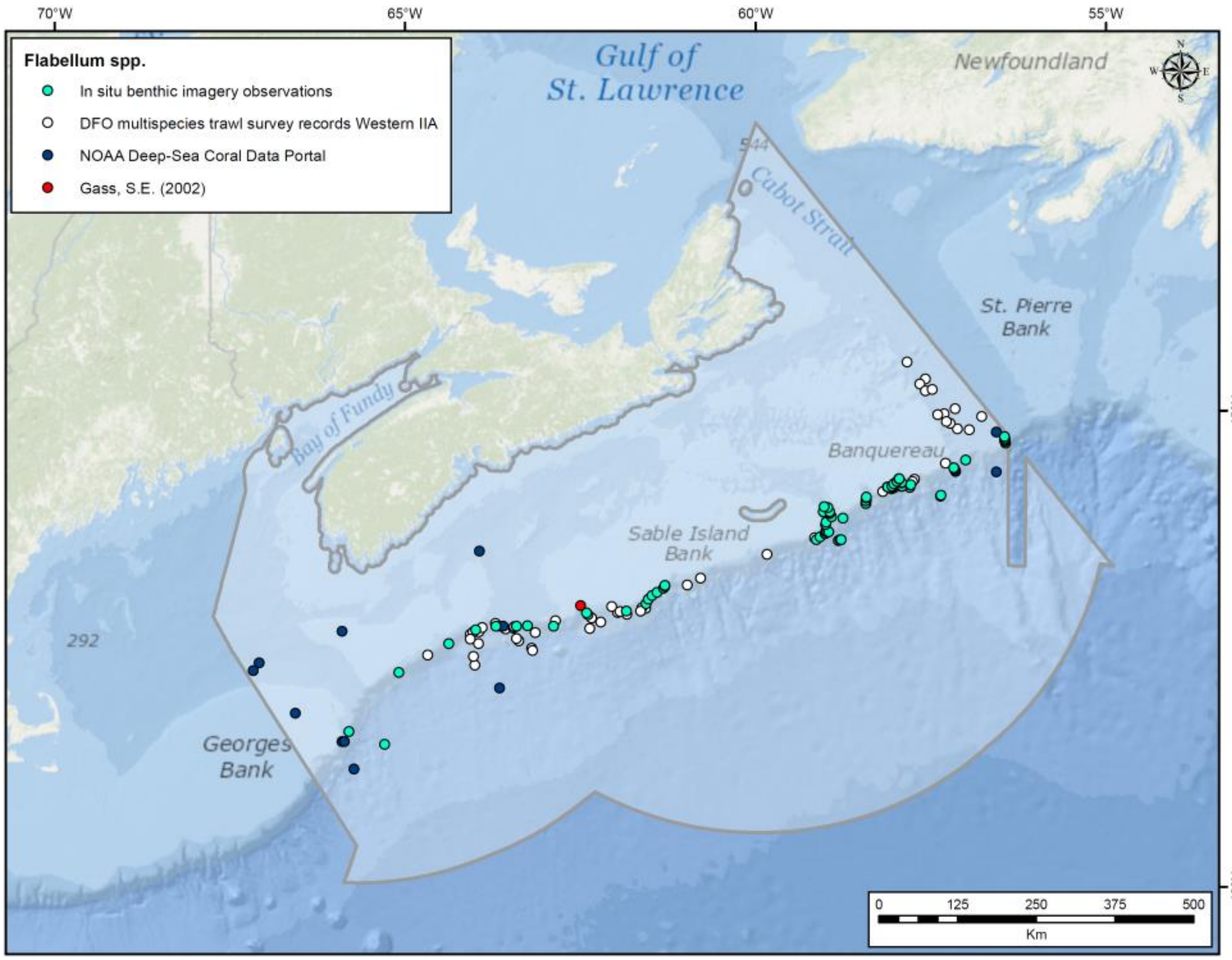

Figure 78. Available *Flabellum* spp. presence data in the Maritimes Region from the Gass (2002) report, DFO *in situ* benthic imagery observations, the NOAA Deep-Sea Coral Data Portal, and DFO multispecies trawl survey records.

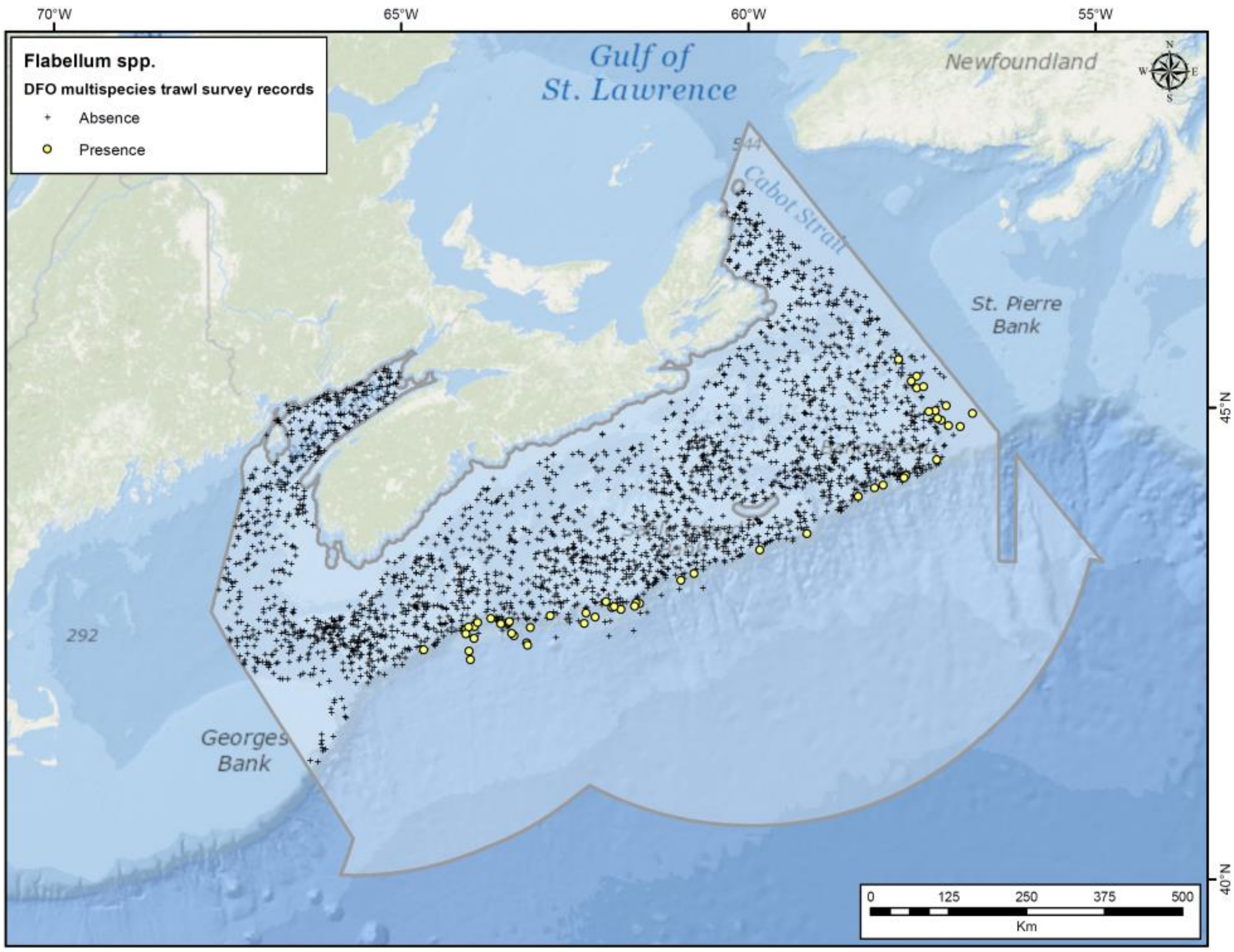

Figure 79. Presence and absence records of *Flabellum* spp. catch recorded from DFO multispecies trawl surveys conducted between 2003 and 2014 in the Maritimes Region.

Table 23. Number of presence and absence records of *Flabellum* catch in the Maritimes Region from DFO multispecies trawl surveys from 2003 to 2014.

Year	Total number of presences	Total number of absences
2003	2	214
2004	1	80
2005	4	277
2006	6	172
2007	5	254
2008	1	161
2009	5	192
2010	17	285
2011	2	253
2012	2	209
2013	3	210
2014	5	237
Total:	**53**	**2544**

Random Forest Model Results

The random forest model on the presence-absence data of *Flabellum* spp. had an excellent performance rating, with an average AUC of 0.965 ± 0.036 SD (Table 24). Class error for both the presence and absence classes was miniscule. Sensitivity and specificity of this model were high at 0.943 and 0.904, respectively.

The predicted presence probability surface of *Flabellum* spp. from the model is shown in Figure 80. Areas of high presence probability were limited in their spatial extent, with patches occurring in the southern Laurentian Channel and along the central/western Scotian Slope. Most of the shelf was predicted with zero *Flabellum* spp. presence probability. The areas of high presence probability corresponded well with the location of presence observations (Figure 81), although there appears to be slight under-prediction of the presence observations located along the eastern Scotian Slope.

Figure 82 depicts the predicted distribution of *Flabellum* spp. based on a prevalence threshold of 0.02. In this map, all presence probability values greater than 0.02 were classified as presence, while values less than 0.02 were classed as absence. The Laurentian Channel, slope, and canyons were predicted as suitable habitat (presence) for *Flabellum* spp, while the majority of the Scotian Shelf was predicted as unsuitable habitat (absence) for this taxonomic group.

Depth, a non-interpolated variable, was the top predictor in this model (Figure 83). This variable was followed more distantly by Bottom Salinity Average Range, Bottom Salinity Average Minimum, and the other variables in the model. Summer primary production and surface chlorophyll *a* variables also ranked high in this model. The partial dependence of *Flabellum* spp. on the top 6 most important variables is shown in Figure 84. *Flabellum* spp. presence probability was highest at 500 m depth and greater. Along the gradient in Bottom Salinity Average Range presence probability was highest at the lowest values (≤ 0.5). As values in this range were located along the slope (Beazley et al., in prep.) where *Flabellum* observations were found, this variable may be acting as a proxy for Slope (9th most important variable in the model). Presence probability was highest at the highest values of Bottom Salinity Average Minimum (> 34).

Table 24. Accuracy measures and confusion matrix from 10-fold cross validation of a random forest model of presence and absence of *Flabellum* spp. within the Maritimes Region. Observ. = Observations; Sensit. = Sensitivity, Specif. = Specificity.

Model Fold	**AUC**	**Observ.**	**Predictions**		**Total n**	**Class error**	**Sensit.**	**Specif.**
1	0.871		**Absence**	**Presence**				
2	0.950	**Absence**	2301	243	2544	0.096	0.943	0.904
3	0.995	**Presence**	3	50	53	0.057		
4	0.965							
5	0.978							
6	0.965							
7	0.988							
8	0.982							
9	0.971							
10	0.988							
Mean	**0.965**							
SD	**0.036**							

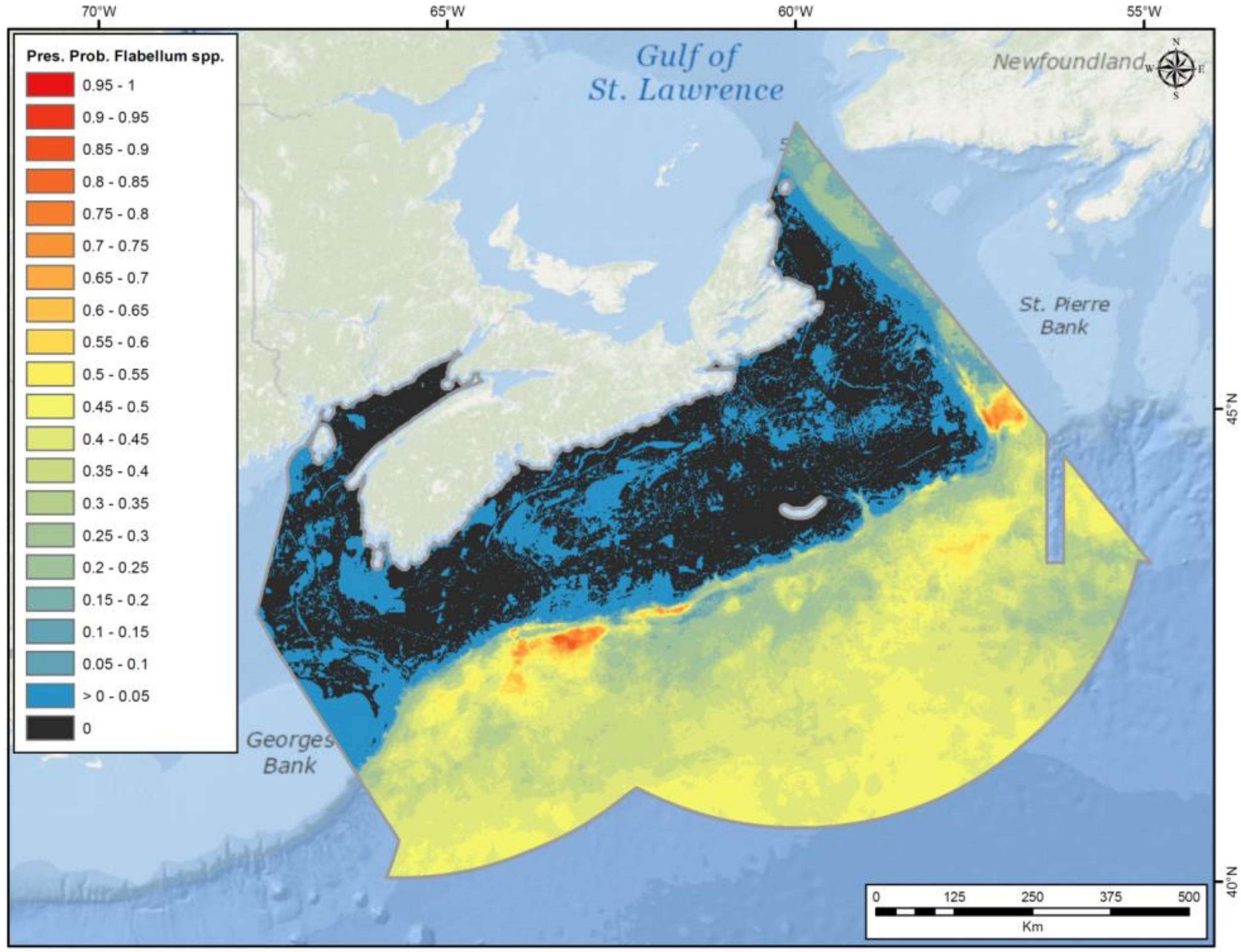

Figure 80. Predictions of presence probability (Pres. Prob.) of *Flabellum* spp. based on a random forest model on *Flabellum* spp. presence and absence data collected from DFO multispecies trawl surveys conducted in the Maritimes Region between 2003 and 2014.

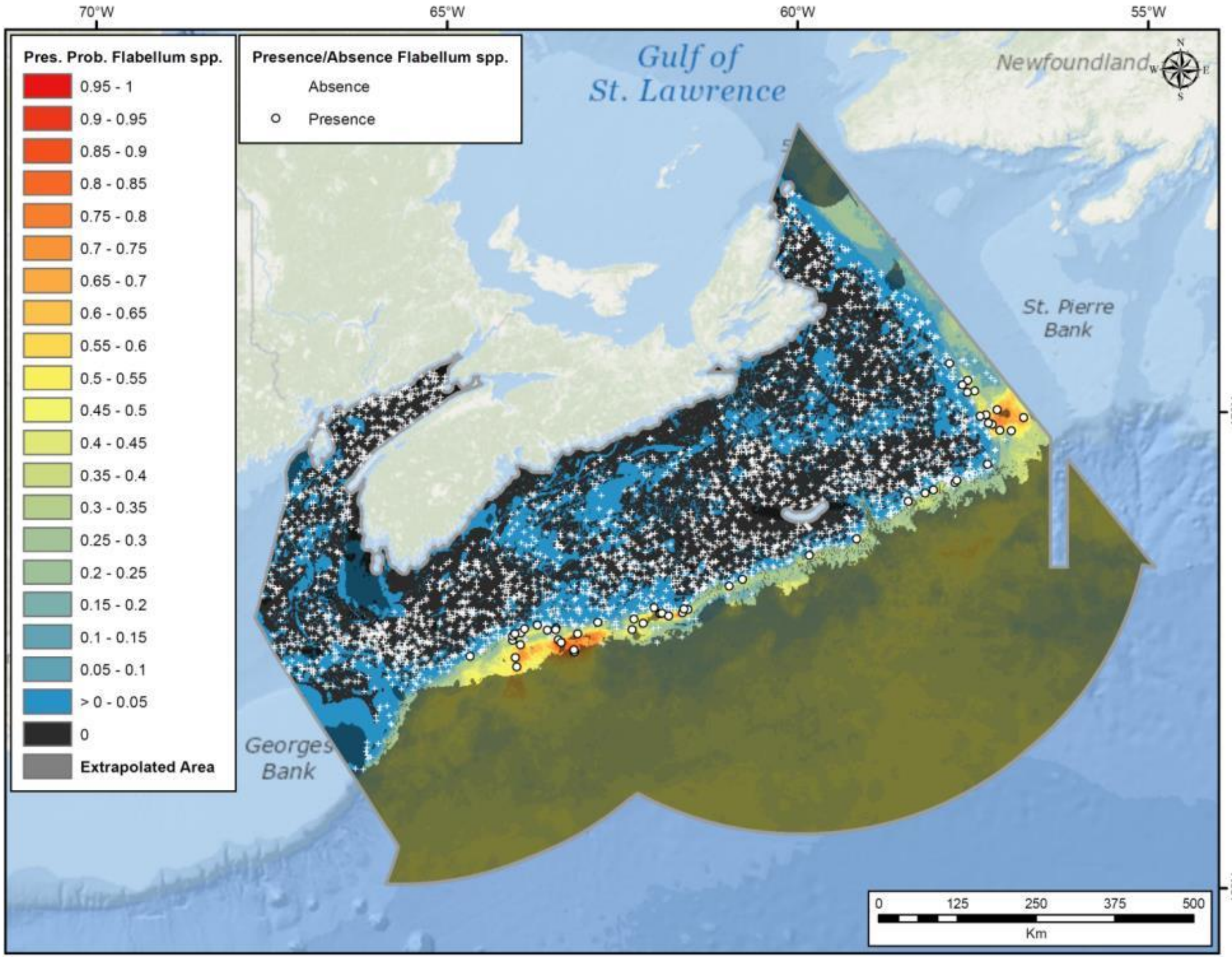

Figure 81. Presence and absence observations and predictions of presence probability (Pres. Prob.) of *Flabellum* spp. based on a random forest model on presence and absence *Flabellum* spp. catch data collected from DFO multispecies trawl surveys conducted in the Maritimes Region between 2003 and 2014. Also shown are the areas of model extrapolation.

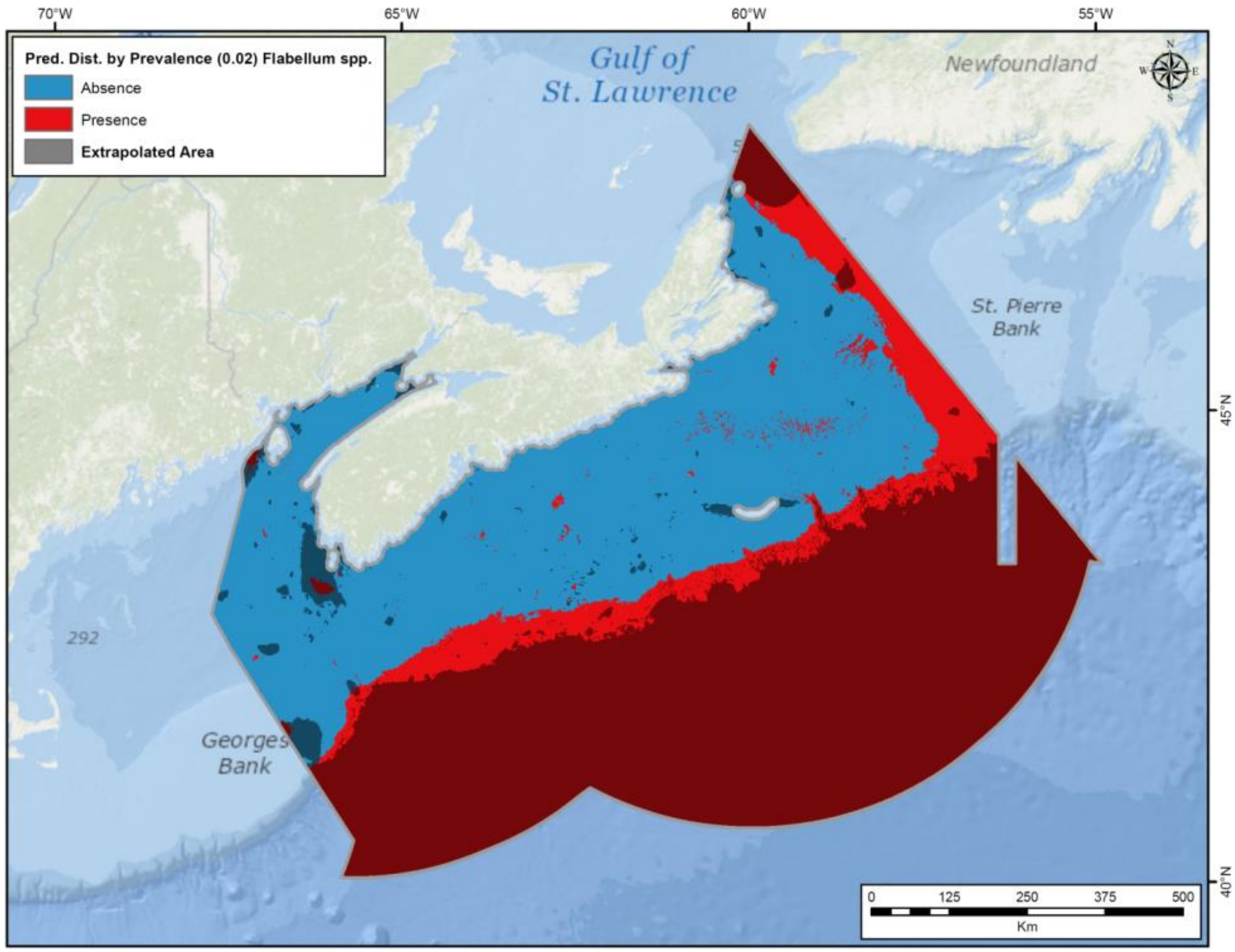

Figure 82. Predicted distribution (Pred. Dist.) of *Flabellum* spp. in the Maritimes Region based on the prevalence threshold (0.02) of *Flabellum* spp. presence and absence data used in the random forest model. Also shown are the areas of model extrapolation (grey polygon may appear red or blue).

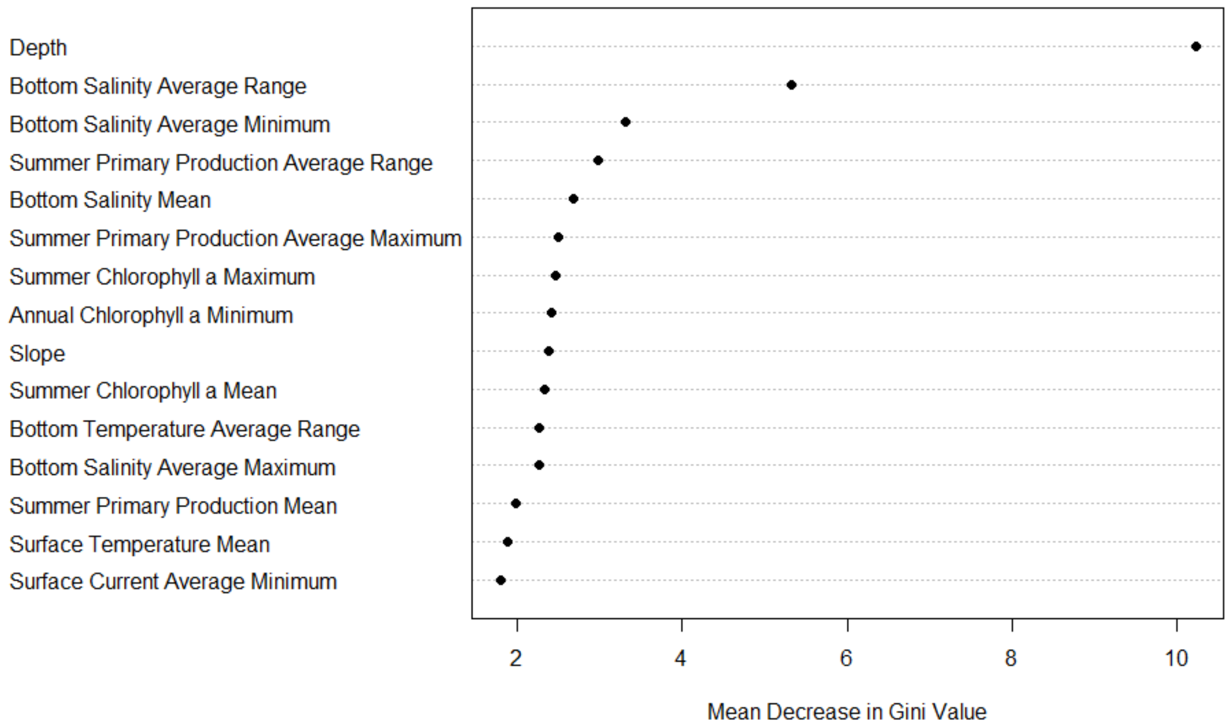

Figure 83. Importance of the top 15 predictor variables measured as the Mean Decrease in Gini value of the random forest model on *Flabellum* spp. presence and absence data in the Maritimes Region. The higher the Mean Gini value the more important the variable is for predicting the response data.

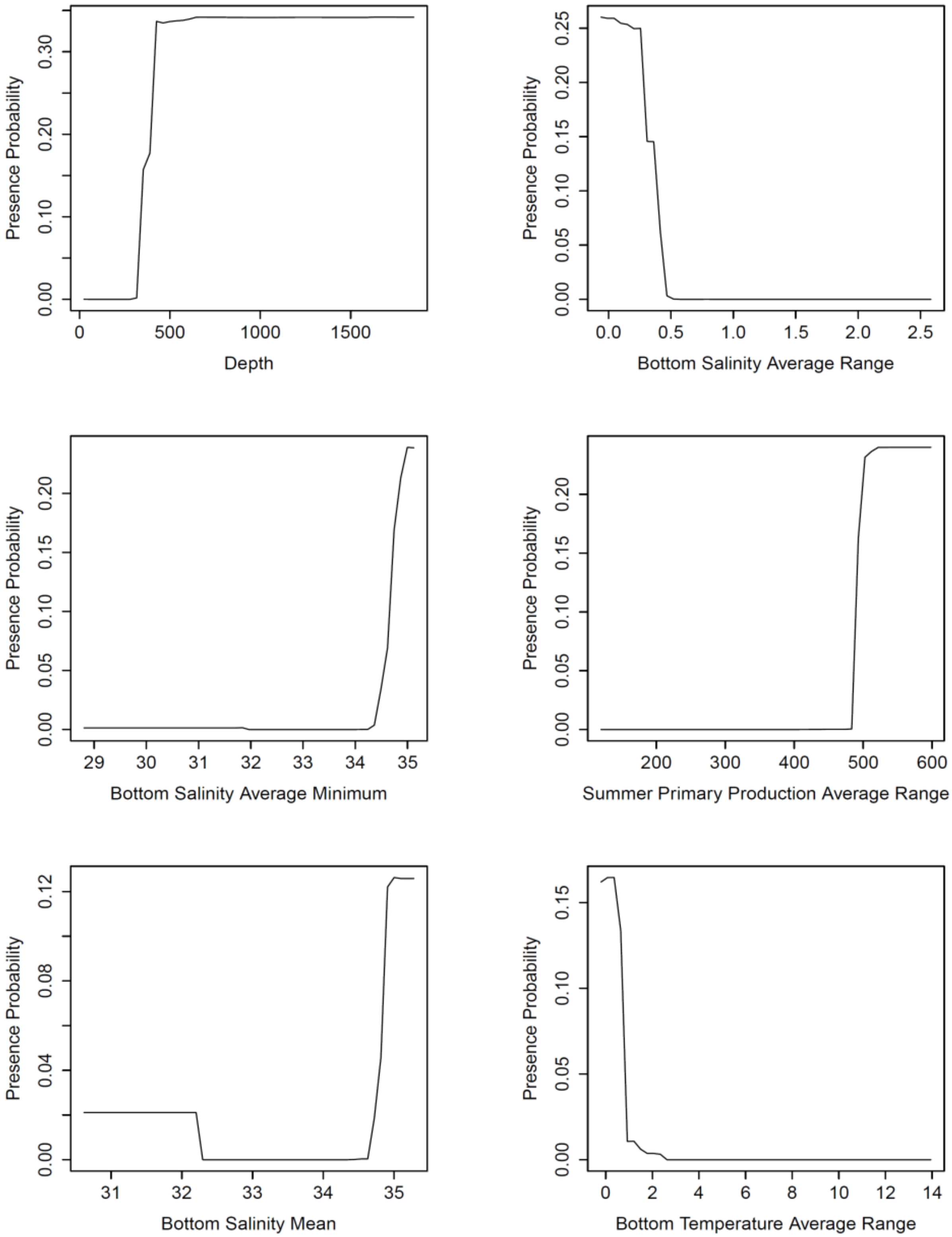

Figure 84. Partial dependence plots of the top six predictors from the random forest model of *Flabellum* spp. presence and absence data collected within the Maritimes Region, ordered left to right from the top. Presence probability is shown on the *y*-axis.

Random Forest Model with Addition of *In Situ* Benthic Imagery Records

Given the relatively low number of *Flabellum* spp. records across the study extent, the DFO multispecies trawl survey data were augmented with 83 additional *in situ* benthic imagery records of *Flabellum* collected in the Maritimes Region between 1965 and 2008 (see Table 25). These records were comprised of *Flabellum alabastrum* and *Flabellum* spp. and were expected to extend the environmental range represented in the data. When added to the DFO multispecies trawl survey presence and absence records, the final dataset consisted of 135 presences and 2541 absences after filtering the data so that only one presence or absence record occurs in a single cell. This combined dataset was remodelled using an unbalanced design and a threshold equal to species prevalence (0.05).

The accuracy measures for this random forest model are shown in Table 26. The average AUC computed from 10-fold cross validation was 0.987 ± 0.005 SD, the highest of the two models. Class error for the presence and absence classes was comparable to that of the former model based on the multispecies trawl survey records. Sensitivity and specificity were the highest of both models.

The additional presence records expanded the area of high *Flabellum* spp. presence probability, particularly along the eastern slope and in the deep-water canyons (Figure 85). The Gully submarine canyon east of Sable Island and the Northeast Channel on the western Scotian Shelf showed much higher *Flabellum* spp. presence probability compared to the former model.

Table 25. Number of *Flabellum* spp. *in situ* benthic imagery records collected between 1965 and 2014 in the Maritimes Region.

Year	Gear	Total number of presences/absences
1965	NRCan Drop Camera	1
1967	NRCan Drop Camera	1
2001	Campod	22
2001	NRCan Drop Camera	3
2003	Campod	4
2005	Campod	1
2006	DSIS	1
2006	ROPOS	4
2007	ROPOS	23
2008	Campod	31
2008	NRCan Drop Camera	6

Table 26. Accuracy measures and confusion matrix from 10-fold cross validation of a random forest model of presence and absence of *Flabellum* spp. within the Maritimes Region. Observ. = Observations; Sensit. = Sensitivity, Specif. = Specificity.

Model Fold	**AUC**	**Observ.**	**Predictions**		**Total n**	**Class error**	**Sensit.**	**Specif.**
1	0.984		**Absence**	**Presence**				
2	0.984	**Absence**	2342	199	2541	0.078	0.956	0.922
3	0.986	**Presence**	6	129	135	0.044		
4	0.987							
5	0.985							
6	0.994							
7	0.991							
8	0.990							
9	0.993							
10	0.977							
Mean	**0.987**							
SD	**0.005**							

These areas of higher probability corresponded well with the location of additional *in situ* benthic imagery observations that were added to this model. The deep portion of the study extent beyond the slope was predicted with a moderate presence probability. This area is still considered extrapolated area by the model (Figure 86).

Figure 87 depicts the predicted distribution of *Flabellum* spp. based on a prevalence threshold of 0.05. In this map, all presence probability values generated from the model that were greater than 0.05 were classified as presence, while values less than 0.05 were classed as absence. The areas predicted as suitable and unsuitable habitat for *Flabellum* spp. are very comparable to that of the former model. The largest difference occurred in deeper water, where the addition of the deep-water *in situ* records extended the areas predicted as suitable habitat further down the slope.

Like the previous model, Depth was the top predictor in this model. Slope was the 2^{nd} most important variable, likely due to the addition of the *in situ* benthic imagery records that are concentrated along the slope and in the deep canyons (Figure 88). Slope was followed by Bottom Salinity Average Range and the remaining variables in the model. The partial dependence plots of the top 6 environmental variables are shown in Figure 89. Presence probability was highest at depths of 500 m and greater and at slopes greater than 10˚.

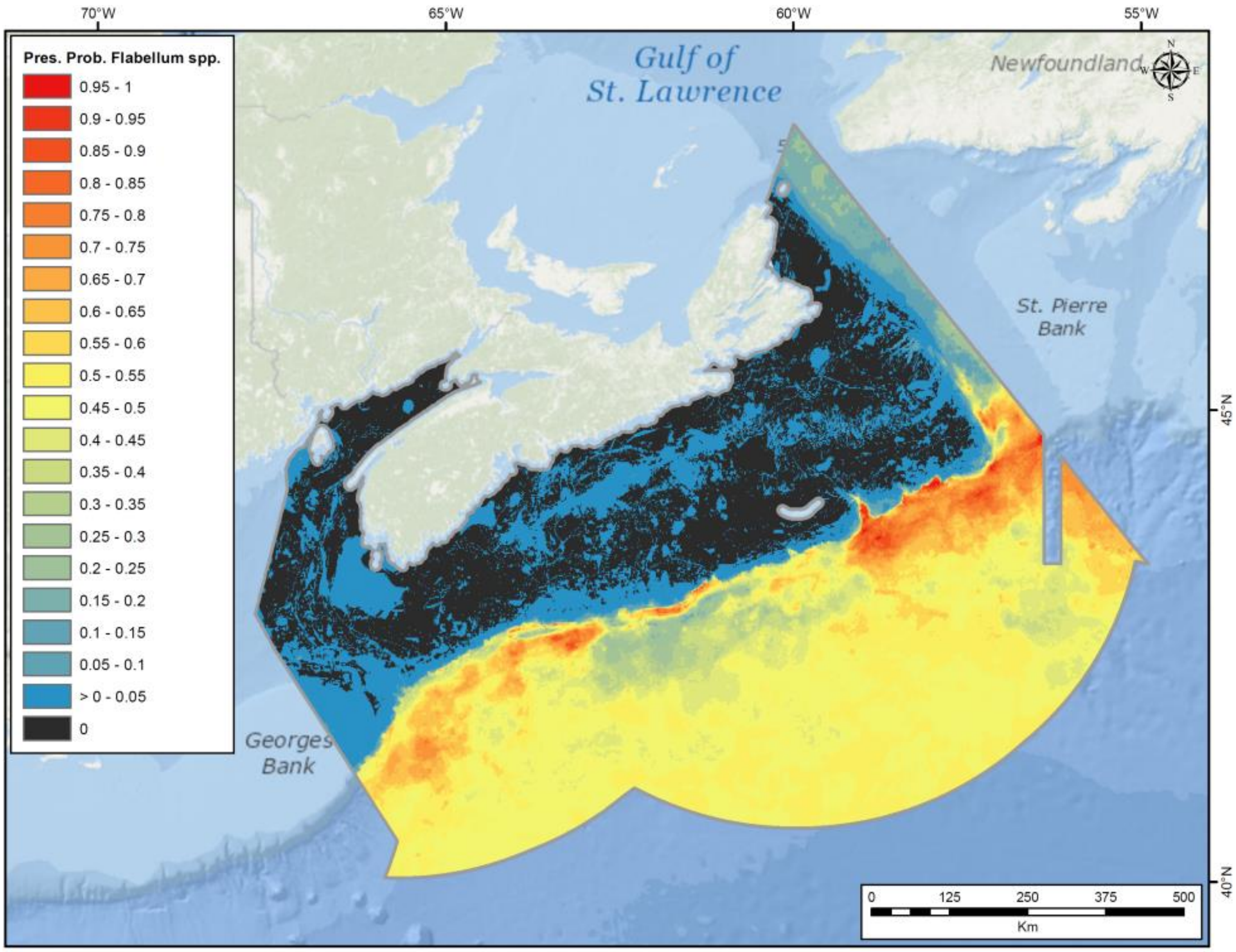

Figure 85. Predictions of presence probability (Pres. Prob.) of *Flabellum* spp. based on a random forest model on presence and absence *Flabellum* spp. catch data collected from DFO multispecies trawl surveys and *in situ* benthic imagery observations of *Flabellum* spp. collected from various surveys conducted in the Maritimes Region between 1965 and 2014.

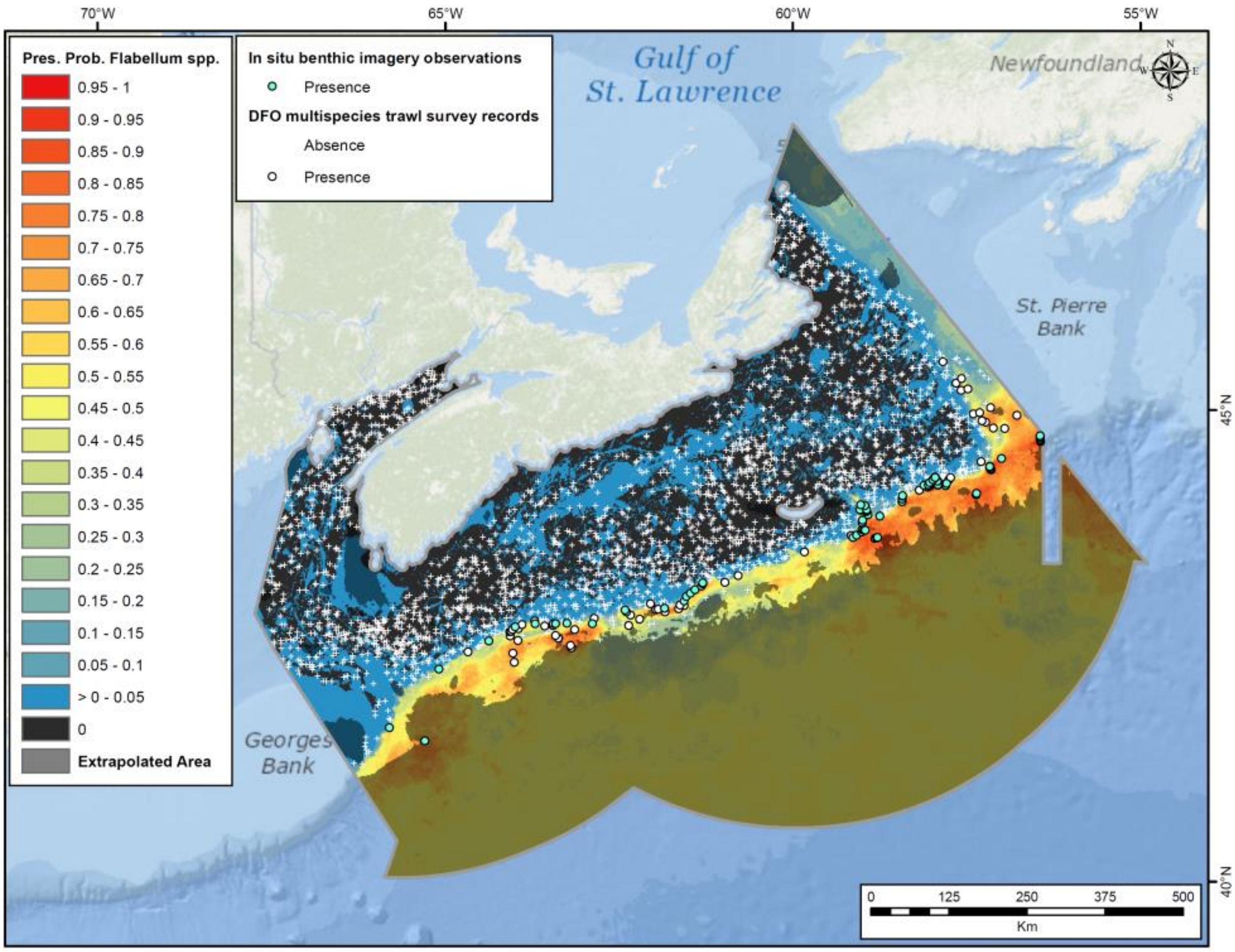

Figure 86. Presence and absence observations and predictions of presence probability (Pres. Prob.) of *Flabellum* spp. based on a random forest model on presence and absence *Flabellum* spp. catch data collected from DFO multispecies trawl surveys and *in situ* benthic imagery observations of *Flabellum* spp. collected from various surveys conducted in the Maritimes Region between 1965 and 2014. Also shown are areas of model extrapolation.

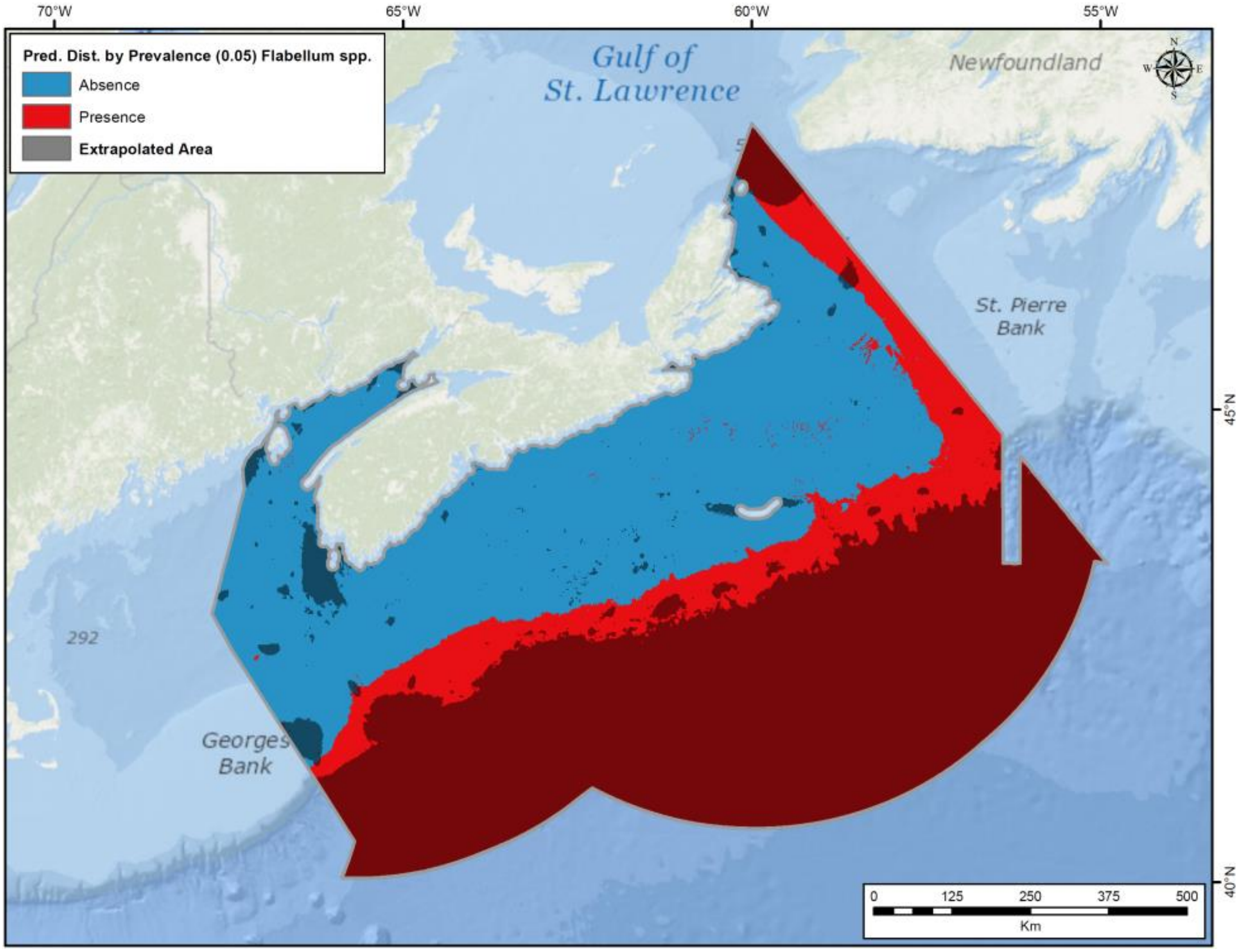

Figure 87. Predicted distribution (Pred. Dist.) of *Flabellum* spp**.** in the Maritimes Region based on the prevalence threshold (0.05) of *Flabellum* spp. presence and absence data used in the random forest model. Also shown are the areas of model extrapolation (grey polygon may appear red or blue).

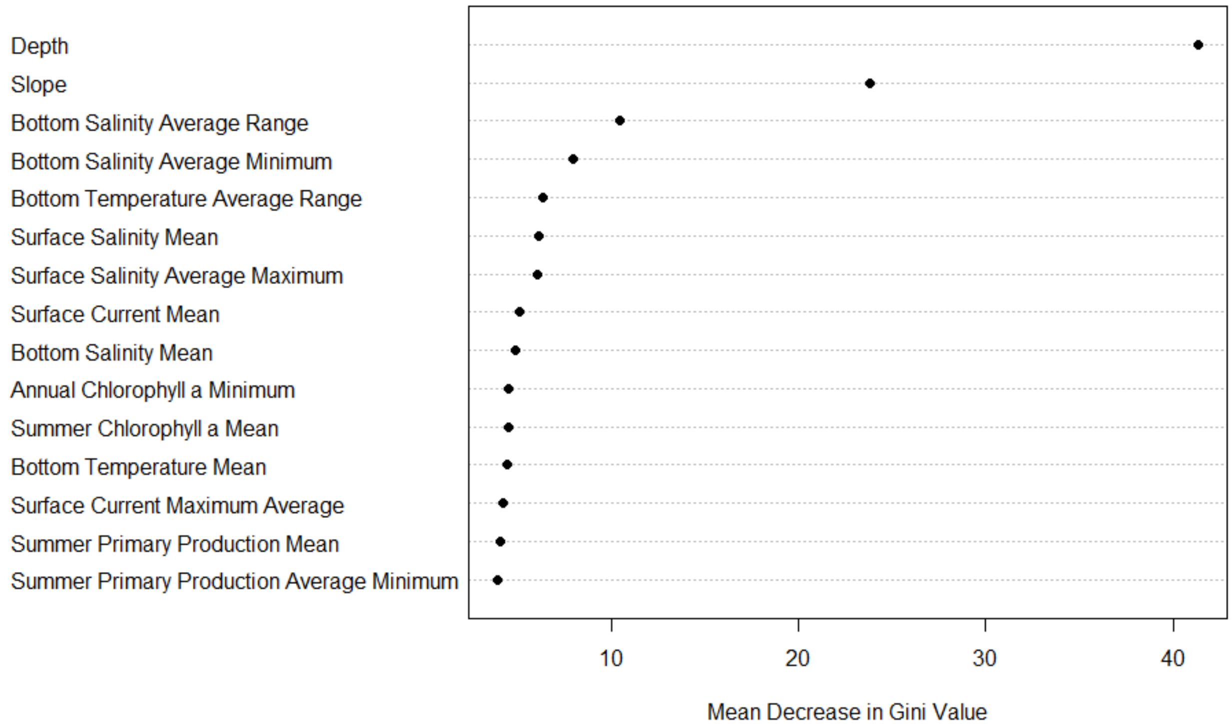

Figure 88. Importance of the top 15 predictor variables measured as the Mean Decrease in Gini value of the random forest model on *Flabellum* spp**.** presence and absence data from DFO multispecies trawl surveys and *in situ* benthic imagery observations collected from various surveys conducted in the Maritimes Region between 1965 and 2014. The higher the Mean Gini value the more important the variable is for predicting the response data.

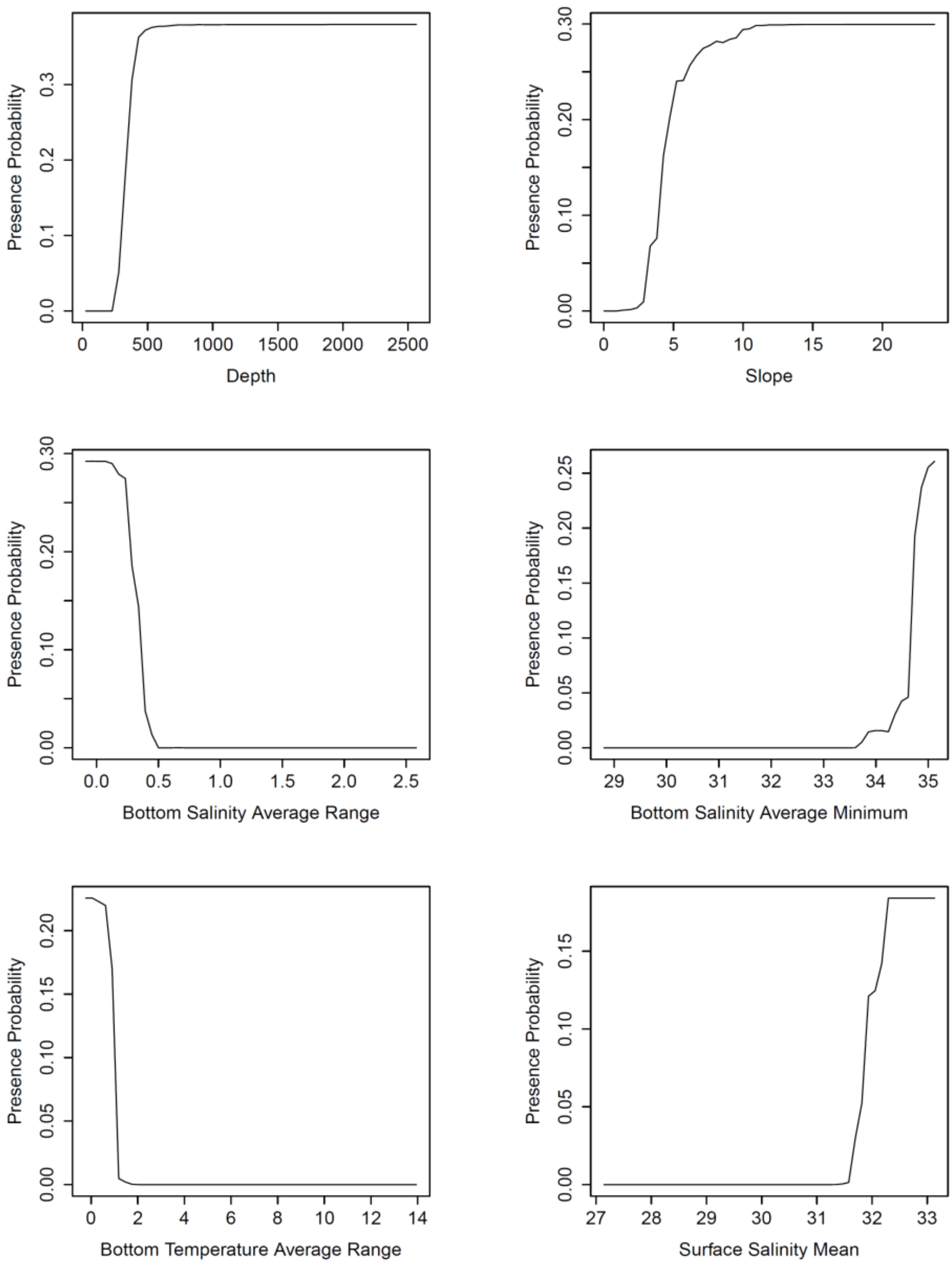

Figure 89. Partial dependence plots of the top six predictors from the random forest model of *Flabellum* spp. presence and absence observations collected from DFO multispecies trawl surveys and *in situ* benthic imagery observations collected within the Maritimes Region between 1965 and 2014, ordered left to right from the top. Presence probability is shown on the *y*-axis.

Model Selection

The model using only records from the DFO multispecies trawl survey was chosen as the best predictor of the distribution of *Flabellum* spp. in the Maritimes Region. There was a high degree of overlap between the multispecies trawl survey records and the *in situ* benthic imagery observations (see Figure 78). Although the accuracy measures for the model with the trawl and *in situ* science records combined were slightly higher than those of the model based only on trawl records, we feel that the inclusion of the *in situ* camera records did not result in a significant enough improvement in the predicted presence probability and prevalence surfaces to warrant including records with a different catchability and no associated absences. Instead, we have chosen to use the *in situ* camera observations as a validation dataset to determine how well the selected model predicts them.

Validation of Selected Model Using Independent Data

Flabellum spp. records from the NOAA Deep-Sea Coral Data Portal, and *F. alabastrum* and *Flabellum* spp. records from DFO *in situ* benthic imagery observations were used to validate the selected model (Figure 90). Considering that there are likely differences in catchability between records used to train the model (collected using Western IIA trawl gear) and the validation records, caution must be taken when interpreting the results of model validation. There were 13 *Flabellum* spp. records from the NOAA database after filtering to 1 per cell. Of these records, 5 (38%) were predicted as absence by the model (Figure 90, upper panel. These records were located off southwestern Nova Scotia and in the Gulf of Maine in a broad area of unsuitable habitat that extends across much of the Scotian Shelf.

Of the 83 *in situ* benthic imagery records, only 5 (0.06%) were predicted as absence by the model (Figure 90, lower panel). These were located on the edge of the Scotian Shelf on the eastern side of the Gully on the edge of Banquereau Bank, and in Haldimand Canyon and where the model delineates the upper depth distribution of the *Flabellum* habitat. Several *Flabellum* records were located in deeper water in areas considered extrapolated by the model. The records predicted as presence were located along the slope and are consistent with the distribution of presence records from the DFO multispecies trawl surveys.

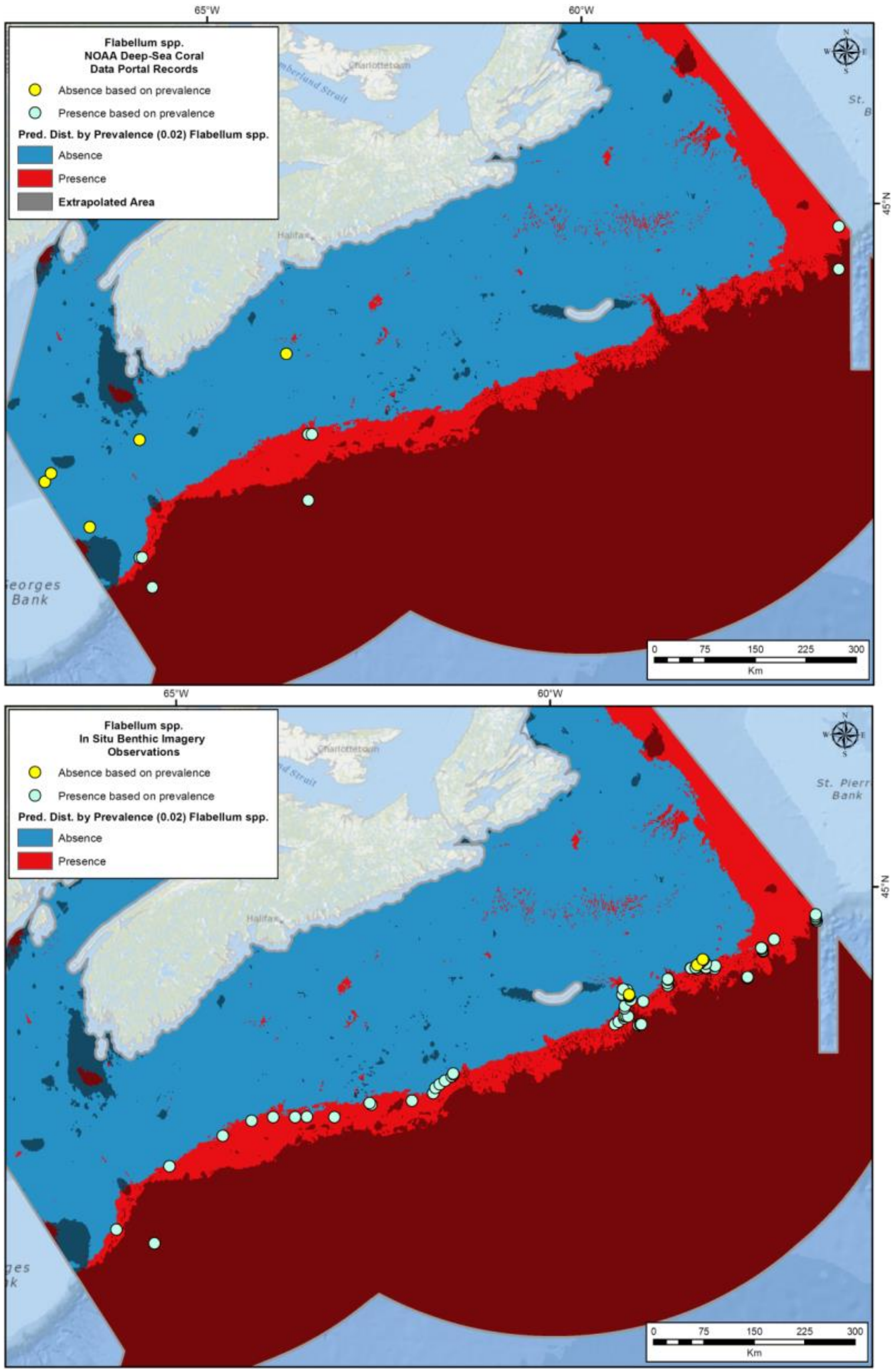

Figure 90. Validation of the predicted distribution by prevalence of *Flabellum* spp. using records from the NOAA Deep-Sea Coral Data Portal (upper panel) and from DFO *in situ* camera surveys (lower panel).

Prediction of *Flabellum* spp. Biomass Using Random Forest

The accuracy measures of the regression random forest model on mean *Flabellum* spp. biomass per grid cell from DFO multispecies trawl surveys are presented in Table 27. This model had relatively good performance, with an average R^2 of 0.234 ± 0.165 SD. The Normalized Root-Mean-Square Error was 0.037 ± 0.014 SD. The percent variance explained was fair (9.56% ± 4.43 SD).

Figures 91 and 92 show the predicted biomass surface of *Flabellum* spp. The majority of the Scotian Shelf was predicted to have zero biomass of *Flabellum* spp. Areas of high predicted biomass occurred on the slope and were associated with the location of high catches (Figure 92). These areas are not considered extrapolated area by the model. An area of moderate predicted biomass in the northeast corner of the study extent off Cape Breton was considered extrapolated area by the model.

Like the presence-absence random forest models, Depth was the top predictor of the biomass distribution of *Flabellum* spp. in the Maritimes Region (Figure 93). This variable was followed distantly in importance by Maximum Average Winter Mixed Layer Depth, Bottom Salinity Average Range, and the other variables in the model. Variables related to surface chlorophyll *a* concentration ranked strongly in this model. The partial dependence of *Flabellum* spp. biomass on the top six most important predictor variables is shown in Figure 94. Like the presence-absence models, predicted biomass rapidly increased at ~500 m depth, but did not fully peak until ~1500 m. Predicted biomass was highest in areas where the Maximum Average Winter Mixed Layer Depth was less than 30 m.

Table 27. Accuracy measures from 10-fold cross validation of a random forest model of average *Flabellum* spp. biomass (kg) per grid cell recorded from DFO multispecies trawl surveys in the Maritimes Region. RMSE = Root-Mean-Square Error; NRMSE = Normalized Root-Mean-Square Error.

Model Fold	**R^2**	**RMSE**	**NRMSE**	**Percent (%) variance explained**
1	0.521	0.026	0.051	5.93
2	0.157	0.018	0.034	16.10
3	0.059	0.014	0.026	15.49
4	0.503	0.016	0.031	3.09
5	0.299	0.016	0.031	6.36
6	0.123	0.015	0.029	10.64
7	0.162	0.034	0.065	4.72
8	0.095	0.025	0.047	11.18
9	0.135	0.022	0.041	10.10
10	0.289	0.010	0.020	11.96
Mean	**0.234**	**0.020**	**0.037**	**9.56**
SD	**0.165**	**0.007**	**0.014**	**4.43**

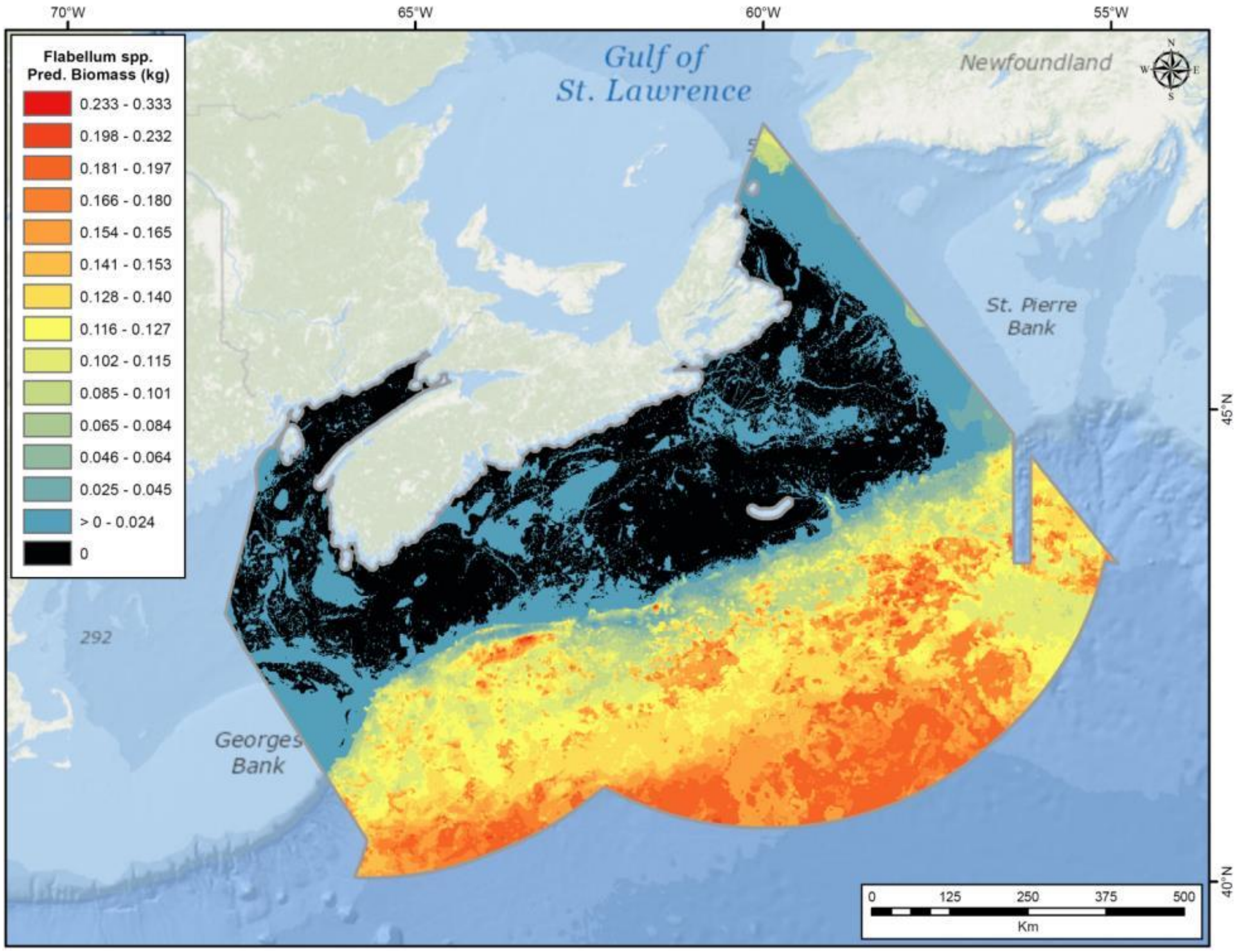

Figure 91. Predictions of biomass (kg) per grid cell of *Flabellum* spp. from catch data recorded in DFO multispecies trawl surveys conducted in the Maritimes Region between 2003 and 2014.

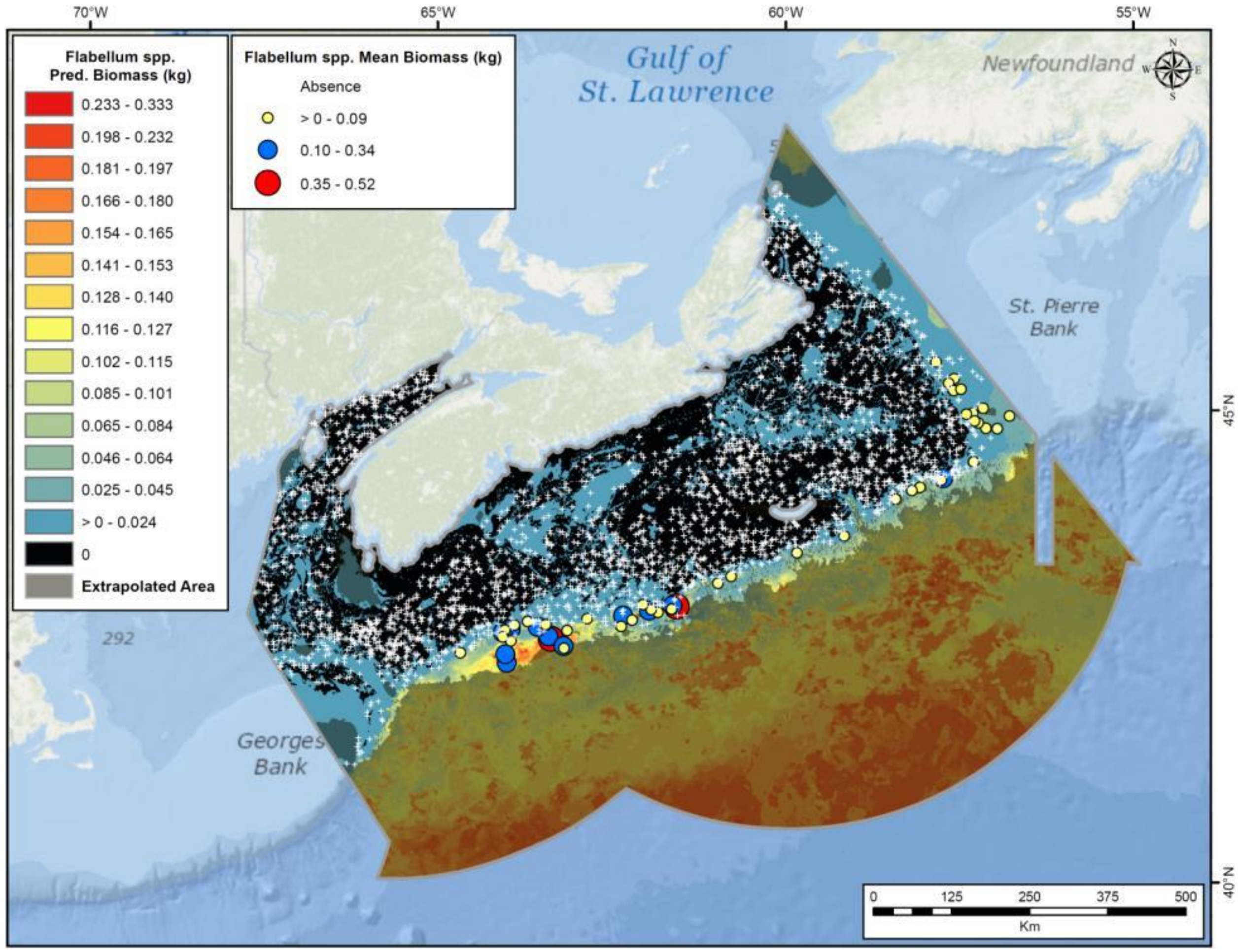

Figure 92. Predictions of biomass (kg) of *Flabellum* spp. from catch data recorded in DFO multispecies trawl surveys conducted in the Maritimes Region between 2003 and 2014. Also shown are the mean biomass values per grid cell and areas of model extrapolation.

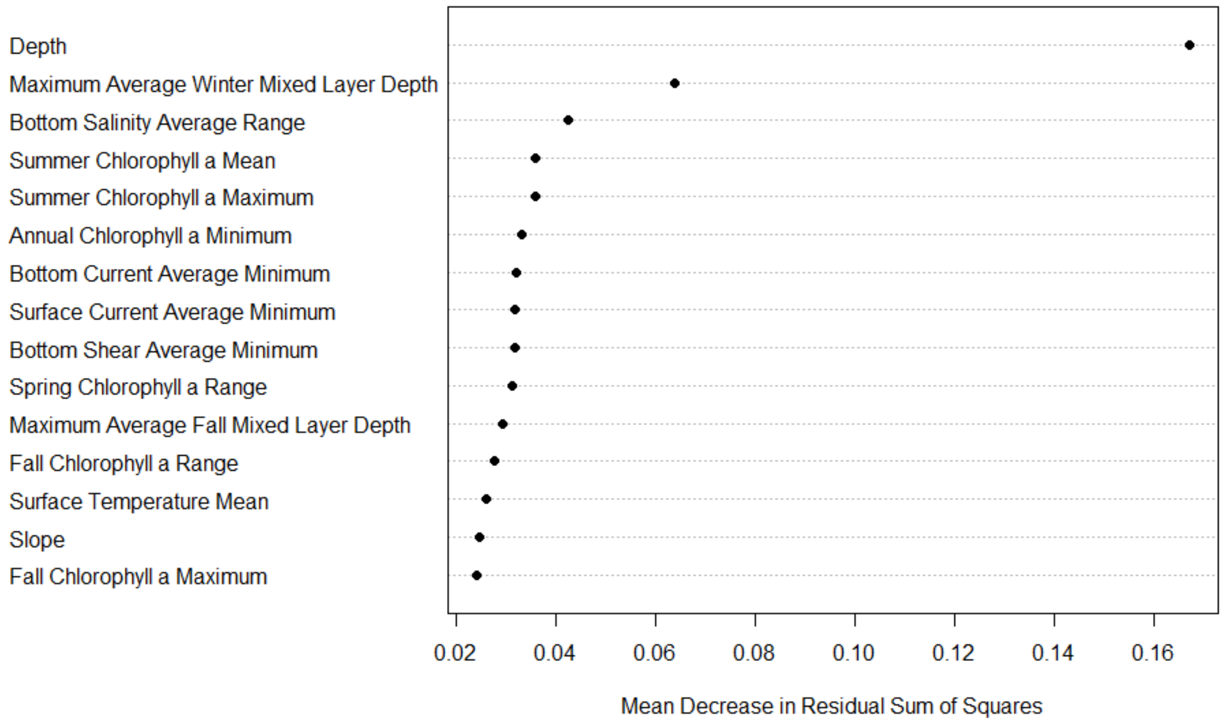

Figure 93. Importance of the top 15 predictor variables measured as the Mean Decrease in Residual Sum of Squares of the regression random forest model on *Flabellum* spp. biomass data from the Maritimes Region. The higher the Mean Decrease in Residual Sum of Squares, the more important the variable is for predicting the response data.

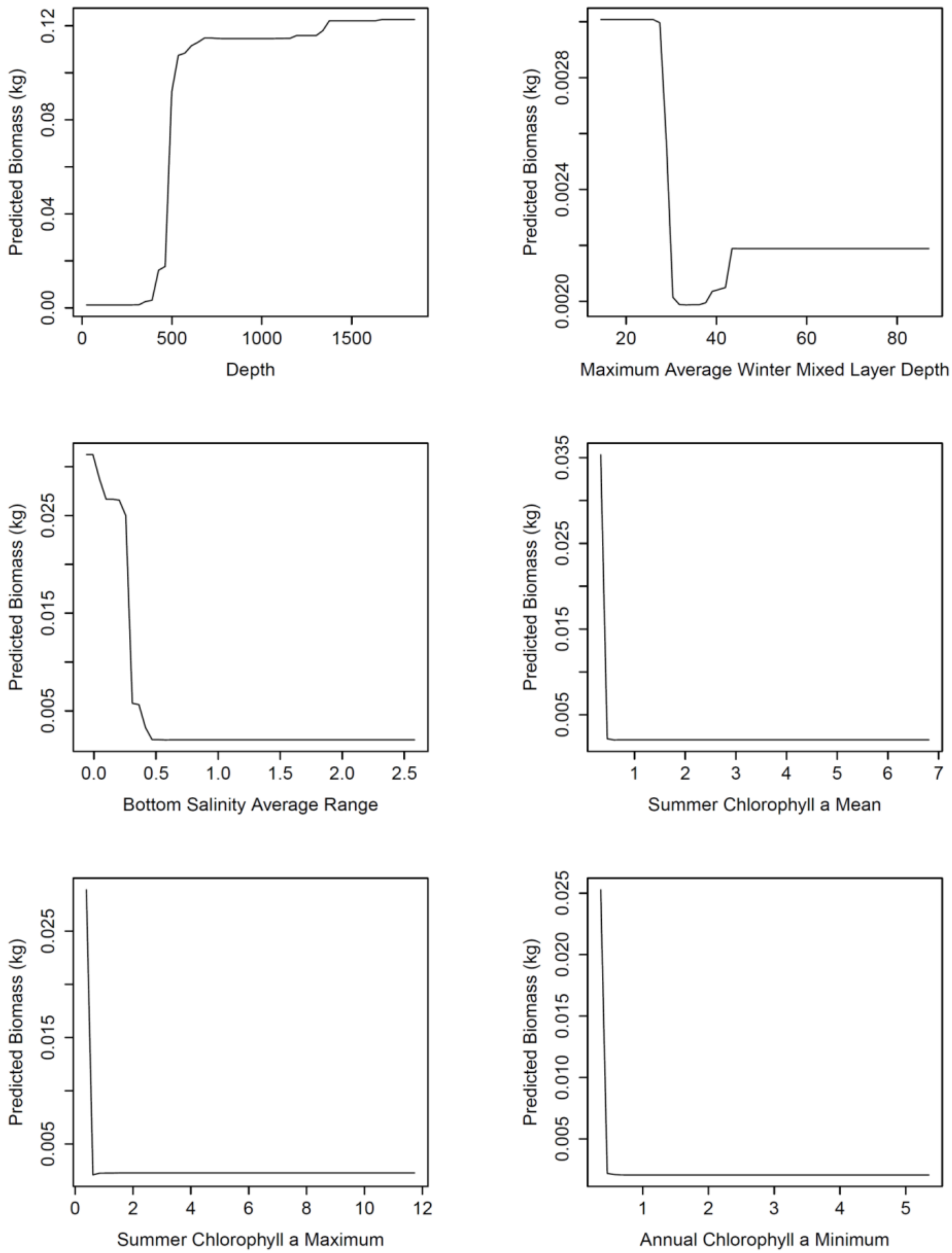

Figure 94. Partial dependence plots of the top six predictors from the random forest model of *Flabellum* spp. biomass data collected within the Maritimes Region, ordered left to right from the top. Predicted biomass (kg) is shown on the *y*-axis.

Kernel Density Analysis of *Flabellum* spp. Catch Data

Kernel density analysis of the *Flabellum* spp. biomass data showed concentrations along the Scotian Slope and in the southern Laurentian Channel (Figure 95). The largest concentrations were observed along the central Scotian Slope southwest of Sable Island Bank and Emerald Basin. The KDE search radius for this group was 11.9 km. Polygons encompassing each of 9 catch thresholds showed that the greatest increase in area between successive weights, after the initial mapping out of the habitat, occurred between 0.04 kg and 0.06 kg, where the area increased by 200.80% (Table 28, Figure 96). There were 8 data points contributing to this expansion (Table 28, Figure 97). Consequently 0.06 kg was considered as the threshold delineating the *Flabellum* spp. habitat. These areas occupy 531 km^2 of seabed. Figure 98 shows the final KDE significant area polygons, the location of significant and non-significant catches, and null records of *Flabellum* spp. from the DFO multispecies trawl survey.

All KDE polygons occurred in areas predicted as suitable habitat when overlain on the predicted distribution by prevalence surface (Figure 99). However, the random forest model predicted a much larger range distribution of *Flabellum* spp. along the slope and beyond than is covered by the KDE polygons, likely due to the high importance of Depth in the model. The KDE polygons were not compared to the RF biomass results due to the poor performance of that model.

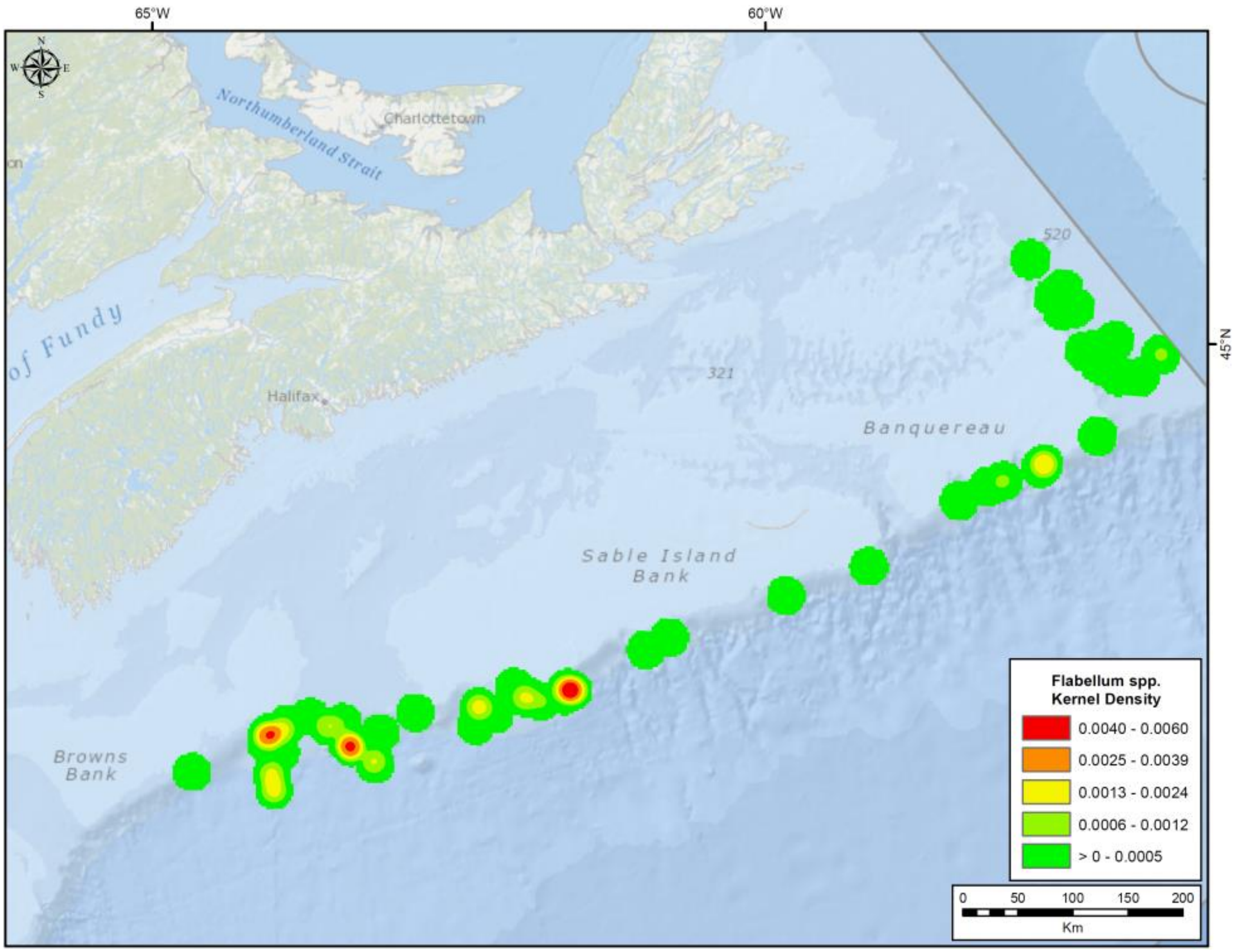

Figure 95. Kernel density analysis of *Flabellum* spp. catches in the Maritimes Region.

Table 28. Summary characteristics for nine *Flabellum* spp. catch weight thresholds. The threshold corresponding to the habitat delineation is indicated by shading.

Number of Points in Polygon	Area (km^{2})	Catch Threshold (kg)	Percent Change in Area (%)
2	16	0.5	63.45
5	25	0.3	825.43
10	235	0.2	16.97
14	275	0.1	92.83
20	531	0.06	200.80
28	1597	0.04	86.19
42	2974	0.02	26.82
51	3772	0.01	40.65
55	5305	0.002	

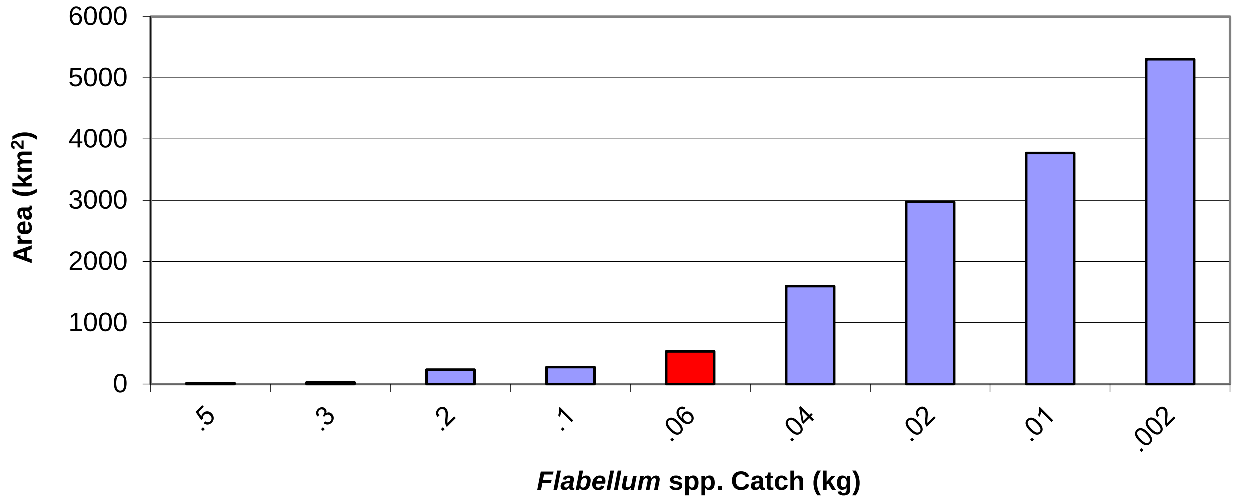

Figure 96. Area occupied by polygons encompassing catches of *Flabellum* spp. greater than or equal to 9 catch thresholds (0.5 kg to 0.002 kg).

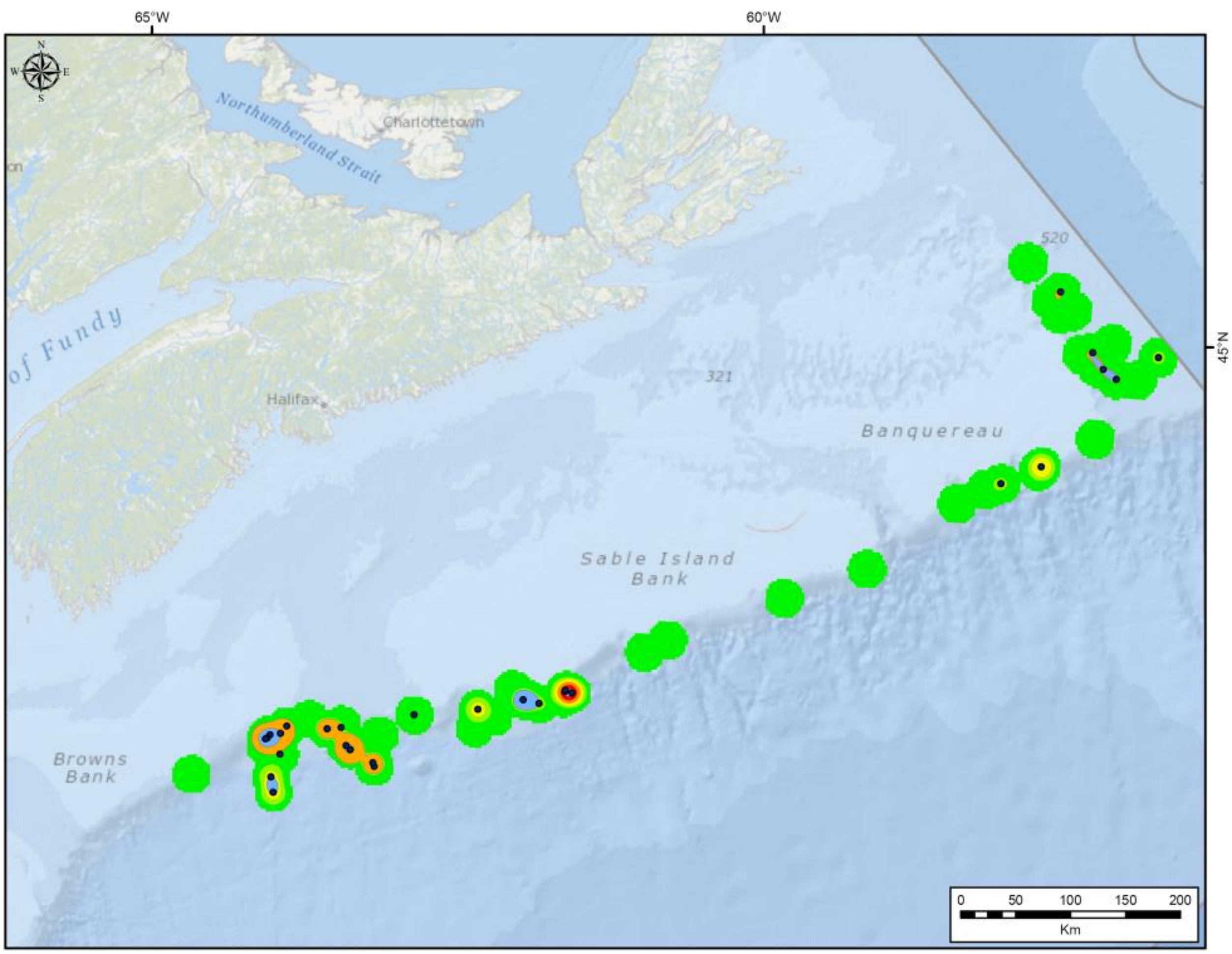

Figure 97. Polygons encompassing *Flabellum* spp. catch weights greater than or equal to 0.06 kg (blue) and 0.04 kg (orange) overlain on the kernel density surface. Tow locations associated with biomass values above these weight thresholds are also shown. The 0.06 kg threshold was used to delineate significant concentrations.

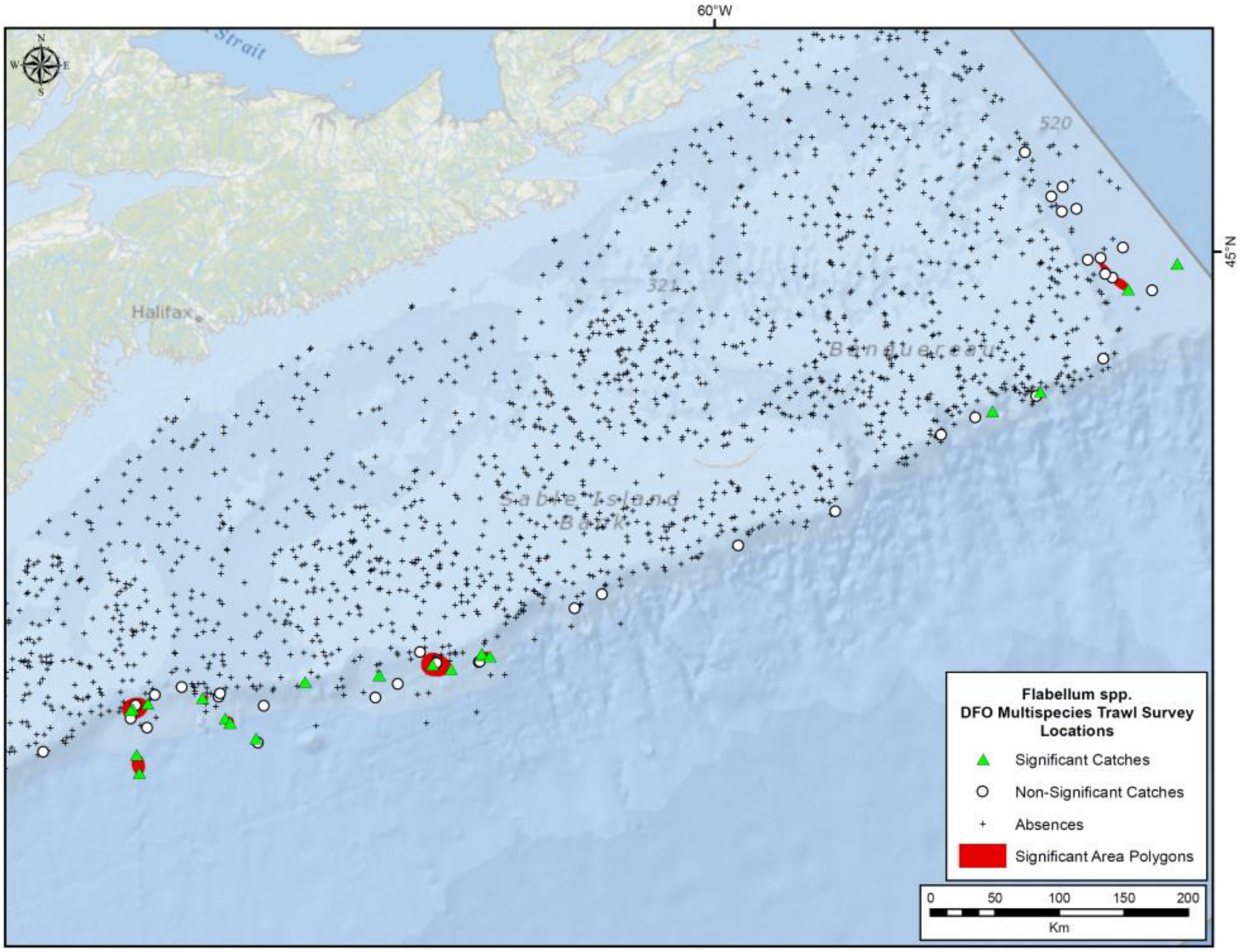

Figure 98. Location of polygons identifying significant *Flabellum* spp. concentrations in the Maritimes Region. Also shown are the significant and non-significant catches of *Flabellum* spp. and absence data (nulls) from the DFO multispecies trawl surveys.

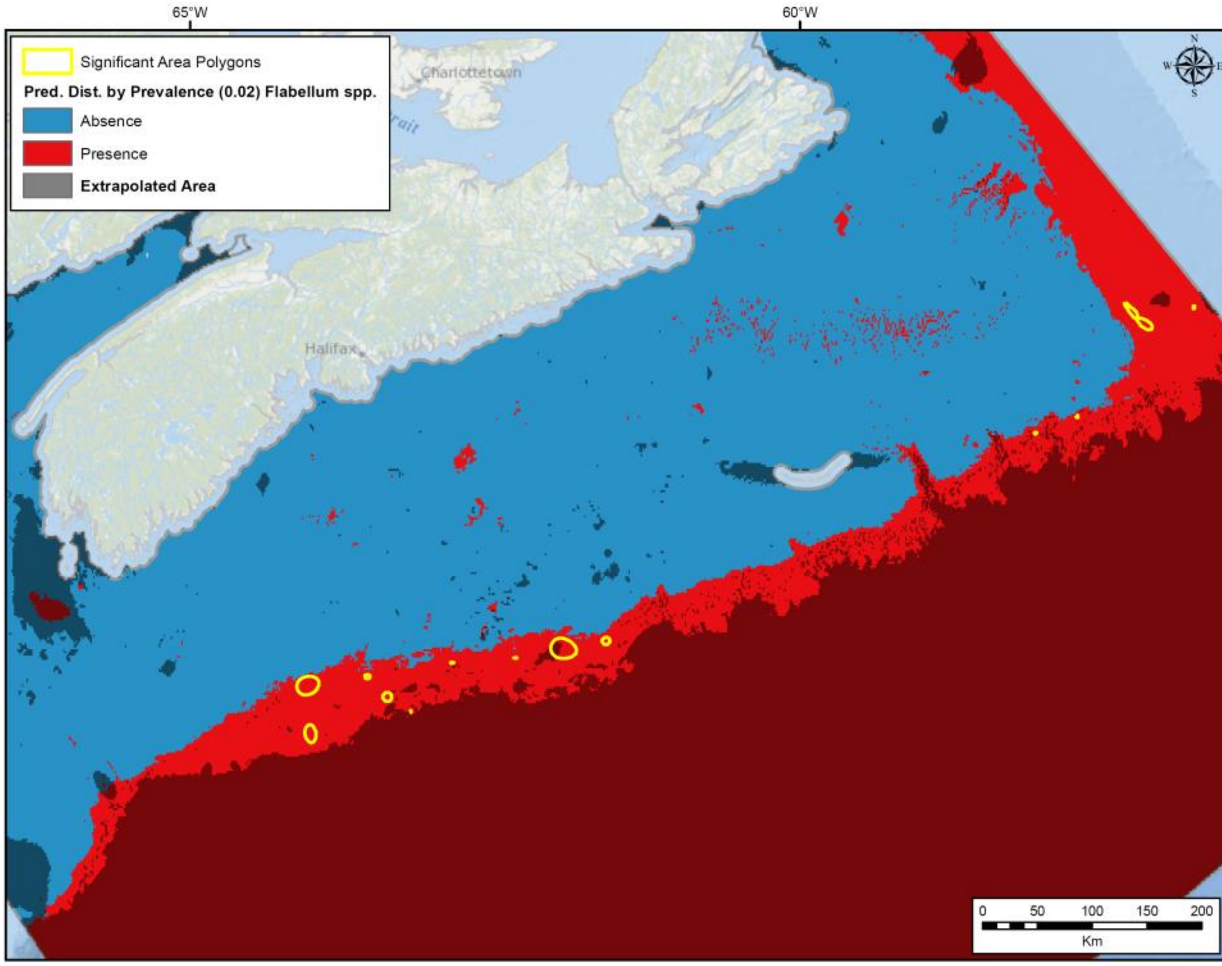

Figure 99. KDE polygons denoting significant concentrations of *Flabellum* spp. overlain on the predicted distribution by prevalence of *Flabellum* spp. generated from the selected random forest model.

DISCUSSION

This study is the first to use random forest modelling to predict the distribution of horse mussels, stalked tunicates, sand dollars, soft corals and *Flabellum* cup corals in the Maritimes Region. Table 29 shows a summary of the accuracy measures for the best runs, when applicable, of the presence-absence random forest models for each of the five taxa. The models were given either a very good or excellent performance rating, with cross-validated AUC values ranging from 0.868 to 0.965. The highest performing model was of *Flabellum* Cup Corals, a group containing at least one of the three *Flabellum* species known to occur in the region (Buhl-Mortensen et al., 2007; Cogswell et al., 2009). The high performance of this model can likely be attributed to the narrow environmental envelope of *Flabellum* in the region (narrow spatial and depth distribution with records occurring mainly along the shelf break and slope) (McPherson and Jetz, 2007; Tsoar et al., 2007). When applied to predict the biomass distribution of these taxa, random forest performed poorly. This is consistent with other applications of regression random forest to trawl catch data in the Maritimes (Beazley et al., 2016a), Arctic (Beazley et al., 2016c), and Newfoundland Regions (Guijarro et al., 2016a,b), and can possibly be attributed to the highly imbalanced design and large number of zero catches included in the models, and/or high variability in the positive catches (Li and Heap, 2008).

Table 29. Summary of the random forest presence-absence mean accuracy measures and model prevalence (i.e. the proportion of presences relative to the entire dataset) values for use as cut-offs for random forest predicted probabilities.

Taxonomic Group	Prevalence	AUC	Sensitivity	Specificity
Horse Mussels (*Modiolus modiolus*) Beds	0.05	0.918	0.839	0.874
Stalked Tunicate Fields	0.14	0.868	0.787	0.795
Sand Dollar Beds	0.27	0.886	0.848	0.783
Soft Coral Gardens	0.14	0.878	0.793	0.779
Flabellum Cup Corals	0.02	0.965	0.943	0.904

In species distribution modelling, predictions of presence probability exhibit bias towards the majority class when species prevalence (i.e., species frequency, or the proportion of presences in relation to the total dataset) is skewed (Real et al., 2006; Hanberry and He, 2013). Jiménez-Valverde et al. (2009) found that the effects of prevalence were only significant for highly unbalanced datasets (prevalence <0.01 or >0.99), suggesting that the impact of unbalanced prevalence for the datasets modelled in our study may be minimal (lowest prevalence was 0.02; see Table 28). Nonetheless, the authors stated that selection of an appropriate cut-off threshold to convert the probability map into a presence-absence map is crucial for avoiding the drawbacks of unbalanced prevalence. The selection of this threshold has implications for conservation management, where boundaries indicating species presence are often required from model outputs. A common approach is to convert the continuous probabilities into discrete presence and absence classes by adjusting the threshold that specifies the predicted probability above which a species is determined present to match species prevalence (Liu et al., 2005; Hanberry and He, 2013). This approach has been adopted in the current works and in previous random forest applications in the Northwest Atlantic (see Beazley et al., 2016a; Guijarro et al. 2016 a,b; Murillo et al., 2016). Howell et al. (2016) used a similar method on MaxEnt model outputs of sponge species in the northwest Atlantic, defining areas of presence and absence based on a user-defined threshold, but further differentiated those values above the threshold into areas of high and low presence probability to highlight the most significant areas. As the 'prevalence' surfaces generated in this report do not differentiate significant areas of a taxon from their broader, low-level distribution, they are best used to complement kernel density analyses and as habitat templates for areas where no other data are available. To date, prevalence surfaces have been used to modify the boundaries of coral and sponge KDE-derived VME in the NAFO region (see NAFO, 2016) and KDE-derived significant area polygons to create Significant Benthic Areas domestically (Kenchington et al., 2016a).

Given that the trawl catchability of the taxa modelled in this report is largely unknown and considered low for some (e.g. sand dollars and horse mussels that can bury in soft sediment), validation of both the SDM surfaces and KDE-derived significant benthic area polygons using *in situ* camera and/or video surveys is recommended. In 2013, NAFO's Working Group on Ecosystem Science and Assessment (WGESA) generated KDE polygons for large sea squirts (which included only *Boltenia ovifera*) and erect bryozoans on the Tail of Grand Bank (see NAFO, 2015). WGESA recommended that ground-truthing of the polygons be done using *in situ* camera and/or video surveys prior to formal adoption of these polygons as VME. In 2015 DFO conducted *in situ* photographic surveys for this purpose on the Tail of Grand Bank, but did not observe *B. ovifera* within the large sea squirt KDE polygons. The absence of this species was thought to be due to its small patch size and preference for rocky outcrops and hard substrate that occurs on scales of less than 1 km (less than the length of the trawl tows; NAFO, 2015). As a result, WGESA recommended that the location of the catches, and not the full KDE polygons, be adopted as 'significant concentrations' only (NAFO, 2015). Whether the same patch sizes occur in the Maritimes Region is unknown. The high density of points in the KDE polygon for *B. ovifera* off Cape Breton suggests that more extensive habitat may occur there. Furthermore, the KDE significant area polygons for soft corals should undergo further validation given that the biomass distribution may be skewed by the inclusion of hard attachment substrate in the weight calculations. The maximum mean catch biomass of soft corals in the Maritimes Region (5.72 kg) was within the range of that found in other regions that use similar trawl gear (6.60 kg for *Gersemia rubiformis* in the Gulf Region; Murillo et al., 2016) and do not include rocks in the weight calculations, suggesting no highly erroneous biomass values. Also, there was good spatial congruence between the KDE polygons and presence-absence SDM results (with the latter being unaffected by these weight issues), giving more validity to the KDE analysis of this taxon.

Significant concentrations as defined by KDE encompass all catches above the delimiting threshold, and smaller, non-significant catches that could be indicative of new recruitment, different species composition, or areas thinned by fishing impacts (Kenchington et al., 2016a). In some cases, the KDE polygons are comprised of only one significant tow with its associated 12 - 21 km radius around the start position. This was most prominent in horse mussels in this report, and large gorgonian corals in Beazley et al. (2016a). These single catches are not indicative of a continuous habitat, particularly if they are surrounded by null records. In these cases NAFO does not consider these polygons to meet VME criteria (NAFO, 2013), but suggested that further research using benthic imagery could validate them. Most single-tow KDE polygons for the taxa modelled in this report were within the vicinity of other positive catches, indicating smaller patch sizes that possibly relate to the patchy distribution of their attachment substrate and/or fishing disturbance. Nonetheless, without further validation of these areas we recommend that these single-tow KDE significant area polygons not be used to drive site selection in any conservation management applications, and instead be used in a descriptive sense to add value after a potential conservation area is selected.

ACKNOWLEDGMENTS

We thank Marty King (DFO) for his guidance throughout the development of this report. We thank Javier Murillo and Leslie Nasmith (both DFO) for their review of the report, and J. Murillo and Javier Guijarro (DFO) for their input on the model outputs. This project was funded in part by a one year project under DFO’s Strategic Program for Ecosystem-Based Research and Advice (SPERA) to EK and through financial support by DFO’s Oceans and Coastal Management Division, Ecosystem Management Branch.

REFERENCES

Beazley, L., Kenchington, E., Murillo, F.J., Lirette, C., Guijarro, J., McMillan, A., and Knudby, A. 2016a. Species Distribution Modelling of Corals and Sponges in the Maritimes Region for Use in the Identification of Significant Benthic Areas. Can. Tech. Rep. Fish. Aquat. Sci. 3172: vi + 189p.

Beazley, L., Lirette, C., Sabaniel, J., Wang, Z., Knudby, A., and Kenchington, E. 2016b. Characteristics of Environmental Data Layers for Use in Species Distribution Modelling in the Gulf of St. Lawrence. Can. Tech. Rep. Fish. Aquat. Sci. 3154: viii + 357p.

Beazley, L., Murillo, F.J., Kenchington, E., Guijarro, J., Lirette, C., Siferd, T., Treble, M., Wareham, V., Baker, E., Bouchard Marmen, M., and Tompkins MacDonald, G. 2016c. Species Distribution Modelling of Corals and Sponges in the Eastern Arctic for Use in the Identification of Significant Benthic Areas. Can. Tech. Rep. Fish. Aquat. Sci. 3175: vii + 210p.

Breiman, L. 2001. Random forests. Machine Learning 45: 5–32.

Brown, C.J., Sameoto, J.A., and Smith, S.J. 2012. Multiple methods, maps, and management applications: Purpose made seafloor maps in support of ocean management. J. Sea. Res. 72: 1-13.

Buhl-Mortensen, L., Mortensen, P.B., Armsworthy, S., and Jackson, D. 2007. Field observations of *Flabellum* spp. and laboratory study of the behavior and respiration of *Flabellum alabastrum*. Bull. Mar. Sci. 81(3): 543-552.

Cogswell, A.T., Kenchington, E.L.R., Lirette, C.G., MacIsaac, K., Best, M.M., Beazley, L.I., and Vickers, J. 2009. The current state of knowledge concerning the distribution of coral in the Maritimes Provinces. Can. Tech. Rep. Fish. Aquat. Sci. 2855, v + 66p.

DFO, 2004. Identification of Ecologically and Biologically Significant Areas. DFO Can. Sci. Advis. Sec. Ecosystem Status Rep. 2004/006.

DFO, 2012. Marine Protected Area Network Planning in the Scotian Shelf Bioregion: Objectives, Data, and Methods. Can. Sci. Advis. Sec. Sci. Advis. Rep. 2012/064.

DFO, 2014a. Maritimes Research Vessel Summer Survey Trends. DFO Can. Sci. Advis. Sec. Sci. Resp. 2014/017.

DFO, 2014b. Offshore Ecologically and Biologically Significant Areas in the Scotian Shelf Bioregion. DFO Can. Sci. Advis. Rep. 2014/041.

DFO, 2015. Information on Potential Sensitive Benthic Areas in the Bay of Fundy: Head Harbour/West Isles/Passages and the *Modiolus* Reefs, Nova Scotia Shore. DFO Can. Sci. Advis. Sec. Sci. Resp. 2014/044. (Erratum: January 2016).

ESRI. 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

Etnoyer, P., and Morgan, L. 2003. Occurrences of Habitat-forming Deep Sea Corals in the Northeast Pacific Ocean. Technical Report, NOAA Office of Habitat Conservation, Marine Biology Conservation Institute, Washington, Redmond. 33pp.

Francis, F. T-Y., Filbee-Dexter, K., and Scheibling, R.E. 2014. Stalked tunicates *Boltenia ovifera* form biogenic habitat in the rocky subtidal zone of Nova Scotia. Mar. Biol. 161: 1375-1383.

Gass, S.E. 2002. An Assessment of the Distribution and Status of Deep Sea Corals in Atlantic Canada by Using both Scientific and Local Forms of Knowledge. MES thesis, Dalhousie University, Halifax, Canada.

Gordon Jr., D.C., and Kenchington, E.L.R. 2007. Deep-water corals in Atlantic Canada: A review of DFO Research (2001-2003). Proc. N.S. Inst. Sci. 44(1): 27-50.

Government of Canada, 2011. National Framework for Canada's Network of Marine Protected Areas. Fisheries and Oceans Canada, Ottawa. 31pp.

Guijarro, J., Beazley, L., Lirette, C., Kenchington, E., Wareham, V., Gilkinson, K., Koen-Alonso, M., and Murillo, F.J. 2016a. Species Distribution Modelling of Corals and Sponges from Research Vessel Survey Data in the Newfoundland and Labrador Region for Use in the Identification of Significant Benthic Areas. Can. Tech. Rep. Fish. Aquat. Sci. 3171: vi + 126p.

Guijarro, J., Kenchington, E., Murillo, F.J., Beazley, L., Lirette, C., Wareham, V., Koen-Alonso, M. 2016b. Species Distribution Modelling of Crinoids, Bryozoans and Ascidians in the Newfoundland and Labrador Region. Can. Tech. Rep. Fish. Aquat. Sci. 3181: v + 60p.

Hanberry, B.B. and He, H.S. 2013. Prevalence, statistical thresholds, and accuracy assessment for species distribution models. Web Ecol. 13: 13-19.

Howell, K-L., Piechaud, N., Downie, A-L., and Kenny, A. 2016. The distribution of deep-sea sponge aggregations in the North Atlantic and implications for their effective spatial management. Deep-Sea Res. I 115: 309-320.

Jiménez-Valverde, A., Lobo, J.M., and Hortal, J. 2009. The effect of prevalence and its interaction with sample size on the reliability of species distribution models. Comm. Ecol. 10(2): 196-205.

Kenchington, E. 2014. A General Overview of Benthic Ecological or Biological Significant Areas (EBSAs) in Maritimes Region. Can. Tech. Rep. Fish. Aquat. Sci. 3072: iv + 45p.

Kenchington, E., L. Beazley, F. J. Murillo, G. Tompkins MacDonald, E. Baker. 2015. Coral, Sponge, and Other Vulnerable Marine Ecosystem Indicator Identification Guide, NAFO Area. NAFO Sci. Coun. Studies, 47: 1–74. doi:10.2960/S.v47.m1.

Kenchington, E., Beazley, L., Lirette, C., Murillo, F.J., Guijarro, J., Wareham, V., Gilkinson, K., Koen Alonso, M., Benoît, H., Bourdages, H., Sainte-Marie, B., Treble, M., and Siferd, T. 2016a. Delineation of Coral and Sponge Significant Benthic Areas in Eastern Canada Using Kernel Density Analyses and Species Distribution Models. DFO Can. Sci. Advis. Sec. Res. Doc. 2016/093. vi + 178p.

Kenchington, E., Cogswell, A., Lirette, C. and Murillo-Perez, F.J. 2009. The Use of Density Analyses to Delineate Sponge Grounds and Other Benthic VMEs from Trawl Survey Data. Serial No. N5626. NAFO SCR Doc. 09/6, 15p.

Kenchington, E.L.R., Gilkinson, K.D., MacIsaac, K.G., Bourbonnais-Boyce, C., Kenchington, T.J., Smith, S.J., and Gordon, D.C. 2006. Effects of experimental otter trawling on benthic assemblages on Western Bank, Northwest Atlantic Ocean. J. Sea Res. 56: 249–270.

Kenchington, E., Lirette, C., Cogswell, A., Archambault, D., Archambault, P., Benoit, H., Bernier, D., Brodie, B., Fuller, S., Gilkinson, K., Lévesque, M., Power, D., Siferd, T., Treble, M., and Wareham, V. 2010. Delineating Coral and Sponge Concentrations in the Biogeographic Regions of the East Coast of Canada Using Spatial Analyses. DFO Can. Sci. Advis. Sec. Res. Doc. 2010/041. vi + 202pp.

Kenchington, E., Lirette, C., Murillo, F.J., Beazley, L., Guijarro, J., Wareham, V., Gilkinson, K., Koen Alonso, M., Benoît, H., Bourdages, H., Sainte-Marie, B., Treble, M., Siferd, T. 2016b. Kernel Density Analyses of Coral and Sponge Catches from Research Vessel Survey Data for Use in Identification of Significant Benthic Areas. Can. Tech. Rep. Fish. Aquat. Sci. 3167: viii + 207p.

Kenchington, E., Murillo, F.J., Lirette, C., Sacau, M., Koen-Alonso, M., Kenny, A., Ollerhead, N., Wareham, V., and Beazley, L. 2014. Kernel Density Surface Modelling as a Means to Identify Significant Concentrations of Vulnerable Marine Ecosystem Indicators. PLoS ONE 10(1): e0117752. doi:10.1371/journal.pone.0117752.

Kostylev, V.E., Parrott, D.R., Dickson, R. and Todd, B.J. 2009. Distribution and morphology of horse mussel beds in the Bay of Fundy identified using multibeam sonar. 8th International Conference, GeoHab - marine geological and biological habitat mapping; Trondheim; NO; May 5-7, 2009,49p.

Li, J. and Heap, A.D. 2008. A Review of Spatial Interpolation Methods for Environmental Scientists. Geoscience Australia, Record 2008/23, 137pp.

Liu, C., Berry, P.M., Dawson, T.P., and Pearson, R.G. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28: 385-393.

McPherson, J.M. and Jetz, W. 2007. Effects of species' ecology on the accuracy of species distribution models. Ecography 30: 135-151.

Murillo, F.J., Kenchington, E., Beazley, L., Lirette, C., Knudby, A., Guijarro, J., Benoît, H., Bourdages, H., and Sainte-Marie, B. 2016. Distribution Modelling of Sea Pens, Sponges, Stalked Tunicates and Soft Corals from Research Vessel Survey Data in the Gulf of St. Lawrence for Use in the Identification of Significant Benthic Areas. Can. Tech. Rep. Fish. Aquat. Sci. 3170: vi + 132p.

NAFO, 2013. Report of the 6th Meeting of the NAFO Scientific Council Working Group on Ecosystem Science and Assessment (WGESA). NAFO SCS Doc. 13/024, Serial No. N6277, 209pp.

NAFO, 2015. Report of the 8th Meeting of the NAFO Scientific Council (CS) Working Group on Ecosystem Science and Assessment (WGESA). NAFO SCS Doc. 15/19, Serial No. N6549, 176pp.

NAFO. 2016. Report of the Scientific Council Meeting, 03 - 16 June 2016. NAFO SCS Doc. 16/14, Serial No. N6587, 296p.

Plough, H.H. 1969. Genetic polymorphism in a stalked ascidian from the Gulf of Maine. J. Hered. 60: 193–205.

Real, R., Barbosa, A.M., and Vargas, J.M. 2006. Obtaining environmental favourability functions from logistic regression. Environ. Ecol. Stat. 13: 237-245.

Rincón, B. and Kenchington, E.L. 2016. Influence of Benthic Macrofauna as a Spatial Structuring Agent for Juvenile Haddock (*Melanogrammus aeglefinus*) on the Eastern Scotian Shelf, Atlantic Canada. PLoS ONE 11(9): e0163374. doi:10.1371/journal.pone.0163374.

Secretariat of the Convention on Biological Diversity (SCBD).2010. COP-10 Decision X/2. Secretariat of the Convention on Biological Diversity, Nagoya. https://www.cbd.int/doc/decisions/cop-10/cop-10-dec-02-en.pdf.

Stanley, D. J. and James, N.P. 1971. Distribution of *Echinarachnius parma* (Lamarck) and Associated Fauna on Sable Island Bank, Southeast Canada. Smithsonian Contrib. Earth Sci. No. 6, 24pp.

Tremblay, M.J., Black, G.A.P., and Branton, R.M. 2007. The distribution of common decapod crustaceans and other invertebrates recorded in annual ecosystem surveys of the Scotian Shelf 1999-2006. Can. Tech. Rep. Fish. Aquat. Sci. 2762, iii + 74p.

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D., and Kadmon, R. 2007. A comparative evaluation of presence-only methods for modelling species distribution. Biod. Res. 13: 397 - 405.

WoRMS Editorial Board 2017. World Register of Marine Species. Available from http://www.marinespecies.org at VLIZ. Accessed 2017-01-19. doi:10.14284/170.

APPENDIX A - Locations of Tow Positions Used to Delineate KDE Significant Areas

Table A1. Mission number and set and position information of DFO multispecies trawl survey records used to identify significant area polygons for horse mussels (*Modiolus modiolus*). Set number is the last 3 digits of the Mission Number and Set.

Year	Mission Number and Set	Start Lat. (DD)	Start Long. (DD)	End Lat. (DD)	End Long. (DD)	*Modiolus modiolus* Weight (kg)
1999	NED1999929004	44.0423	-60.9040	44.0183	-60.8792	27.72
2010	NED2010027051	45.1020	-65.1162	45.0900	-65.1458	6.360
2013	NED2013028139	44.0263	-61.0980	44.0552	-61.1053	3.312
2009	NED2009841007	41.9335	-66.3163	41.9572	-66.2935	1.252
2011	NED2011025140	44.0980	-60.6392	44.0992	-60.5983	0.732
2013	NED2013002052	41.8118	-66.4785	41.7848	-66.4672	0.690
2013	NED2013028154	44.1208	-59.5532	44.1275	-59.5142	0.589
2012	NED2012022076	45.0812	-65.3475	45.0955	-65.3108	0.522
2010	NED2010027050	45.1145	-65.2382	45.1345	-65.2100	0.335
2010	NED2010027052	45.3120	-65.3157	45.2835	-65.3090	0.272
2009	NED2009002021	43.7167	-61.5990	43.7400	-61.6242	0.258

Table A2. Mission number and set and position information of DFO multispecies trawl survey records used to identify significant area polygons for *Boltenia ovifera*. Set number is the last 3 digits of the Mission Number and Set.

Year	Mission Number and Set	Start Lat. (DD)	Start Long. (DD)	End Lat. (DD)	End Long. (DD)	*Boltenia ovifera* Weight (kg)
2008	TEM2008830169	45.6180	-59.0822	45.6427	-59.0585	51.000
2007	TEM2007686051	44.9590	-58.7900	44.9307	-58.7833	25.650
2010	NED2010027116	46.9020	-60.2387	46.9173	-60.2023	4.950
2008	TEM2008830148	46.3177	-59.4907	46.3305	-59.4523	3.350
2007	TEL2007745166	46.4025	-60.1830	46.4300	-60.1717	3.000
2009	NED2009027148	46.2045	-59.3258	46.1832	-59.2985	2.700
2013	NED2013022230	46.4740	-59.7068	46.4507	-59.7338	2.660
2009	NED2009027160	45.4572	-58.1577	45.4405	-58.1922	2.210
2008	TEM2008830134	45.1840	-59.1453	45.1995	-59.1107	2.204
2007	TEL2007745162	46.9560	-60.3102	46.9837	-60.3102	2.200

2008	TEM2008830165	45.2900	-58.5465	45.3192	-58.5495	2.044
2007	TEL2007745167	46.3902	-59.8822	46.4055	-59.8463	1.900
2009	NED2009027136	46.0608	-59.3123	46.0592	-59.3543	1.836
2007	TEL2007745130	45.0295	-58.9555	45.0550	-58.9723	1.700
2012	NED2012022168	46.0590	-59.2898	46.0883	-59.2878	1.609
2007	TEL2007745169	45.8573	-59.5880	45.8620	-59.5478	1.600
2007	TEL2007745175	45.3807	-58.9185	45.4085	-58.9152	1.600
2008	TEL2008805036	45.3408	-58.7557	45.3200	-58.7855	1.600
2011	NED2011002031	41.8490	-66.3945	41.8560	-66.4347	1.540
2008	TEM2008830133	45.0953	-59.0385	45.1152	-59.0683	1.476
2010	NED2010002034	44.9477	-58.1713	44.9182	-58.1762	1.390
2008	TEM2008830139	46.2240	-59.6958	46.2117	-59.6565	1.350
2010	NED2010002031	45.0945	-58.5418	45.0673	-58.5598	1.115
2008	TEM2008830138	45.9887	-59.4047	45.9638	-59.4157	1.100
2013	NED2013022231	46.5840	-59.4998	46.6050	-59.5210	1.092
2012	NED2012022173	46.5657	-59.8293	46.5503	-59.8657	1.079
2014	NED2014018175	46.7367	-59.8445	46.7190	-59.8777	1.078
2011	NED2011025207	46.1885	-59.2348	46.1592	-59.2318	1.047
2007	TEM2007686026	45.0988	-58.1730	45.1158	-58.2013	1.040
2014	NED2014018172	46.4443	-59.8620	46.4702	-59.8400	1.026
2007	TEL2007745168	45.8793	-59.7175	45.8603	-59.7362	1.000

Table A3. Mission number and set and position information of DFO multispecies trawl survey records used to identify significant area polygons for sand dollars. Set number is the last 3 digits of the Mission Number and Set.

Year	Mission Number and Set	Start Lat. (DD)	Start Long. (DD)	End Lat. (DD)	End Long. (DD)	Sand Dollar Weight (kg)
2013	NED2013022205	45.0603	-57.6922	45.0420	-57.7237	30.850
2014	NED2014018213	44.1975	-58.6712	44.1803	-58.7063	24.000
2010	NED2010002053	44.6208	-57.9668	44.5960	-57.9893	19.660
2003	NED2003003054	44.1012	-59.9263	44.1015	-59.9657	19.350
2010	NED2010027174	44.0653	-60.2305	44.0683	-60.2702	15.500
2003	NED2003003059	44.2452	-60.5528	44.2748	-60.5515	13.570
2009	NED2009027172	44.7048	-58.0338	44.7330	-58.0437	12.660
2009	NED2009002037	44.3508	-60.5917	44.3467	-60.5522	9.940
2008	TEL2008805011	44.0503	-59.9793	44.0527	-59.9397	9.900
2002	NED2002003005	44.2170	-60.7425	44.2320	-60.7800	9.100
2009	NED2009002045	44.5658	-61.0588	44.5375	-61.0650	8.270
2003	NED2009027168	44.4667	-58.7855	44.4375	-58.7835	7.440
2003	NED2003042044	44.8528	-57.6300	44.8537	-57.5873	6.800

2003	NED2003003057	44.3272	-60.2322	44.3033	-60.2517	6.780
2003	NED2003003015	44.7390	-59.6647	44.7343	-59.6240	6.780
2002	NED2002003097	43.7885	-60.5833	43.7738	-60.6183	6.090
2013	NED2013028018	44.7515	-57.8432	44.7302	-57.8722	5.990
2013	NED2013022157	44.6060	-59.3057	44.6053	-59.3470	5.730
2003	NED2003042054	44.5890	-58.1312	44.6120	-58.1040	5.670
2010	NED2010002034	44.9477	-58.1713	44.9182	-58.1762	5.520
2009	NED2009002027	43.6467	-60.5478	43.6473	-60.5075	5.500
2010	NED2010027135	44.7815	-57.6145	44.7527	-57.6170	5.370
2009	NED2009002033	43.7407	-60.3992	43.7135	-60.4163	4.970
2003	NED2003003037	44.2663	-58.1913	44.2502	-58.2242	4.790
2003	NED2003003060	44.2650	-60.9262	44.2755	-60.9648	4.760
2002	NED2002003094	43.5763	-60.4092	43.5953	-60.4403	4.760
2010	NED2010027210	43.7987	-59.9333	43.7788	-59.9040	4.430
2013	NED2013022162	44.0687	-60.3838	44.0772	-60.4223	4.280
2009	NED2009002043	44.4915	-60.6812	44.4797	-60.7178	4.200
2000	NED2000966007	44.4205	-60.8432	44.4493	-60.8458	3.950
2002	NED2002003095	43.6737	-60.3930	43.7010	-60.3760	3.950
2013	NED2013022154	44.6310	-58.8162	44.6433	-58.8527	3.760
2010	NED2010002051	44.7455	-57.8113	44.7332	-57.8485	3.740
2013	NED2013022152	44.2327	-58.8673	44.2608	-58.8555	3.672
2003	NED2003003047	44.3913	-58.8135	44.4017	-58.7747	3.650
2013	NED2013022168	44.6103	-60.4360	44.6375	-60.4482	3.554
2010	NED2010002052	44.6202	-57.8210	44.5910	-57.8140	3.520
2006	NED2006036046	44.6387	-58.2740	44.6475	-58.2327	3.450
2003	NED2003042060	44.4828	-58.7373	44.4708	-58.7008	3.440
2002	NED2002003089	43.8107	-59.6423	43.8110	-59.6825	3.260
2011	NED2011025060	45.3715	-65.2595	45.3957	-65.2378	3.202
2010	NED2010027180	44.6297	-60.4378	44.6570	-60.4552	3.150
2013	NED2013022203	44.6698	-57.4088	44.6967	-57.4260	3.040
2013	NED2013022149	44.4810	-59.0100	44.4748	-58.9718	3.040
2013	NED2013022145	44.0093	-59.2697	44.0105	-59.3113	3.014
2000	NED2000431083	46.2885	-59.3070	46.2852	-59.2645	3.000
2003	NED2003036137	44.0637	-59.7470	44.0702	-59.7088	2.792
2003	NED2003003071	43.7212	-59.7717	43.7308	-59.7333	2.770
2009	NED2009002038	44.3542	-60.3610	44.3545	-60.3202	2.750
2005	NED2005002050	44.2255	-60.7788	44.2025	-60.7533	2.730
2011	NED2011025238	44.4782	-58.4218	44.4863	-58.4613	2.704
2008	TEM2008830129	44.5197	-59.0628	44.5102	-59.0257	2.680
2003	NED2003042061	44.4330	-58.5495	44.4440	-58.5108	2.630
2010	NED2010027228	42.6547	-63.4120	42.6415	-63.4838	2.610
2000	NED2000966104	43.8107	-59.8772	43.8143	-59.8368	2.600
2002	NED2002003058	44.7695	-58.2197	44.7707	-58.2615	2.560
2011	NED2011025145	44.1012	-60.1372	44.1072	-60.0977	2.540
1999	NED1999929007	43.6455	-60.8383	43.6300	-60.8733	2.480
2008	TEL2008805006	44.5027	-60.7840	44.5305	-60.8033	2.450

2013	NED2013028015	45.0760	-57.6805	45.0513	-57.7032	2.432
2003	NED2003003055	44.2497	-60.0688	44.2760	-60.0867	2.395
2003	NED2003003024	44.8463	-58.0490	44.8758	-58.0458	2.335
2009	NED2009002019	43.5272	-61.2755	43.5275	-61.3158	2.310
2000	NED2000966008	44.4987	-60.6760	44.5275	-60.6850	2.300
2010	NED2010027188	44.1668	-61.0012	44.1957	-61.0067	2.280
2011	NED2011025061	45.2945	-65.3043	45.3163	-65.2783	2.232
2002	NED2002003065	44.5293	-59.4955	44.5085	-59.4655	2.230
2006	TEL2006615063	43.8083	-60.0090	43.8117	-59.9680	2.200
2010	NED2010027233	42.6043	-63.9540	42.6200	-63.8818	2.190
2006	TEL2006615064	43.8473	-59.8838	43.8502	-59.8433	2.050
2003	NED2003003023	44.9662	-58.0627	44.9572	-58.1025	2.045
2000	NED2000966039	44.9573	-57.7012	44.9320	-57.7215	2.040
2003	NED2003042019	46.0503	-59.1813	46.0677	-59.1465	2.000

Table A4. Mission number and set and position information of DFO multispecies trawl survey records used to identify significant area polygons for soft corals. Set number is the last 3 digits of the Mission Number and Set.

Year	Mission Number and Set	Start Lat. (DD)	Start Long. (DD)	End Lat. (DD)	End Long. (DD)	Soft Coral Weight (kg)
2006	NED2006036046	44.6387	-58.2740	44.6475	-58.2327	5.720
2007	TEL2007745153	45.5552	-57.8778	45.5837	-57.8920	1.650
2010	NED2010027216	43.0557	-61.2592	43.1065	-61.2840	1.090
2006	NED2006030031	42.7228	-66.1503	42.7058	-66.1642	0.880
2006	NED2006036047	44.6837	-57.9337	44.6912	-57.8933	0.780
2005	TEL2005633044	44.9428	-57.7225	44.9677	-57.7002	0.510
2006	NED2006036017	45.6703	-59.7888	45.6992	-59.7807	0.290
2006	NED2006030030	42.7355	-66.0493	42.7058	-66.0595	0.235
2006	TEL2006615088	45.5202	-59.6840	45.5123	-59.7245	0.205
2005	TEL2005633058	44.3855	-58.2937	44.4143	-58.2860	0.160
2009	NED2009027168	44.4667	-58.7855	44.4375	-58.7835	0.152
2006	NED2006030032	42.7562	-66.6948	42.7857	-66.6935	0.150
2006	NED2006036048	44.9275	-57.8630	44.9555	-57.8443	0.140
2013	NED2013022149	44.4810	-59.0100	44.4748	-58.9718	0.128
2005	TEL2005633010	45.5920	-60.1375	45.6033	-60.0992	0.122
2005	TEL2005633011	45.6278	-59.9620	45.6505	-59.9345	0.122
2005	TEL2005633089	44.1983	-58.8233	44.1922	-58.7830	0.118
2011	NED2011025190	45.8203	-59.6868	45.8352	-59.6498	0.118
2006	TEL2006615090	45.4468	-60.1335	45.4262	-60.1632	0.110
2006	NED2006036055	44.3818	-57.3452	44.3848	-57.3963	0.110
2007	TEL2007745169	45.8573	-59.5880	45.8620	-59.5478	0.106

2006	NED2006036019	45.8542	-59.4680	45.8677	-59.4313	0.105
2013	NED2013022150	44.4035	-58.9863	44.4125	-59.0223	0.104
2003	NED2003042026	46.5453	-59.9902	46.5713	-59.9730	0.100
2007	TEM2007686017	44.2698	-58.6987	44.2730	-58.6627	0.100
2012	NED2012022059	42.7112	-66.1968	42.6968	-66.2318	0.100
2007	TEL2007745162	46.9560	-60.3102	46.9837	-60.3102	0.088
2013	NED2013022229	46.6453	-59.9703	46.6727	-59.9542	0.088
2006	TEL2006615089	45.5197	-59.8828	45.5195	-59.9228	0.085
2006	NED2006036068	44.3497	-58.7775	44.3567	-58.7388	0.080
2007	TEL2007745167	46.3902	-59.8822	46.4055	-59.8463	0.080
2007	TEL2007745168	45.8793	-59.7175	45.8603	-59.7362	0.080
2013	NED2013022217	45.5965	-60.0277	45.6185	-60.0002	0.078
2011	NED2011025187	45.2245	-59.8717	45.2122	-59.9085	0.074
2003	NED2003042053	44.4602	-58.1483	44.4893	-58.1475	0.072
2011	NED2011025177	45.0115	-60.4413	45.0255	-60.4058	0.068
2009	NED2009027117	44.6880	-60.5528	44.6675	-60.5425	0.066
2009	NED2009027162	45.1585	-58.0728	45.1387	-58.0767	0.066
2006	NED2006036020	45.9790	-59.3170	45.9852	-59.2907	0.065
2006	NED2006036012	45.2133	-60.4427	45.2325	-60.4105	0.064
2010	NED2010002017	44.5628	-59.7268	44.5837	-59.6962	0.061
2006	NED2006036038	45.2680	-58.8477	45.2703	-58.8208	0.060
2009	NED2009027114	45.2085	-60.6390	45.1843	-60.6613	0.060
2010	NED2010027028	42.7057	-65.9688	42.6765	-65.9728	0.060
2005	TEL2005633020	46.5325	-59.6533	46.5630	-59.6452	0.057
2014	NED2014018159	45.1302	-59.3845	45.1118	-59.4132	0.054
2005	TEL2005633059	44.5090	-58.3372	44.5362	-58.3523	0.052
2005	TEL2005633060	44.5867	-58.4765	44.6173	-58.4717	0.052
2008	TEM2008830138	45.9887	-59.4047	45.9638	-59.4157	0.052
2011	NED2011025258	44.1540	-58.8420	44.1618	-58.8827	0.052
2008	TEM2008830160	44.6452	-58.0573	44.6695	-58.0308	0.051
2003	NED2003042030	46.7052	-59.9083	46.7250	-59.8767	0.050
2006	NED2006036040	45.3477	-57.8048	45.3382	-57.7665	0.050
2006	NED2006036072	44.5793	-58.9852	44.6087	-58.9812	0.050
2010	NED2010027159	44.8870	-59.7403	44.8662	-59.7110	0.050

Table A5. Mission number and set and position information of DFO multispecies trawl survey records used to identify significant area polygons for *Flabellum* cup corals. Set number is the last 3 digits of the Mission Number and Set.

Year	Mission Number and Set	Start Lat. (DD)	Start Long. (DD)	End Lat. (DD)	End Long. (DD)	*Flabellum* Weight (kg)
2010	NED2010027218	42.9715	-61.5662	42.9905	-61.5347	0.522

2010	NED2010027234	42.6300	-63.3773	42.6418	-63.3032	0.521
2005	TEL2005633050	44.3053	-57.7317	44.2835	-57.7567	0.336
2005	TEL2005633111	42.9777	-61.6238	42.9662	-61.6598	0.324
2009	NED2009027091	42.7008	-64.0563	42.7182	-64.0257	0.322
2010	NED2010027231	42.3745	-64.0082	42.3837	-64.1860	0.254
2007	TEL2007745086	42.8732	-62.3370	42.8743	-62.3745	0.250
2006	NED2006030086	42.7287	-63.9498	42.7322	-63.9038	0.240
2010	NED2010027232	42.4682	-64.0270	42.4898	-63.9542	0.230
2012	NED2012022119	42.6973	-64.0700	42.7103	-64.0345	0.211
2009	NED2009027210	42.9292	-61.9673	42.9322	-62.0060	0.192
2010	NED2010027228	42.6547	-63.4120	42.6415	-63.4838	0.192
2009	NED2009027093	42.7567	-63.5690	42.7563	-63.5928	0.174
2005	NED2010027235	42.5523	-63.1947	42.5387	-63.2690	0.152
2005	NED2005034111	42.9858	-61.6217	42.9742	-61.6583	0.095
2007	TEL2007745087	42.9070	-61.8372	42.9063	-61.8748	0.090
2010	NED2010002048	44.9417	-56.7758	44.9662	-56.7995	0.090
2014	NED2014018205	44.2078	-58.0618	44.1897	-58.0665	0.088
2014	NED2014018097	42.8400	-62.8557	42.8462	-62.8175	0.063
2007	TEM2007686034	44.8157	-57.1183	44.7960	-57.0897	0.060