Fisheries and Oceans Canada | Pêches et Océans Canada
Ecosystems and Oceans Science | Sciences des écosystèmes et des océans

Canadian Science Advisory Secretariat (CSAS)

Research Document 2018/066

Maritimes Region

St. Anns Bank Framework Assessment

Jae S. Choi, Angelia S.M. Vanderlaan, Gordana Lazin,
Mike McMahon, Ben Zisserson, Brent Cameron, and Jenna Munden

Fisheries and Oceans Canada
Population Ecology Division
Bedford Institute of Oceanography
PO Box 1006, 1 Challenger Drive
Dartmouth, Nova Scotia B2Y 4A2

December 2018

Canada

Foreword

This series documents the scientific basis for the evaluation of aquatic resources and ecosystems in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Published by:

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

http://www.dfo-mpo.gc.ca/csas-sccs/
csas-sccs@dfo-mpo.gc.ca

ISSN 1919-5044

Correct citation for this publication:

Choi, J.S., Vanderlaan, A.S.M., Lazin, G., McMahon, M., Zisserson, B., Cameron, B., and Munden, J. 2018. St. Anns Bank Framework Assessment. DFO Can. Sci. Advis. Sec. Res. Doc. 2018/066. vi + 65 p.

Aussi disponible en français :

Choi, J.S., Vanderlaan, A.S.M., Lazin, G., McMahon, M., Zisserson, B., Cameron, B. et Munden, J. 208. Cadre d'évaluation du banc de Sainte-Anne. Secr. can. de consult. scient. du MPO, Doc. de rech. 2018/066. vi + 68 p.

TABLE OF CONTENTS

ABSTRACT

This analysis is intended to begin the dialogue required to develop a framework for monitoring and assessing spatially managed areas such as Marine Protected Areas (MPAs). In particular, we begin with the *Oceans Act* requirement to describe productivity, biodiversity, habitat and species of interest by identifying pragmatic metrics of these ecosystem-level attributes from pre-existing monitoring systems in the area of interest: St. Anns Bank in the Maritimes Region of Canada. We also identify a few human influences/pressures that are also readily quantified, namely fishing and vessel activity and some of the data gaps that are evident in the area. The nature of these data, and the steps required to use them properly, are made explicit in an open-sourced and revision-controlled environment (R and git) for the purpose of developing a transparent, vetted data and code system from which future assessment and modeling attempts can be staged. This should reduce the replication of effort in future. The rationale for the methods chosen to quantify the characteristic space and time scales of processes and features are identified and discussed. Ultimately, the intent is to develop approaches similar to the risk-based approach frequently encountered in fishery stock assessments, such that we can begin to express the "status" of an area or MPA. It is perhaps even possible to attempt to define reference points and obtain a better sense of the "health" of these areas, or at least assess the relative influence of area-based management closures. This first report will focus only upon the "Data" side of the issues and clarify the proposed "Methods".

INTRODUCTION

TERMS OF REFERENCE

The Health of the Oceans (2007) initiative and the National Conservation Plan (2014) support the conservation and restoration of lands and waters in Canada. In this context, the Science Branch of the Department of Fisheries and Oceans (DFO), has been tasked with developing a monitoring approach for Marine Protected Areas (MPAs) and, if possible, to assess their effectiveness in meeting their objectives.

SCOPE OF THIS REPORT

An Marine Protect Area (MPA) is defined in the *Oceans Act 35(1)* as, "An area of the sea that forms part of the internal waters of Canada, the territorial sea of Canada or the Exclusive Economic Zone of Canada and has been designated under this section for special protection for one or more of the following reasons:

- *conservation and protection* of commercial and non-commercial fishery resources, including marine mammals, and their habitats;
- *conservation and protection* of endangered or threatened marine species, and their habitats;
- *conservation and protection* of unique habitats;
- *conservation and protection* of marine areas of high biodiversity or biological productivity; and
- *conservation and protection* of any other marine resource or habitat as is necessary to fulfill the mandate of the Minister."

As a result, for the purposes of this report, we will likewise focus upon these key ecosystem attributes or components, namely: **productivity, biodiversity, habitat, and species of interest**. In reality, however, there are many other attributes or components of ecosystems that are also known to be important and relevant, including: ecological integrity and health, trophic structure and balance, ecosystem function, complexity, network structure, resilience, sustainability, as well as an open-ended number of species or life-history stages of the resident species. These other components will be touched upon where possible or necessary.

ST. ANNS BANK

The St. Anns Bank Marine Protected Area (herein, SAB) is an area of interest for eventual designation as an MPA. It is located east of Cape Breton Island, Nova Scotia, Canada (Figure 1). Previous advisory processes (DFO 2012; Kenchington 2013), identified the primary objectives of SAB as being to conserve, protect and, where appropriate, restore ecologically distinctive or significant areas and, overall, the ecosystem "health" of SAB. As in the *Oceans Act*, the focus was upon the above four ecosystem components: **productivity, biodiversity, habitat and species of interest**.

Other MPA goals were also expressed in DFO (2012) and Kenchington (2013), but these were less emphatic:

- *contribute* to the health, resilience and restoration of the Eastern Scotian Shelf ecosystem;
- *contribute* to the recovery and sustainability of commercial fisheries; and
- *promote* scientific research and monitoring to further understand and protect SAB.

Ford and Serdynska (2013) make more precise the ecological components that the SAB area of interest might help to protect and conserve, especially in the context of the definition of MPAs in the *Oceans Act*:

- commercial and non-commercial fishery resources, including marine mammals, and their habitats (e.g., habitat for Atlantic Cod, redfish, American Plaice, sea urchins, White Hake, Witch Flounder, sea anemones, sponges, and sea pens);
- Endangered or Threatened marine species, and their habitats (e.g., habitat for depleted species such as Atlantic Wolffish, Atlantic Cod, and Leatherback Turtles);
- unique habitats (it is the only major bank on the Inner Scotian Shelf); and
- marine areas of high biodiversity or biological productivity of invertebrates and fish.

OBJECTIVES

The **primary objectives** of this report are to:

- Develop an Assessment Framework that can
  - monitor the status of an MPA; and
  - assess the effectiveness of an MPA in meeting its conservation objectives.
- Identify data gaps and sources of uncertainty.

The method by which this can be accomplished, however, is anything but straightforward. This is primarily due to the fact that the SAB is:

- a large ecosystem and as such complex, operating at various space, time and organizational scales;
- connected in various ways to the surrounding environment and so cannot be treated as an isolated system; and
- measures of system components of interest, namely, productivity, biodiversity, habitat and species of interest, are ambiguous and imperfect at best, and usually non-existent or poor in information quality and/or quantity.

As such, we emphasize that this report is a simplistic first attempt at developing a general approach towards assessing MPAs. Indeed, given the above significant challenges, it is best viewed as a work in progress that will require further precision and improvement. To this end, we will, in the following, describe the data used for the assessment; outline the methods and assumptions associated with the modeling of this data; summarize the primary results of this analysis; provide some discussion of salient points; and conclude with general recommendations. The technical aspects of data Quality Assurance (QA)/Quality Control (QC), associated assumptions and methods are identified in the Appendices.

NOTE: The primary purpose of this document is the Data and Methods. Results and Discussions will be the focus of a subsequent paper.

DATA

In this section, we focus upon a description of the data chosen for inclusion in this assessment. The purpose of the section is to clearly identify the data, sampling design, and the associated assumptions and methods required/used to filter and integrate them in an informative manner.

For the sake of transparency, all data assimilation and QA/QC methods have been encoded in an open-sourced analytical environment R (R Core Team 2015) and made publicly accessible at GitHub (packages under development) and to permit flexible and adaptive multiuser contributions through the **Git** revision control system and a uniform data interface system. This approach permits the development of a coherent and vetted approach that is completely open source in nature. In this way, we see this project as a real and flexible structural and collaborative framework in the sense of a real scaffolding, to build a monitoring and assessment system that can be easily transferred to other regions, domains and mandates, as well as, fostering collaborations and communication with universities and the general public.

STUDY AREA

Evaluating MPA status and effectiveness in meeting conservation objectives, requires explicit reference to changes both within and without the area of interest. Even in the most basic BACI-type design, this requirement is explicit (Green 1979; Underwood 1992). For this reason, and also to facilitate evaluations of other potential MPAs in the region, a much larger surrounding area was chosen for analytical purposes. This area is the continental shelf region of Nova Scotia (Figure 2), bounded by latitudes 37°N to 48°N and longitudes from 48°W to 71°W. [Note: It should be emphasized that this will not alleviate problems associated with pseudo-(spatial, temporal) replication (Hurlbert 1984), although an assumption of a Gaussian process (see Methods) may potentially alleviate this problem.]

DATA SELECTION CRITERIA

Exhaustive surveys of available data have been compiled by Ford and Serdynska (2013). Their conclusions were that most biological data and environmental conditions are poorly sampled in the SAB area. The decision criteria for inclusion of data in this study were as follows:

- Part of an **on-going sampling** program. This is because the design principle behind this project is that the underlying assessment must smoothly transition into a routine monitoring approach into the future.
- Sufficient and regular **spatial** coverage (> 100 sampling locations) inside MPA and throughout the study area.
- Sufficient and regular **temporal** coverage (approximately annually, > 10 years) inside MPA and throughout the study area.
- **Informative** – high data quality that is in some manner related to productivity, biodiversity, habitat and species of interest.

The same decision criteria were applied to human usage data. The result was to include the following data streams for MPA characterisation:

- Atlantic Zone Monitoring Program (AZMP)/chlorophyll-a and nutrients: BioChem bottle data (Devine et al. 2014).
- AZMP/Zooplankton: BioChem database (Devine et al. 2014).
- Remote Sensing Data: ocean colour and SST (Remote Sensing Group).
- Groundfish: DFO's Groundfish Research Vessel surveys focus upon demersal fish species, since approximately 2000, upon invertebrates as well.
- Snow Crab survey, focus upon benthic invertebrates.

- Clam survey data in Banquereau and Western banks (though it does not pass the temporal coverage conditions, it offers very high resolution multispecies data on the banks).
- Temperature records: from various sources, especially, groundfish, Snow Crab and AZMP surveys.
- Salinity (Groundfish surveys/AZMP, BioChem).
- Oxygen and pH (once the data have been reloaded; Groundfish surveys/AZMP, BioChem).
- Bathymetry (CHS, Groundfish survey, Snow Crab survey).

To characterise human usage patterns, the following have been chosen for inclusion:

- Logbook records of catch and effort (MARFIS/ZIFF).
- AIS tracks – Radio-based Automatic Identification System.
- VMS potentially – Satellite-based Vessel Monitoring System.

DISCRETE BOTTLE DATA: CHLOROPHYLL-A AND NUTRIENTS

- Relevance: productivity, biodiversity, habitat and species of interest (in relative order).
- Sampling: AZMP surveys, Groundfish surveys, pelagic net tows and water profiles.
- Spatial coverage: variable no. stations, 143,499 records, 829 missions.
- Temporal coverage: 1955 to present, annual surveys.
- Source code: https://github.com/jae0/aegis/
- Discrete bottle data consisting of chlorophyll-a and nutrient records (nitrate, phosphate and silicate) were obtained by laboratory analysis of water samples collected at discrete depths. For this study all available nutrient and chlorophyll-a discrete bottle data were extracted from DFO's BioChem database for the study area. This dataset contains more than 500,000 records with the earliest records starting in 1955. After QA/QC, the discrete bottle data retained for analysis was comprised of 143,499 profiles, collected on 829 missions (Figure 3; Appendix 1).

The number of profiles available in each year (Figure 4) shows that there were few profiles taken until the mid-1960s, and a relatively steady number of yearly profiles after the initiation of the AZMP in 1999. The peak sampling during the period 1976-1982 corresponds to DFO's Scotian Shelf Ichthyoplankton Program (SSIP) and foreign research vessels sampling programs that were obtained from the National Oceanic and Atmospheric Administration's (NOAA) National Oceanographic Data Center (Pierre Clement, personal communication). Monthly distribution of the profiles (Figure 5) demonstrates that most of the data was collected in July (mostly during DFO's groundfish surveys), followed by the months of September and April. Note that spatial distribution of the sampling varies among months with most data collected on the Scotian Shelf in July and the fewest data in January (Figure 3 and Figure 6).

ZOOPLANKTON DATA

- Relevance: productivity, biodiversity, species of interest, habitat (in relative order).
- Sampling: AZMP surveys, groundfish surveys, pelagic net tows, 400 taxa.
- Spatial coverage: 2367 net deployments, 126 missions.
- Temporal coverage: 1999 to 2014, annual surveys.

- Source code: https://github.com/jae0/aegis/

Number of net deployments for each month is shown in Figure 7 and the corresponding spatial distribution of the net deployments are shown in Figure 8. Note that most of the net data were collected in July during summer groundfish survey missions and in April and October on AZMP spring and fall missions, while winter months contain mostly fixed station data (Halifax 2 and Prince 5). Abundance patterns are found in Figure 9. The QA/QC issues are identified in Appendix 1.

REMOTE SENSING DATA

Ocean Colour

- Relevance: productivity, habitat, biodiversity and species of interest (in relative order).
- Sampling: MODIS.
- Spatial coverage: 39°N to 62.5°N and 42°W to 71°W, resolution of 1.5 km.
- Temporal coverage: August 2002 to March 2015, 610 quarter-monthly (8-day) composite images.
- Source code: https://github.com/jae0/aegis/

Ocean colour refers to the spectral distribution of light emerging from the ocean that carries information about water constituents, particularly about biologically useful chlorophyll concentration in the surface layer. When measured from satellites, it provides unique synoptic view of chlorophyll spatial distribution over large areas of the ocean on a daily time scale.

The nominal uncertainty of chlorophyll products derived from ocean colour satellites is 35%, with better agreement with *in-situ* chlorophyll for the open ocean (Moore et al. 2008), while overestimation is often observed in the coastal ocean (Darecki and Stramski 2004). This is due to the inability of the algorithms to distinguish chlorophyll from suspended particulate matter and colored dissolved organic matter often present in the coastal waters as, for example, in the Bay of Fundy and Northumberland Strait.

Ocean colour satellite data for this study was provided by the Remote Sensing Unit (RSU) from the Bedford Institute of Oceanography (BIO) as 8-day composite chlorophyll images, which are routinely produced by the unit for the AZMP. The dataset was created using the Moderate Resolution Imaging Spectroradiometer (MODIS-Aqua) data, where the chlorophyll-a values are based on the 2012 reprocessing carried by the National Aeronautics and Space Administration (NASA) using OC3 chlorophyll algorithm. Composite images were created from daily Level-2 MODIS-Aqua files downloaded from NASA by averaging valid chlorophyll-a values for each pixel using all available daily images within that time period (Caverhill et al. 2015; Feldman and McClain 2015). The dataset is comprised of years 2002 to 2015 and area 39°N to 62.5°N and 42°W to 71°W, with spatial resolution of about 1.5 km per pixel.

Even though there is ocean color data available before the MODIS launch, it was decided to limit our dataset to a single sensor to avoid potential biases between the sensors. Due to the frequent cloud coverage of the Northwest Atlantic, it was decided to use 8-day composite images as daily images would not provide a sufficient number of valid pixels.

The values of chlorophyll-a pixels within St. Anns Bank polygon were extracted from each 8-day composite image and average chlorophyll-a concentration was computed for the polygon. The time series for the polygon and the associated climatology that was computed from time series data show characteristic spring blooms in March and April, with varying intensity and timing

throughout the years (Figures 10 and 11). An example of MODIS semi-monthly chlorophyll-a products showing spring bloom progression in the St. Anns Bank area in 2012 is shown on Figure 12.

Primary Production

- Relevance: productivity, habitat, biodiversity and species of interest (in relative order).
- Sampling: MODIS.
- Spatial coverage: 39°N to 62.5°N and 42°W to 71°W, resolution of 1.5 km.
- Temporal coverage: July 2002 to December 2014, 150 monthly images.
- Source code: https://github.com/jae0/aegis/

Marine primary production plays an important role in biogeochemical cycles, in food web dynamics, and in marine fisheries. It may be defined as the amount of organic material (or organic carbon) produced per unit area (or volume) per unit time by photosynthetic plants, predominately by phytoplankton.

Primary production of the ocean on synoptic scale is estimated by models that use satellite data (Sea Surface Temperature (SST), ocean colour chlorophyll, and available light estimates at the surface), shipborne *in-situ* information on the vertical distribution of phytoplankton in the water column, and the phytoplanktons photosynthetic response to light.

Monthly primary production data were provided by the Remote Sensing Unit at BIO that routinely generates production maps for the Northwest Atlantic as part of the AZMP. The general approach for the production computation is described in Platt et al. (2008) and employs chlorophyll and light estimates from MODIS, SST produced by the unit, and DFO's archive of shipborne measured parameters. This particular production algorithm has been validated with *in-situ* measured production (Platt and Sathyendranath 1988) and also has performed very well in global comparisons (Carr et al. 2006).

Primary production for each pixel within St. Anns Bank polygon was extracted from the monthly images, and average production was computed for the polygon. The time series for the polygon and the associated monthly climatology are showing peaks in primary production in spring and summer, with varying intensity and timing throughout the years (Figure 13 and Figure 14).

Sea Surface Temperature

- Relevance: productivity, habitat, biodiversity and species of interest (in relative order).
- Sampling: AVHRR.
- Spatial coverage: 39°N to 62.5°N and 42°W to 71°W, resolution of 1.5 km.
- Temporal coverage: December 1997 to March 2015, 845 8-day composite images.
- Source code: https://github.com/jae0/aegis/

Sea Surface Temperature (SST) from space was estimated using infrared channels of the Advanced Very High Resolution Radiometer (AVHRR) on board the polar-orbiting satellites.

The SST data for this study was provided by the Remote Sensing Unit from the BIO that has been downlinking AVHRR data from NOAA satellites since 1997 on an L-band satellite receiver that resides on the roof of the BIO. They routinely process the received data and supplement it with data stream from the AVHRR onboard the European satellites. Composite SST images of

different periods are operationally produced as part of the AZMP. Here we used 8-day composite images with the same spatial coverage and spatial resolution as the ocean colour chlorophyll data.

The SST pixels within St. Anns Bank polygon were extracted from each 8-day composite image and average SST was computed for the polygon. The time series for the polygon and the associated climatology that was computed from time series data are shown on Figures 15 and 16. An example of semi-monthly SST products corresponding to the spring bloom progression in the St. Anns Bank area in 2012 is shown in Figure 17.

BOTTOM TEMPERATURES

- Relevance: productivity, habitat, biodiversity and species of interest.
- Sampling: Groundfish survey, Snow Crab survey, AZMP profiles.
- Spatial coverage: full extent, varied sampling.
- Temporal coverage: 1950 - present (more historical data present but coverage is variable).
- Source code: https://github.com/jae0/aegis/

Numerous data sources have been compiled by Ocean Sciences Division, DFO. The data were QA/QC controlled and then modeled in a two-part process, temporal (first order harmonic analysis) and then spatial interpolation as indicated in the Methods (Figure 18).

DEMERSAL FISH AND MACRO-INVERTEBRATES

- Relevance: productivity, habitat, biodiversity and species of interest.
- Sampling: Groundfish survey, Snow Crab survey.
- Spatial coverage.
    - Groundfish: full extent, random stratified, variable number of stations.
    - Snow Crab: Colder water environment, geostatistical grids of approximately 10 minutes and approximately 400 stations.
- Temporal coverage.
    - Groundfish: 2000 - present (started in 1970, but consistent sampling since 2000).
    - Snow Crab: 2005 - present (started in 1996, but consistent sampling since 2005).
- Source code:
    - Groundfish: https://github.com/jae0/aegis/
    - Snow Crab: https://github.com/jae0/bio.snowcrab/

Groundfish Survey

The Groundfish survey (Figure 19, left) utilizes a Western II-A Otter Trawl net with a wingspread that is **assumed** to be 12.5 m and a target distance of 1.75 nautical miles (3.24 km) and/or a 20-30 min tow. This net was used from 1982 to the present. Prior to this period, a Yankee 36 ft trawl was used with unmeasured net configuration data. It operates night and day. Sampling occurrence as a function of time and season are shown in Figure 19 (right). The consistent identification of invertebrates in this survey began in approximately the year 2000. All species assemblage analyses will use data from the post 2000 period.

Net Mensuration

Sensors measuring trawl net configuration and state are ubiquitous in modern surveys and commercial fishing practices. In the Snow Crab survey, net configuration has been monitored and used to determine swept area manually (since 1996) and also with an automated procedure (since 2004). Unfortunately, in the groundfish survey, this information is ignored, even though the survey series is a major source of information for many stock assessments. Net configuration data have been collected in the groundfish surveys sporadically since 2004 (Figure 20). However, swept area estimates have never been directly computed. Instead, the catch data is used almost invariably with the assumption that each tow is equivalent in swept area (a "standard tow"). Alternatively, it has also been used by some assessments by "adjusting" catch data by the linear distance from "start" and "end" times and locations and so implicitly assuming the net configuration to be a constant (Don Clark, personal communication). Both approaches are problematic for the reasons identified below.

To address these important and incorrect assumptions, algorithms were developed to automatically determine lift off and touch down times, locations and net width (Munden, J., and J.S. Choi. (in prep.) Calculating Swept Area for the Maritimes Region Trawl Survey). Based upon random visual inspection, the skill was determined to be reasonable, with >90% of the cases having estimates within 15 seconds of visual determination of touch down/lift off locations. Where bias was observed, this was mostly to underestimate total contact time due to an over-smoothing of the lift off or touch down profiles.

From this re-analysis, the actual time and distance of bottom contact was found to range from approximately 10 to 40 minutes and distances from 1.75 to 6 km. Historical studies assumed a "standard tow" of 3.24 km in length. Even when compared to the "length" of a tow determined from positions recorded of "start" and "stop", times or locations were frequently in error relative to actual locations of net touch down and lift-off determined from the Global Positioning System (GPS) information (Figure 21).

Net structure also varies along tows and between tows as it encounters differing substrate, bathymetric and hydrodynamic conditions and vessel speeds, currents, surface sea state, and net fullness/filtration efficiency due to contact with rocks, boulders and mud. Assuming net width to be a constant at 12.5 m is, therefore, problematic (Figures 22 and 23).

When the variable nature of the length and width of survey tracks are both accounted for, the potential error in swept area estimates are evidently large (Figure 24). Indeed, the range in variation of swept area was approximately the same in magnitude as the swept area of a standard tow. This means that an unnecessary and large "observation error", potentially as large in magnitude as the magnitude of the catch in a set, may be needlessly introduced to survey indices. Further, swept area estimates based upon standard tows were also biased down relative to estimates based upon net mensuration and actual tow tracks, meaning that there is a high probability of over-estimating catch densities.

Unfortunately, net mensuration data is still not consistently recorded nor used in the groundfish surveys (Figure 20). As a result, it is not possible to satisfactorily estimate swept area for all historical sets. Statistical methods of recalibration were used to impute swept area in these latter sets (Figure 24), using GAM-based estimates based on relationships with location, depth and salinity (R^2 ranged from 40 to 60%, depending upon availability of covariates). However, due to issues with a significant bias detected in wingspread sensors since 2013 (Figure 23), this approach **cannot** be extended for the period post-2013.

To conclude this section, the observation error (uncertainty) associated with this data series, used in so many research programs, is high and biased high. Removal of these errors by a

coherent representation of net configuration is simple to address and a major data gap that needs to be addressed before this data series can be used quantitatively.

Snow Crab Survey

The Snow Crab survey uses a Bigouden Nephrops trawl (Figure 19), a net originally designed to dig into soft sediments for the capture of lobsters in Europe (headline of 20 m, 27.3 m foot rope mounted with a 3.2 m long 8 mm chain, with a mesh size of 80 mm in the wings and 60 mm in the belly and 40 mm in the cod-end). Tows were conducted for approximately 5 minutes in duration with duration of bottom contact being monitored by Netmind and Seabird sensors. The width of the mouth of the net was also monitored with Netmind sensors. The ship speed was maintained at approximately 2 knots. The warp length was approximately 3 times the depth. Positional information, as well as, water temperature measurements, was collected using a global positioning system and Minilog and Seabird data recorders. The surface area swept by the net was calculated from swept distance and net width information.

Supplemental SAB Stations

The 2015 Snow Crab trawl survey increased sampling in the St. Anns Bank area to provide additional information about the marine macro-fauna. Fourteen stations (14) in, and adjacent to, the proposed MPA location were included in this additional sampling. These locations were close to previous Snow Crab survey stations and represent varied depths, bottom-types and habitats. The species composition of the catch at these stations varied greatly, as expected with differences in depth and bottom type. The sampling at these stations included:

- all catch identified to species level;
- all species counted and weighed to a tenth of a kilogram;
- all finfish and crab species measured and weighed individually; and
- stomach samples taken from finfish for diet studies.

An overview (in Google Earth format) of the latest catch and sampling at these stations can be found at: http://www.enssnowcrab.com/mpa/mpatows.kmz.

FISHERIES ACTIVITY

- Relevance: productivity, habitat, biodiversity and species of interest.
- Sampling: MARFIS and ZIFF.
- Spatial coverage: full extent.
- Temporal coverage: 1999 – present.
- Source code:
  - Net Mensuration: https://github.com/jae0/netmensuration
  - Data Wrangling: https://github.com/Maritimes/Mar.datawrangling/

Commercial-fishing activities can modify the habitat and ecosystem and contribute to changes in the structure and functioning of exploited marine communities. Fishing impacts can be direct, such as the reduction of targeted and non-targeted species, as well as truncations in age and size distributions. Other direct effects due to fishing activities include habitat alterations and substrate modifications through interactions with fishing gear. Fishing can also cause indirect impacts via changes in foodweb structure within an ecosystem.

Direct impacts of commercial fishing can be measured using data from the Marine Fish (MARFIS) database that provides information on commercial-fishing activities. For most fisheries, the fishing position, gear type, catch per unit effort, and estimated weight of catch by species information is available from the database. The MARFIS database details information for all fishing trips where a landing is reported within the DFO Maritimes Region and includes data from 2002 through 2015. The exploitation of marine species and entanglement threat to cetaceans and sea turtles will be quantified from data derived from the MARFIS database.

Trawling and dredging disturbances to the sea floor will require different estimation techniques. Vessel Monitoring System (VMS) point locations have been used to estimate fishing-effort distributions (e.g., Lee et al. 2010) and to estimate impacts on the seabed (Gerritsen et al. 2013). Similar techniques can be developed within R to estimate the impacts to benthos from trawling and dredging activities within the St Anns Bank area.

The St. Anns Bank Area of Interest has four zones within it with various levels of fishing restrictions (Figure 1). Commercial fishing is restricted in Zone 1, the largest area, with the exception of the seal harvest. The MPA requires monitoring to ensure that fishers are complying with these regulations. Monitoring could be done through a variety of techniques and data sources, including data reported in the MARFIS database and VMS data that allow the direct monitoring of fishing activities. Automatic Identification System (AIS) data may also be used to monitoring fishing activities if the vessel is large enough to require AIS (≥ 300 gross registered tonnes) or if a fishing vessel has voluntarily installed an AIS system onboard.

MARFIS Data Extraction

Prorated landings for all species from 2002 onwards were extracted from MARFIS. The proration process distributes the actual reported weights across the reported efforts as identified within the fishers' logs.

In addition to the landings, we also included several other forms of catch so as to better reflect the removal of biomass. These catches were identified by their CATCH USAGECODE and include biomass used as bait or discarded (sometimes identified as dead). Live discards were not included. These catches were self-reported in a variety of units, so they have been converted to kilograms as necessary to match the logged landings.

Most log records included the spatial location of the catch, and some catches have multiple sets of coordinates available within the table MARFISSCI, LOGEFRTSTDINFOID, ENTLATITUDE and ENTLONGITUDE are physically entered into the logbook by the fisher, and they were used preferentially over DETLATITUDE and DETLONGITUDE, which are determined from other sources (like Loran-C).

Many logs have no usable coordinates since they are either left blank or are clearly incorrect (i.e. on land). Rather than discarding this data, we still extracted it and attempted to account for this biomass in the next section.

MARFIS Data Quality Control and Aggregation

Quality Control (QC) occurs in 2 stages and is quite simple. First, all records without coordinates are identified and retained, but they are removed from the main dataset. Next, all remaining data are compared against a high resolution coastline (the same as is shown in Figure 2), and those points falling on land are identified and retained but removed from the main dataset.

The remaining data are considered good and are aggregated. The aggregation level is user-defined. A scale of 3 minutes is used by default since it is an even division of a degree, with no

potential for rounding errors. The aggregation process outputs a single record with a single position for all of the catch in the area.

Following aggregation, the proportion of the total catch attributable to each cell is calculated. The data that failed the QC tests are then summed into a single value, representing the total catch that cannot be positioned. These data are then added to all of the cells in the same proportion as was calculated in the previous step.

For example, one cell might have a total catch of 3269.7 kg, and this cell represents 0.002167 of the total catch. If there are 5954.1 kgs of un-positioned data, then the corrected value for the weight attributed to this cell would be calculated as: 3269.7 kg + (55954.1 kg * 0.002167) = 3390.9 kg.

An example of the annually aggregated (2010) reported commercial catches for sea scallops (*Placopecten magellanicus*) is presented in Figure 25. Nominal catches ranged to 14,073.8 kg in Bay of Fundy, Georges Bank, and the Scotian Shelf.

Similarly, an example of annually aggregated (2011) reported commercial catches for Atlantic Halibut (*Hippoglossus hippoglossus*) is shown in Figure 26. Nominal catches ranged to approximately 9626.9 kg in the same area.

VESSEL ACTIVITY

- Relevance: habitat, biodiversity and species of interest.
- Sampling: AIS.
- Spatial coverage: Global for satellite AIS, coastal (approximately 100 km) for Canadian Coast Guard terrestrial AIS network.
- Temporal coverage: 2013 – present.
- Source code: https://github.com/jae0/aegis/

Commercial shipping can have various direct and indirect effects on an ecosystem. Direct effects including the contamination of the ecosystem from the discharge of contaminates, radiated underwater noise, the introduction of aquatic invasive species, and vessel-strike risk to marine mammals and sea turtles. Spatial-temporal data on vessel traffic is needed to determine the probability and/or magnitude of these effects on ecosystems. Such information is available through AIS data.

The International Maritime Organization (IMO) requires AIS transponders on all international vessels ≥ 300 gross tonnage and all passenger vessels. Many studies have used AIS data to examine risk of lethal vessel collisions to large whales (e.g., Vanderlaan and Taggart 2009; Wiley et al. 2011; Redfern et al. 2013; Guzman et al. 2013) or to assess and monitor ship noise and assess the impact on marine mammals (Hatch et al. 2008; McKenna et al. 2012; Hatch et al. 2012; Merchant et al. 2014). Similar exercises can be undertaken in the St. Anns Bank area with AIS data.

Fisheries and Oceans Canada (DFO) has access to two different sources of AIS data. The first is from the Canadian Coast Guard (CCG) terrestrial system that was developed to track and monitoring coastal shipping and provides a real-time, continuous stream of AIS vessel positions. Archived historical data from this system is available for January 2012 through December 2015, and data from 2016 are currently streaming and archiving to a server within DFO. Decoding routines have been developed using native R methods for these data. Both sources of data provide dynamic and static data, where the dynamic data includes information on vessel

identity, speed, and location, and static data includes information vessel identity, name, size, and type.

The CCG terrestrial system have 21 AIS coastal receiving stations in the Maritime region and 19 AIS coastal receiving stations in Newfoundland and Labrador. These receiving stations have limited range for detecting vessels (Figure 27) as AIS transmission detectability are primarily a function of the receiving tower height above sea level and the height of the AIS antenna on the transmitting vessel. AIS data is transmitted via Very High Frequency (VHF) marine radio on two channels (161.9765 and 162.025 MHz). Based on the height of the associated towers and a vessel with an AIS antenna 100 m high, line of sight calculations for VHF provide a reception range of 57 to 113 m in and around St. Anns Bank (Figure 28). However, there are several other factors that will contribute to transmission range, including weather conditions. Simard et al. (2014) estimated that coastal antennas within this network generally provide a reception range of 100 km (Figure 29). In either case, the terrestrial network is insufficient to monitor vessels across the entire AOI and just north of the AOI. These data can be combined with satellite AIS data.

The satellite AIS data are available globally for the years 2013, 2014, and 2015 with on-going data collection for 2016. Although satellite AIS data coverage is global, data are limited temporally as large sections of vessel transits are unavailable due to a limited number of satellites (n = 8) and their orbital paths (see Figure 30). Spatial interpolation must be completed to fill in missing data. Spatial interpolation is achieved using an A* function (Hart et al. 1968) that estimates the minimum cost to get from one point to another based on a cost map. Using seasonally aggregated annual density distributions of satellite AIS data for the years 2013 through 2015 (Figure 31), cost maps were estimated (Figure 32). Grid resolution for this analysis was initially set to 0.01 degrees and, within each grid square, the number of unique vessels identified by the vessel's Maritime Mobile Service Identity (MMSI) number was counted daily and summed through time. Cost maps were estimated quarterly. Two different cost maps were developed for the St. Anns Bank area to interpolate vessel transits north of Cape Breton into and out of the Gulf of St. Lawrence. Interpolation was heavily influenced by the ferries transiting between Cape Breton and Newfoundland and, therefore, a cost map was developed without the data derived from these ferries. A bathymetric restriction can also be built into the cost maps.

DATA GAPS

Feeding Relationships – Stomach Database

Bottom trawl surveys have been conducted by DFO on the Scotian Shelf annuallt since 1970 using a stratified random design. Sampling protocols changed in the late 1990s with the focus changing from commercially important finfish species to more comprehensive ecosystem monitoring that included the sampling of macro invertebrates (Tremblay et al. 2007). Stomach contents samples were collected from finfish using a length-stratified sampling protocol. Prey were quantified by weight and number, and they are often identified to the genus or family level, or to the species level when possible.

For the purpose of determining any change in the diet of finfish, pre- and post-implementation of the MPA on St. Anns Bank, or if there were differences within the area compared to other areas on the Scotian Shelf, we explored the stomach database to determine if these dietary differences could be described and detected. Within the database, 54% of the prey number observations were missing. Due to the large interannual variation in prey weights, estimating the prey number consistently from non-missing data where both the prey number and weight was available was determined to be impracticable and unreliable. Due to the sampling stratification

both by depth and length classes, it was further determined that there would be insufficient samples available to reach the asymptote of prey species accumulation curves (Cook and Bundy 2010) and total diet composition could not be detected. Prey species identified in the stomach samples could be used for quantifying biodiversity and species richness (Cook and Bundy 2012) for the proposed MPA and comparing it to similar ecosystem or pre- and post-implementation.

Other Ecosystem Metrics

On the side of ecosystem characterization, data gaps in the following are evident: pelagic fish (small and large bodied) and invertebrates (e.g., squid, jellyfish), substrate characterisation via multibeam surveys, marine mammals, reptiles, birds and genetic diversity. They are gaps in that they are expensive and/or difficult to monitor and/or contain information that is not readily available at present.

Other Human Usage Metrics

A large number of variables and ecosystem descriptors are being ignored. In particular, these include human influences such as seismic activity, pollution, ballast water, etc. They may be addressed once the basic biological features have been fully addressed.

METHODS

All methods have been implemented in R, an open-sourced programming environment. The methods are shared via GitHub using the **git** revisioning system. These architectural choices were adopted to enhance the transparency and ease of sharing and collaborating with all interested parties. It is structured such that any additional data series can be easily added to the system to permit adaptive change. In this way, the approaches developed represent a true structural **framework** in which to further develop methods and approaches.

The main methods used/developed in this report will be described in this section.

BIODIVERSITY AND TAXONOMIC RICHNESS

Biodiversity is seemingly a simple concept that is, in fact, fundamentally complex. This is because it ranges in focus from genetic and phenotypic variations within a local population, to breeding populations, to biome or even larger scaled genetic and phylogenetic and community variability, both in terms of their number and relative dominance. Any and all of these aspects of biodiversity can be estimated. However, it is the number of unique kinds of organisms found in a given location (commonly called taxonomic richness) that is most readily quantified and monitored.

Taxonomic richness is known to increase asymptotically with sampling intensity. As such, a statistical correction ("rarefaction") for spatial and temporal sampling intensity must be applied to be meaningfully comparable across locations and time. Specifically, a simple regression model was used to predict an expected richness *R* at a standard surface area and time depth:

$$R \sim \text{Lognormal}((\beta + \alpha_s \log(\text{SA}) + \alpha_t \log(\text{TS})), \tau)$$

The other terms are β a constant, SA surface area (ranging from 1 to 50 km, radial length scale), TS, the number of years entering into the count (ranging from 0 to 5 years) and τ is a lognormal error. The autocorrelation in SA and TS are ignored for the present but will eventually be modeled as well via a Poisson process.

The intent is to model the spatial/temporal patterns and then integrate them in a risk-based, probability model to permit formal statements of risk and probability of exceeding thresholds.

PRODUCTIVITY

Total system standing biomass is generally used as a proxy for productivity. They are not the same; however, they will be used to describe aggregate abundance of: various categories of organisms such as total bottom biomass, macroinvertebrates, zoopankton, phytoplanton, chlorophyll-a, etc.

To estimate true production, a modeled approach is necessary. These indices can then be coupled with spatially explicit total landings to estimate the biomass and secondary production associated with the biota and fishery exploitation/footprint.

The intent is, therefore, to model the spatial/temporal patterns and then integrate them in a risk-based approach to permit formal statements of risk and probability of exceeding thresholds and perhaps even estimate net production.

HABITAT

The basic Hutchinsonian notion of "ecological niche" is closely tied to our current understanding of "habitat". It is a multi-dimensional concept in that it incorporates an undefined set of environmental variables and the associated biological constraints/specialisations/requirements (e.g., nutrients, thermal, oxygen, pH, etc.) pertinent to an organism of interest.

Two notions of habitat can be discriminated, depending upon outlook:

- Functional (H_f) – make increasing more precise habitat definitions by adding more environmental and biological factors for increasingly more precise categories of organisms.
- Integrative (H_i) – the biota living in a given time and location represent a full integration of all relevant environmental and biological factors at their proper space time scales and so, in effect characterises the full system-level concept of "habitat space".

Functional-habitat Modeling

A utilitarian way of describing the **Functional-habitat** (H_f) space of an organism is to examine the presence-absence or relative abundance of organisms as a function of environmental gradients and biological/life history constraints. From such information, the likelihood of a given location to be potential habitat for an organism of interest can be derived. The most robust method is to develop a probability model under the assumption that the presence or absence of an organism is a Bernoulli process. This can be readily parameterised using standard Generalized Linear Models. However, as environmental constraints are almost always modal in influence given a wide enough environmental gradient, a nonlinear model is more useful. Generalized Additive Models and Stochastic Partial Differential Equation Models are two methods that can deal with these environmental constraints in a simple and efficient manner (Choi 2010; Appendix 3).

The utility of such an approach is most relevant for organisms with highly specific habitat requirements. They are used in this framework. The intent is to develop such Functional-habitat models for key species of interest: wolffish, Cod, etc., to assess changes in their available "habitat". Examples are provided from those derived in the Snow Crab assessment (see Results). Though this Functional-habitat concept itself does not exclude species interactions, in actual practice, they are generally ignored as they make the statistical estimation impossibly over-parameterized (but see Choi et al. 2012, for one possible solution). Further, as there will

always be factors that are either poorly known, poorly sampled, or poorly parameterized, (e.g., dissolved oxygen, pH, redox, bacteria, jellyfish, squid, pollution, substrate type, etc.), these models will always necessarily be incomplete.

It is relatively straightforward to model the spatial/temporal patterns and then "integrate" them in a risk-based approach to permit more formal statements of risk and probability of exceeding thresholds.

Integral Habitat – Whole System Level

While the Functional-habitat concept is interesting and pragmatic, it is decidedly a reductionistic perspective. The monitoring and assessment requirements for an MPA also demands a whole system (phenomenological) perspective. The *assumption* we make is that **the relative abundance of organisms found in a given location and time defines and mirrors the kind of habitat in which they live**. Sessile organisms that require high flow environments and associated biota tend to exist and flourish with a given group of other organisms similarly adapted, and they are different from those that require cold waters and minimal water flow, etc.

Thus, if we can quantify the observed species composition in a given location and time, we would, in effect, be describing the habitat. We will define this as **Integral-habitat** (H_i): species assemblage information that directly reflects all biological and environmental interactions simultaneously, both measured and unmeasured, and some too fleeting to be measurable.

Fortunately, these associations are readily quantified using multivariate methods of data ordination. Here we focus upon Principal Components Analysis, which focuses upon an eigenanalysis of correlational structure of the species assemblages. An example derived from the Snow Crab assessment is provided in the Results.

The intent is to model the spatial/temporal patterns and then integrate them in a risk-based approach to permit formal statements of risk and probability of exceeding thresholds.

CONNECTIVITY: SPACE AND TIME SCALES

Spatial Scale

Marine Protected Areas exist in a spatial context. The characteristic spatial scale of productivity, diversity and habitat found in the MPA will determine which processes will be relevant to these aspects of an MPA. If the spatial variations in the productivity of a species of interest is small relative to the size of an MPA, the chances of the MPA having an influence upon the species is enhanced. This is usually the case when short-range processes dominate (e.g., less mobile species, weakly dispersing, low currents, habitat heterogeneity at small scales). If, however, the spatial scale is larger than the MPA, then it would mean that broader/larger processes were influencing the productivity of the species (e.g., higher mobility or dispersal processes/current, and stronger spatial connectivity, habitat heterogeneity at larger scales) – resulting in a lower likelihood of the MPA having an influence upon the species or components of interest.

A second important factor is the relationship of the characteristic spatial scale to monitoring. As organisms exist at a given spatial scale in a given area, a sampling/monitoring protocol must reference/address these spatial scales. For example, when a spatial feature (e.g., biodiversity) demonstrates short characteristic spatial scales (i.e., a lot of spatial structure at smaller scales), any sampling approach must respect this and similarly operate at such shorter scales or even smaller, if one is to be able to resolve the patterns and describe properly the subject of interest. Similarly, if a feature (e.g., biodiversity) is long-ranged and one wishes to resolve the patterns properly, then a sampling protocol must be similarly long-ranged to resolve the pattern. A

sampling program much smaller than the characteristic spatial scale would be beneficial, but the accrued benefits relative to cost of sampling would be rapidly diminishing. In that time, effort and resources requirements generally increase more rapidly than any benefit, (e.g., in the simplest case, if one is looking only naively at standard error as a measure of benefit, then it would increase asymptotically with increased effort with a power of $-1/2$.

For these fundamental reasons, defining the spatial scale of a given observation or process is imperative for the development of any assessment or monitoring of MPAs. To this end, we represent any spatially explicit observation as $Y(\mathbf{s})$, which are measured in a coordinate space $\{\mathbf{s} \in S \in \Re^d\}$ and domain S of dimensionality d. In this framework, we will mainly focus upon the case of $d = 2$ spatial dimensions (e.g., longitude and latitude or northing and easting). The observations $Y(\mathbf{s})$ are assumed to be realizations of a **spatial stochastic process**, y, that is, some latent, unobservable but real stochastic generative function. These are also known as (spatial) "Random Fields" in the literature. The manner in which the variability of y changes as a function of distance, $h = \| \mathbf{s} - \mathbf{s}' \|$, is known as the spatial autocorrelation function. This spatial dependence is highly informative in that it defines how the similarity of observations changes with distance and so ultimately defines **spatial scale**.

The spatial model is succinctly expressed as a regression model of a stochastic process (Banerjee et al. 2014):

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + e(\mathbf{s})$$

where, the observations $Y(\mathbf{s})$ are a function of some mean process $\mu(\mathbf{s})$, and a residual error process $e(\mathbf{s})$. The latter are further defined as:

$$\begin{aligned} \mu(\mathbf{s}) &= x(\mathbf{s})^T \beta \\ e(\mathbf{s}) &= \omega(\mathbf{s}) + \varepsilon(\mathbf{s}) \end{aligned}$$

where $x(\mathbf{s})$ are spatially referenced predictors with associated parameters β; and the residual error process is decomposed into spatial $\omega(\mathbf{s})$ and nonspatial $\varepsilon(\mathbf{s})$ components. The latter is also known as "nugget" error in geostatistics and represents error associated with measurement and/or microscale variability/processes.

The error structures are usually assumed to be the following:

$$\begin{aligned} \varepsilon(\mathbf{s}) &\sim \mathrm{N}(0, \tau^2) \\ \omega(\mathbf{s}) &\sim \mathrm{GP}(0, C(\mathbf{s}, \mathbf{s}'; \boldsymbol{\theta})) \\ \mathbf{Y} &\sim \mathrm{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \end{aligned}$$

The nonspatial error is assumed normal with mean 0 and standard deviation τ. The spatial error is assumed to follow a Gaussian Process with mean 0 and covariance $C(\mathbf{s}, \mathbf{s}'; \boldsymbol{\theta})$, that is, a *spatial covariance function* with parameters $\boldsymbol{\theta}$. A multivariate normal likelihood is usually assumed for the observations $\mathbf{Y} = (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))^T$, with mean $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ and covariance $\boldsymbol{\Sigma} = \left[C(\mathbf{s}_i, \mathbf{s}_j; \theta)\right]_{i,j=1}^{n} + \tau^2 I_n$. The matrix of regressors is $\mathbf{X} = [\mathbf{x}(\mathbf{s}_i)^T]_{i=1}^{n}$ and I_n is an identity matrix of size n.

The spatial covariance function $C(h) = C(\mathbf{s}, \mathbf{s}')$ expresses the tendency of observations closer together to be more similar to each other than those further away; $h = \| \mathbf{s} - \mathbf{s}' \|$ is the distance separating observations. Historically, a number of different forms have been used. The most frequently used forms include (for $h > 0$):

$$
\begin{aligned}
C(h)_{\text{Spherical}} &= \begin{cases} \sigma^2(1-\frac{3}{2}h/\phi+\frac{1}{2}(h/\phi)^3); & 0<h<=\phi \\ 0; & h>\phi \end{cases} \\
C(h)_{\text{Exponential}} &= \sigma^2(\exp(-h/\phi)) \\
C(h)_{\text{Gaussian}} &= \sigma^2(\exp(-(h/\phi)^2)) \\
C(h)_{\text{Powered exponential}} &= \sigma^2(\exp(-|h/\phi|^p)) \\
C(h)_{\text{Matérn}} &= \sigma^2\frac{1}{2^{\nu-1}\Gamma(\nu)}(\sqrt{2\nu}h/\phi)^\nu\ K_\nu(\sqrt{2\nu}h/\phi).
\end{aligned}
$$

At zero distance, $C(0) = Cov(Y(s), Y(s)) = Var(Y(s)) = \tau^2 + \sigma^2$ (*i.e.*, global variance), where τ is the nonspatial error, σ is the spatial error, and $\theta = \{\phi, \nu, p, \ldots\}$ are function-specific parameters including ϕ the *range* parameter. $\Gamma(\cdot)$ is the Gamma function and $K_\nu(\cdot)$ is the Bessel function of the second kind with smoothness ν. The latter, Matérn covariance function is frequently used as the shape of this function is more flexible (Figure 33), albeit at the cost of an additional parameter (Appendix 2).

The semivariance, more commonly used in geostatistics, is the covariance reflected on the horizontal axis of the global variance: $\gamma(h) = C(0) - C(h) = \frac{1}{2}\ \text{Var}[Y(\mathbf{s}) - Y(\mathbf{s}')]$.

The **spatial autocorrelation function** is defined as the covariance function scaled by the global variance: $\rho(h) = C(h)/C(0)$. For the purposes of this framework, we will define the **spatial scale** to be the distance at which the spatial autocorrelation decreases asymptotically to $\rho(x) \to 0.05$ ("practical" range). This will be determined in modeled form where possible and, when it is not possible, an empirical estimate will be made derived from the empirical semivariogram. We focus upon spatial patterns larger than 1 km in resolution and smaller than the size of study domain, the "Scotian Shelf Ecosystem" (SSE).

Temporal Scale

Marine Protected Areas also exist in a dynamic frame. As such, similar to the above spatial considerations, there also exists some characteristic temporal scale upon which an MPA and its subcomponents operate. If the overall temporal variations of the biota and environment is small relative to the overall life of an MPA, the chances of the MPA having an influence upon the species is enhanced. Presumably, the longevity of MPAs will be long lasting and so will guarantee some influence, however small that may be; an effect that would be enhanced if the subject is shorter-ranged in spatial scales.

Again, similar to the spatial scale case, this also has a simple and obvious implication in terms of monitoring and assessment. Short-range variations require higher sampling effort to resolve/understand the issues and vice-versa.

As the temporal scale is an informative metric for monitoring and assessment of an MPA, we must be precise in its definition. Similar to the spatial case, we focus upon how the correlation and variability of some quantity changes with greater difference in time. The analogue to the semivariogram in a time series context is known as a **cumulative periodogram**.

A periodogram expresses the amount of variance found at different wavelengths (ω). It is a discrete sample estimate of the continuous concept of spectral density, $\gamma(t)$:

$$\gamma(t) = \int_{-1/2}^{1/2} e^{2\pi i\omega t} f(\omega) d\omega.$$

It is easily obtained from a Fast Fourier Transform of any arbitrary time-series and so the cumulative distribution permits a rapid identification of the time scale at which correlation drops to some arbitrary level. To be approximately comparable to the spatial scale, we define the temporal scale as the time difference by which the Cumulative Power Spectral Density increases to 95% of the total variance.

If the goal is to resolve short-term processes, then sampling must also, of necessity, be more frequent. However, similar to spatial scale issues, there is a point where there will be diminishing returns for any increase in the resolution of a temporal signal. It is the intent of this framework to operate upon timescale of 1 year or greater. Sub-annual signals, where they are available, would be used to decompose the seasonal signals from the inter-annual signals to avoid bias due to discretization errors.

Space-time Models

In reality, spatial and temporal patterns coexist and evolve. They are correlated processes and, as such, a challenge to model properly ("nonseparable" – see below). This renders the independent treatment and estimation of autocorrelation in time and space problematic. Nonetheless, new developments in computational methods are bringing such models within range of use for a framework such as this. This is primarily due to efficient methods associated with numerical modeling of Stochastic Partial Differential Equations (SPDEs), especially in spectral (Fourier) space.

Again, we follow Banerjee et al.'s (2014) development of spatio-temporal models as a simple extension of the spatial regression model. The observations, $Y(\mathbf{s},t)$ are measured in a coordinate space $(\mathbf{s},t) \in S \in \Re^d \times \Re$ in the domain S of dimensionality $d+1$. The space-time regression model can then be specified as:

$$Y(\mathbf{s},t) = \mu(\mathbf{s},t) + e(\mathbf{s},t),$$

where, $\mu(\mathbf{s},t) = x(\mathbf{s},t)^T\beta(\mathbf{s},t)$ is the mean process and $e(\mathbf{s},t)$ the residual error. The parameters β of the spatially referenced predictors x can have variable forms:

- $\beta(\mathbf{s},t)$ – varying in both time and space.
- $\beta(\mathbf{s})$ – spatially varying (fixed in time).
- $\beta(t)$ – temporally varying (and fixed in space).
- β – completely fixed (no variation in time and space).
- $\beta_s(\mathbf{s}) + \beta_t(t)$ – complex, hierarchical (mixed).

The error process can be separated into a spatiotemporally structured component ω and an unstructured component εs:

$$e(\mathbf{s},t) = \omega(\mathbf{s},t) + \varepsilon(\mathbf{s},t).$$

The *unstructured* error is usually assumed to be a white error process: $\varepsilon(\mathbf{s},t) \sim N(0,\sigma_\varepsilon^2)$. However, the manner in which the *spatiotemporally structured* error is handled is not straight-forward:

- $\omega_s(t)$ – temporal effects nested in sites (temporal evolution at each site, space is not modelled).
- $\omega_t(\mathbf{s})$ – spatial effects nested in time (spatial evolution at each time, time is not modelled).

- $\omega(\mathbf{s})\upsilon(t)$ or $\omega(\mathbf{s}) + \upsilon(t)$ – *separable* (structure in space and structure in time are independent).
- $\omega(\mathbf{s}, t)$ – non-separable (both time and space structure evolve in a nonsimple manner).

If spatial and temporal autocorrelation act independently ("separably") then:

$$\begin{aligned} \upsilon(t) &\sim \mathrm{GP}(0, C(t, t'; \boldsymbol{\theta}_t)) \\ \omega(\mathbf{s}) &\sim \mathrm{GP}(0, C(\mathbf{s}, \mathbf{s}'; \boldsymbol{\theta}_s)). \end{aligned}$$

The spatial and temporal errors are usually assumed to follow Gaussian Processes with mean 0 and covariance $C(\cdot,\cdot;\boldsymbol{\theta})$. The spatial covariance can be modelled with any spatial form such as: $C(\mathbf{s}, \mathbf{s}'; \boldsymbol{\theta}_s) = C(h)_{\text{Matérn}}$. Similarly, the temporal covariance can be formulated as any autocorrelation model such as: $C(t, t'; \boldsymbol{\theta}_t) = \sigma_t^2 \exp(-\phi_t |t - t'|)$, or if discrete in time: $\mathrm{AR}(1)$: $\upsilon(t+1) = \rho_t \upsilon(t) + \eta(t)$ with $\eta(t) \overset{iid}{\sim} N(0, \sigma_t^2)$, *etc*.

While computationally appealing, even in this simple case of "separable" models, an evaluation of the likelihood requires the inverse $\Sigma_{n\times n}^{-1}$, which happens to scale with $O(n^3)$ operations. This has been a strong bottleneck to further development of these covariance-based methods in large-scaled problems of space and space-time. Approximations have been suggested to overcome this crippling numerical problem: modeling the spatial process $\omega(s)$ with a lower dimensional process via kernel convolutions, moving averages, low rank splines/basis functions and predictive processes (projection of spatial process onto a smaller subset; Sølna and Switzer 1996, Wikle and Cressie 1999, Hung et al. 2004, Xu et al. 2005, Banerjee et al. 2014); approximating the spatial process as a Markov random field with Laplace and SPDE Approximations (Lindgren and Rue 2015); and approximating the likelihood of the spatial-temporal SPDE process with a spectral domain process (Sigrist et al. 2015).

In this framework, we focus upon Sigrist et al.'s (2015) approach due to the more realistic assumption of non-separability of space and time processes and the high computational performance of operating in spectral space. Specifically, the space-time error $\omega(\mathbf{s}, t)$ is formulated as an advection-diffusion process (SPDE):

$$\frac{\partial}{\partial t}\omega(\mathbf{s}, t) = -\mathbf{u}^T \nabla \omega(\mathbf{s}, t) + \nabla \cdot \Sigma \nabla \omega(\mathbf{s}, t) - \zeta \omega(\mathbf{s}, t) + \epsilon(\mathbf{s}, t).$$

Here, $\mathbf{s} = (x, y)^T \in \Re^2$: $\mathbf{u} = (u_x, u_y)^T$ parameterizes the drift velocity (advection); $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})^T$ is the gradient operator; $\nabla \cdot \mathbf{F} = (\frac{\partial F_x}{\partial x}, \frac{\partial F_y}{\partial y})^T$ is the divergence operator for a given vector field $\mathbf{F} = (F_x, F_y)^T$; $\Sigma^{-1} = \frac{1}{\phi_d^2}\begin{pmatrix} \cos\alpha & \sin\alpha \\ -\gamma \cdot \sin\alpha & \gamma \cdot \cos\alpha \end{pmatrix}^T \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\gamma \cdot \sin\alpha & \gamma \cdot \cos\alpha \end{pmatrix}$ parameterizes the anisotropy in diffusion via ($\gamma > 0$, $\alpha \in [0, \pi/2]$) with $\phi_d > 0$ parameterizing the diffusion range; $\zeta > 0$ parameterizing local damping; and $\epsilon(\mathbf{s}, t)$ parameterizing a Gaussian random field that accounts for source-sink processes with white noise in time and Matérn spatial covariance (aka, "innovation").

The full advection-diffusion model specification in state space form is (Sigrist et al. 2015):

$$\begin{aligned} Y(\mathbf{s}, t) &= x(\mathbf{s}, t)^T \boldsymbol{\beta} + \omega(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t) \\ \omega(\mathbf{s}, t) &= \boldsymbol{\Phi}\boldsymbol{\alpha}(\mathbf{s}, t) \quad \{\text{discretized advection} - \text{diffusion process}\} \\ \alpha(\mathbf{s}, t) &= \mathbf{G}\boldsymbol{\alpha}(\mathbf{s}, t-1) + \hat{\epsilon}\,(\mathbf{s}, t) \quad \{\text{transition model}\} \\ \varepsilon(\mathbf{s}, t) &\sim N(\mathbf{0}, \tau^2 \mathbf{1}) \quad \{\text{unstructured error}\} \\ \hat{\epsilon}\,(\mathbf{s}, t) &\sim N\left(\mathbf{0}, \widehat{\mathbf{Q}}\right) \quad \{\text{innovation}\} \end{aligned}$$

where, $\boldsymbol{\alpha}(\mathbf{s},t)$ are Fourier coefficients; $\boldsymbol{\Phi} = [\boldsymbol{\phi}(\mathbf{s}_1), \ldots, \boldsymbol{\phi}(\mathbf{s}_N)]^T$ is a matrix of spatial basis functions; $\mathbf{G}$ is the transition matrix; and $\widehat{\mathbf{Q}}$ is the innovation covariance matrix (residual errors). See Appendix 3 for more details.

Once the form of the space-time model is formulated, it can be used as a predictive/interpolating method. Such interpolation is required to describe and understand the linkages between the key ecosystem attributes of productivity, biodiversity, habitat, and species of interest, as well as, to estimate the spatial and temporal scales.

We use an R-package implementation of the above space and space-time (stmv, [https://github.com/jae0/stmv]). The stmv library additionally permits localised, hierarchical, model-based interpolations, for a given (local) subdomain, indexed by λ. Note that μ indicates a global regressor in a linear, nonlinear or binomial functional form f, and μ_λ represents a local regressor specific to the subdomain λ. The model specifications for each data series are as follows:

- Bathymetry (pure space model):

$$\begin{aligned} Y_\lambda(\mathbf{s}) &= \omega_\lambda(\mathbf{s}) + \varepsilon_\lambda(\mathbf{s}) \\ \omega_\lambda(\mathbf{s}) &\sim \mathrm{GP}(0, C(\mathbf{s},\mathbf{s}';\boldsymbol{\theta}_\lambda)) \\ \varepsilon_\lambda(\mathbf{s}) &\sim \mathrm{N}(\mathbf{0}, \tau_\lambda(\mathbf{s})^2) \end{aligned}$$

- Substrate grainsize (space with covariates):

$$\begin{aligned} Y_\lambda(\mathbf{s}) &= \mu(\mathbf{s}) + \omega_\lambda(\mathbf{s}) + \varepsilon_\lambda(\mathbf{s}) \\ \mu(\mathbf{s}) &= f(\,\mathrm{depth}, \mathrm{slope}, \mathrm{curvature}\,) \\ \omega_\lambda(\mathbf{s}) &\sim \mathrm{GP}(0, C(\mathbf{s},\mathbf{s}';\boldsymbol{\theta}_\lambda)) \\ \varepsilon_\lambda(\mathbf{s}) &\sim \mathrm{N}(\mathbf{0}, \tau_\lambda(\mathbf{s})^2) \end{aligned}$$

- Bottom temperature (temporal effects nested in sites):

$$\begin{aligned} Y_\lambda(\mathbf{s},t) &= \mu(\mathbf{s}) + \omega_\lambda(\mathbf{s}) + \varepsilon_\lambda(\mathbf{s}) \\ \mu(\mathbf{s}) &= f(\,\mathrm{depth}\,) \\ \omega_\lambda(\mathbf{s},t) &= \boldsymbol{\Phi}_\lambda \boldsymbol{\alpha}_\lambda(\mathbf{s},t) \\ \alpha_\lambda(\mathbf{s},t) &= \mathbf{G}_\lambda \boldsymbol{\alpha}_\lambda(\mathbf{s},t-1) + \hat{\epsilon}_\lambda\,(\mathbf{s},t) \\ \varepsilon_\lambda(\mathbf{s},t) &\sim N(\mathbf{0}, \tau_\lambda^2 \mathbf{1}) \\ \hat{\epsilon}_\lambda\,(\mathbf{s},t) &\sim N(\mathbf{0}, \widehat{\mathbf{Q}}_\lambda) \end{aligned}$$

- Ecosystem indicators (including integrative habitat H_i; temporal effects nested in sites):

$$\begin{aligned} Y_\lambda(\mathbf{s},t) &= \mu(\mathbf{s}) + \mu_\lambda(\mathbf{s},t) + \omega_\lambda(\mathbf{s}) + \varepsilon_\lambda(\mathbf{s}) \\ \mu(\mathbf{s}) &= f(\,\mathrm{depth}, \mathrm{slope}, \mathrm{curvature}, \mathrm{substrate\ grainsize}\,) \\ \mu_\lambda(\mathbf{s},t) &= f_\lambda(\,\mathrm{temperature}\,) \\ \omega_\lambda(\mathbf{s},t) &= \boldsymbol{\Phi}_\lambda \boldsymbol{\alpha}_\lambda(\mathbf{s},t) \\ \alpha_\lambda(\mathbf{s},t) &= \mathbf{G}_\lambda \boldsymbol{\alpha}_\lambda(\mathbf{s},t-1) + \hat{\epsilon}_\lambda\,(\mathbf{s},t) \\ \varepsilon_\lambda(\mathbf{s},t) &\sim N(\mathbf{0}, \tau_\lambda^2 \mathbf{1}) \\ \hat{\epsilon}_\lambda\,(\mathbf{s},t) &\sim N(\mathbf{0}, \widehat{\mathbf{Q}}_\lambda) \end{aligned}$$

- Abundance and functional habitat (H_f; temporal effects nested in sites):

$$
\begin{aligned}
Y_\lambda(\mathbf{s},t) &= \mu(\mathbf{s}) + \mu_\lambda(\mathbf{s},t) + \omega_\lambda(\mathbf{s}) + \varepsilon_\lambda(\mathbf{s}) \\
\mu(\mathbf{s}) &= f(\text{ depth, slope, curvature, substrate grainsize }) \\
\mu_\lambda(\mathbf{s},t) &= f_\lambda(\text{ temperature, ecosystem indicators }) \\
\omega_\lambda(\mathbf{s},t) &= \mathbf{\Phi}_\lambda \boldsymbol{\alpha}_\lambda(\mathbf{s},t) \\
\alpha_\lambda(\mathbf{s},t) &= \mathbf{G}_\lambda \boldsymbol{\alpha}_\lambda(\mathbf{s},t-1) + \hat{\epsilon}_\lambda(\mathbf{s},t) \\
\varepsilon_\lambda(\mathbf{s},t) &\sim N(\mathbf{0}, \tau_\lambda^2 \mathbf{1}) \\
\hat{\epsilon}_\lambda(\mathbf{s},t) &\sim N(\mathbf{0}, \widehat{\mathbf{Q}}_\lambda)
\end{aligned}
$$

Tagging and Mark-recapture

A tangible way of quantifying time and space scales (connectivity) is to demonstrate movement and genetic similarity. In the latter, no effort has been made. In the former, due to synergies with the fishing industry, increased tagging efforts have been made in the vicinity of SAB. Most of the effort has been driven by industry, Ocean Tracking Network (OTN), and Emera interest in Snow Crab movement near SAB. However, acoustic tagging of other species of interest have been completed as well. In total, 80 tags were deployed: 58 Cod, 14 Atlantic Striped Wolffish, 1 Shorthorn Sculpin and 7 Snow Crab. These tags will allow us to track these species over the coming years within the MPA (through the two existing receiver lines), as well as outside the MPA with other OTN receiver lines and potentially with an OTN glider. They will help define the spatial connectivity and range of the species of interest. The intent is to develop movement models where possible and estimate spatial range directly.

Mark-recapture information for sea turtles, seals, sharks and various other species exist in the area. These data have not been examined nor are they always available and represent a data gap at present.

RISK MODELING

Risk means many things to many people. We use the term specifically in the sense of having a believable error distribution for some quantity of interest such that probabilistic inferences can be made. Once the error distributions are determined, it is simple to make probability statements related to how likely the current state is different from some previous state or arbitrary threshold.

The estimation of such error distributions requires a reliable method to propagate the errors from observations to predictions. One venerable (**deterministic**) method is to build a mechanistic model and then using approximations or simulations to determine the error distributions of interest. A second (**phenomenological**) method empirically quantifies the errors and propagates them via statistical/correlational methods. The latter is chosen in this framework as it is very general and simple to implement. The former is not chosen as no operating model of an ecosystem nor any subcomponent is known to the authors that can be said to be able to perform with sufficient skill to be able to propagate errors, let alone, magnitudes.

The proposed approach is to use the discrete form of the logistic equation for this purpose. Verhulst, Pearl and Lotka in the late 1800s and early 1900s popularized this equation, using it to describe patterns of asymptotic increase (population growth, economic growth, etc.). The model is sufficiently general that it can be readily applied to most quantities of interest that show some dynamic behaviour, such as, for example: aggregate estimates of biomass (productivity), biodiversity (biodiversity change), and habitat (habitat change). Similar techniques can be applied to the other indicators of interest and may be applied depending upon availability of time. The justification of why it can be so generally applied is developed in Appendix 4.

The discrete form the basic logistic equation is, after normalization to K, the "carrying capacity" (see Appendix 4):

$$y_t \approx r y_{t-1}(1 - y_{t-1}).$$

Perhaps the most powerful and flexible method of estimation of the parameters, $\theta = \{r, K\}$, is a state space representation where an additional observation model is added to connect observed indices O to the unobserved and (usually unobservable) real system state y. The simplest such model is a linear scaling factor, q, though again others are possible, at the cost of more complexity:

$$O_t = q y_t$$

where, $\theta = \{r, K, q\}$.

The use of a Bayesian approach to solve the above nonlinear state space problem is frequently used as it provides an opportunity to: have greater numerical stability; incorporate prior scientific knowledge in a formal manner; realistically propagate credible errors; estimate unobserved ("true") states; and simultaneously estimate model "process" errors and data "observation" errors. [Note that process errors (σ_p^2) are the uncertainties that feeds back into future states via error propagation – e.g., via the recursive form of the logistic equation (i.e., errors in y_{t+1} in the state space of y_t vs y_{t+1}). They are important if predictive risk is being assessed. Observation errors (σ_o^2) refer to the uncertainties associated with measurement and observation (i.e., measurement/data-related errors of both variables in the state space of y_t vs y_{t+1}).]

This latter ability is particularly important as parameter estimates and forecasts based on observation-only errors provide unrealistically optimistic (small and constant) error bounds. Parameter estimates and forecasts based on process-only errors expand rapidly into the future, resulting in potentially unrealistically pessimistic (large and usually growing) error bounds.

The posterior distribution of the parameters of interest, θ, conditional upon the data are estimated via MCMC (Gibbs) sampling using the JAGS platform (Plummer 2003). The JAGS model used for parameter estimation can be found at: https://github.com/jae0/stmvdev/.

ANTHROPOGENIC THREATS

Productivity, habitat and biodiversity monitoring is essential if one is to assess if MPAs are achieving their primary objectives. It is also essential to examine the anthropogenic threats in the St. Anns Bank area and examine the cumulative impacts on productivity, habitat, biodiversity, and endangered or threatened species. From the data available, we can examine trawling and dredging disturbances, exploitation of marine resources by fisheries, fishing-gear entanglement threats to marine mammals and sea turtles, and vessel collision threats due to marine traffic, and vessel-noise disturbances. Each threat can be normalised on a zero-one scale to compare the intensity of threats across the region and then weighted and combined to examine cumulative anthropogenic threats as in Coll et al. (2012).

RESULTS

Most analyses have not been completed. The effort so far has been to assimilate and develop the scaffolding to support the analyses and future monitoring and assessment. These headings are here mostly as place holders. However, some preliminary results can be reported upon to show direction. We highlight a few of these results but emphasize their very preliminary nature.

BIODIVERSITY

No analytical results are available for presentation at this time.

PRODUCTIVITY

At present, only the models for Snow Crab have been completed using this method (Figure 34; see Choi et al. 2012 for more details).

HABITAT

Functional Habitat

At present, only the models for Snow Crab have been completed using this method (Figure 35 and 36; see Choi et al. 2012 for more details).

Integral Habitat

An example of **Integral habitat** (H_i) as expressed through an ordination of taxa found together in various bottom trawls (Figure 37).

CONNECTIVITY

Spatial Scale

This is a first attempt at describing the spatial scale of SAB and outlying areas. Essentially, S_s represents the distance one must walk before one loses memory of where one started (arbitrarily defined at the 5% level of autocorrelation). Using this level of similarity as the standard, depth variations are demonstrably patchy/clustered (Figure 38). In the SAB area, there is a mixture of large autocorrelation scales (exp(6)=400 km) in the north-east areas, while the south-west areas have spatial scales of approximately 20 km or less.

Recall that if S_s is small, then short-range processes dominate (e.g., less mobile species, weakly dispersing, low currents, habitat heterogeneity at small scales). If S_s is large, then broader/larger processes were influencing the productivity of the species (e.g., higher mobility or dispersal processes/current, and stronger spatial connectivity, habitat heterogeneity at larger scales).

As an another example, Snow Crab densities suggest different spatial autocorrelation patterns (Figure 39), with the majority of core fishing grounds being in the range of 60-90 km in length scale. In the vicinity of SAB, these spatial scales decline to about 10-60 km.

Temporal Scale

Similar to spatial scale, we have defined temporal scale (S_t) as the time required to reach a state where autocorrelation drops to an insignificant level (5%). Recall that if S_t is small, short-range variations dominate (higher sampling effort to resolve/describe). If S_t is large, long-range variations dominate (lower sampling effort required to resolve/describe). No analytical results are available for presentation at this time.

Tagging

No analytical results are available for presentation at this time beyond the overviews.

DISCUSSION

No discussions are forthcoming until more analysis can be completed.

CONCLUSIONS AND RECOMMENDATIONS

This report was an exercise in demonstrating what is available and possible. All methods discussed are viable and informative and so should serve as a strong basis for any future monitoring framework in SAB and MPAs in general.

ACKNOWLEDGEMENTS

We highlight the invaluable assistance of a great number of scientists that have been part of the various surveys that inform this document. This report could not have been completed without their guidance and assistance: Shelley Bond (DataShop), Carla Caverhill, George White (Remote Sensing Group), Catherine Johnson, Benoit Casault (Atlantic Zone Monitoring Progarm), Ben Zisserson, Brent Cameron, Snow Crab industry (Snow Crab survey), Scott Wilson, Bill MacEachern, Don Clark (groundfish survey), Dale Roddick (clam survey), Vladimir Kostelev, Charles Hannah (substrate grain size), Canadian Hydrographic Service (bathymetry data), Roger Pettipas (Ocean Sciences, ocean temperature data), Robert Grandy, Greg Croft and Krista Wry (Commercial Data Division), and Tanya Koropatnick and Norman Cochrane (vessel data).

REFERENCES

Banerjee, S., B.P. Carlin, and A.E. Gelfand. 2014. Hierarchical Modeling and Analysis for Spatial Data. CRC Press.

Carr, M-E., M.A.M. Friedrichs, M. Schmeltz, M.N. Aita, D. Antoine, K.R Arrigo, I. Asanuma, O. Aumont, R. Barber, M. Behrenfeld, R. Bidigare, E.T. Buitenhuis, J. Campbell, A. Ciotti, H. Dierssen, M. Dowell, J. Dunne, W. Esaias, B. Gentili, W. Gregg, S. Groom, N. Hoepffner, J. Ishizaka, T. Kameda, C. Le Quéré, S. Lohrenz, J. Marra, F. Mélin, K. Moore, A. Morel, T.E. Reddy, J. Ryan, M. Scardi, T. Smyth, K. Turpie, G. Tilstone, K. Waters, and Y. Yamanaka. 2006. A Comparison of Global Estimates of Marine Primary Production from Ocean Color. Deep Sea Research Part II: Topical Studies in Oceanography 53(5). Elsevier: 741–770.

Caverhill, C., H. Maas, C. Porter, G White III, and C. Fuentes-Yaco. 2015. MODIS/Aqua Ocean Colour Dataset, Remote Sensing Unit. Bedford Institute of Oceanography RSU-BIO Technical Document 5.

Choi, J.S. 2010. Habitat Preferences of the Snow Crab, Chionoecetes Opilio: Where Stock Assessment and Ecology Intersect; pp. 361–376. In: G.H. Kruse, G.L. Eckert, R.J. Foy, R.N. Lipcius, B. Sainte-Marie, D.L. Stram, and D. Woodby (eds.) Biology and Management of Exploited Crab Populations Under Climate Change. Alaska Sea Grant, University of Alaska Fairbanks.

Choi, J.S., B. Zisserson, and B. Cameron. 2012. Assessment of Scotian Shelf Snow Crab in 2011. DFO Canadian Science Advisory Secretariat Research Document 2012/024.

Coll, M., C. Piroddi, C. Albouy, F.B.R. Lasram, W.W.L. Cheung, V. Christensen, V.S. Karpouzi, F. Guilhaumon, D. Mouillot, M. Paleczny, M.L. Palomares, J. Steenbeek, P. Trujillo, R. Watson, and D. Pauly. 2012. The Mediterranean Sea Under Siege: Spatial Overlap Between Marine Biodiversity, Cumulative Threats and Marine Reserves. Global Ecology and Biogeography 21 (4). Wiley Online Library: 465–480.

Constantine, A.G., and P. Hall. 1994. Characterizing Surface Smoothness via Estimation of Effective Fractal Dimension. Journal of the Royal Statistical Society, Series B (Statistical Methodology) 56: 97–113.

Cook, A.M., and A. Bundy. 2010. The Food Habits Database: An Update, Determination of Sampling Adequacy and Estimation of Diet for Key Species. Canadian Technical Report of Fisheries and Aquatic Sciences 2884: 1–140.

Cook, A. M., and A. Bundy. 2012. Use of Fishes as Sampling Tools for Understanding Biodiversity and Ecosystem Functioning in the Ocean. Marine Ecology Progress Series 454. Inter-Research, Nordbuente 23 Oldendorf/Luhe 21385 Germany: 1–18.

Darecki, M., and D. Stramski. 2004. An Evaluation of Modis and Seawifs Bio-Optical Algorithms in the Baltic Sea. Remote Sensing of Environment 89(3). Elsevier: 326–50.

Devine, L., M.K. Kennedy, I. St-Pierre, C. Lafleur, M. Ouellet, and S. Bond. 2014. BioChem: The Fisheries and Oceans Canada Database for Biological and Chemical Data. Canadian Technical Report of Fisheries and Aquatic Sciences 3073: iv + 40 pp.

DFO. 2012. Conservation Priorities, Objectives, and Ecosystem Assessment Approach for the St. Anns Bank Area of Interest (AOI). DFO Canadian Science Advise Secretariat Science Advisory Report 2012/034.

Diggle, P.J., and P.J. (Jr.) Ribeiro. 2007. Model Based Geostatistics. New York: Springer.

Feldman, G.C., and C.R. McClain. 2015. Ocean Color Web. Edited by N. Kuring and S.W. Bailey. MODIS Aqua Reprocessing 2013.1. NASA Goddard Space Flight Center. https://oceancolor.gsfc.nasa.gov/.

Finley, A.O., S. Banerjee, and B.P. Carlin. 2007. SpBayes: An R Package for Univariate and Multivariate Hierarchical Point-Referenced Spatial Models. Journal of Statistical Software 19(4): 1–24.

Ford, J., and A. Serdynska. 2013. Ecological Overview of St. Anns Bank. Canadian Technical Report of Fisheries and Aquatic Sciences 3023.

Gerritsen, H.D., C. Minto, and C. Lordan. 2013. How Much of the Seabed Is Impacted by Mobile Fishing Gear? Absolute Estimates from Vessel Monitoring System (Vms) Point Data. ICES Journal of Marine Science. Oxford University Press, fst017, 1–9.

Green, R.H. 1979. Sampling Design and Statistical Methods for Environmental Biologists. Wiley Interscience, Chichester, England.

Guzman, H.M, C.G. Gomez, C.A. Guevara, and L. Kleivane. 2013. Potential Vessel Collisions with Southern Hemisphere Humpback Whales Wintering Off Pacific Panama. Marine Mammal Science 29(4): 629–642.

Hart, P.E., N.J. Nilsson, and B. Raphael. 1968. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. Systems Science and Cybernetics, IEEE Transactions 4(2). IEEE: 100–107.

Hatch, L.T., C.W. Clark, S.M. Van Parijs, A.S. Frankel, and D.W. Ponirakis. 2012. Quantifying Loss of Acoustic Communication Space for Right Whales in and Around a US National Marine Sanctuary. Conservation Biology 26(6): 983–94.

Hatch, L.T., C. Clark, R. Merrick, S. Van Parijs, D. Ponirakis, K. Schwehr, M. Thompson, and D. Wiley. 2008. Characterizing the Relative Contributions of Large Vessels to Total Ocean Noise Fields: A Case Study Using the Gerry E. Studds Stellwagen Bank National Marine Sanctuary. Environmental Management 42(5): 735–52.

Hung, C.-W., H.-H. Hsu, and M.-M. Lu. 2004. Decadal Oscillation of Spring Rain in Northern Taiwan. Geophysical Research Letters 31(22). Wiley Online Library.

Hurlbert, S.H. 1984. Pseudoreplication and the Design of Ecological Field Experiments. Ecological Monographs 54(2). Wiley Online Library: 187–211.

Kenchington, T.J. 2013. A Monitoring Framework for the St. Anns Bank Area of Interest. Canadian Science Advisory Secretariat Research Document 2013/117.

Lee, J., A.B South, and S. Jennings. 2010. Developing Reliable, Repeatable, and Accessible Methods to Provide High-Resolution Estimates of Fishing-Effort Distributions from Vessel Monitoring System (VMS) Data. ICES Journal of Marine Science: Journal du Conseil. Oxford University Press, fsq010.

Lindgren, F., and H. Rue. 2015. Bayesian Spatial Modelling with R-INLA. Journal of Statistical Software 63(1): 1–25.

Lotka, A.J. 1925. Elements of Physical Biology. Williams and Wilkins.

McKenna, M.F., D. Ross, S.M. Wiggins, and J.A. Hildebrand. 2012. Underwater Radiated Noise from Modern Commercial Ships. The Journal of the Acoustical Society of America 131(1): 92–103.

Merchant, N.D., E. Pirotta, T.R. Barton, and P.M. Thompson. 2014. Monitoring Ship Noise to Assess the Impact of Coastal Developments on Marine Mammals. Marine Pollution Bulletin 78(1): 85–95.

Moore, S.K., V.L. Trainer, N.J. Mantua, M.S. Parker, E.A. Laws, L.C. Backer, and L.E. Fleming. 2008. Impacts of Climate Variability and Future Climate Change on Harmful Algal Blooms and Human Health. Environmental Health 7(2). BioMed Central: S4.

Platt, T., and S. Sathyendranath. 1988. Oceanic Primary Production: Estimation by Remote Sensing at Local and Regional Scales. Science 241: 1613–20.

Platt, T., S. Sathyendranath, M.-H. Forget, G.N. White III, C. Caverhill, H. Bouman, E. Devred, and S. Son. 2008. Operational Estimation of Primary Production at Large Geographical Scales. Remote Sensing of Environment 112: 3437–3448.

Plummer, M. 2003. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling.

R Core Team. 2015. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Redfern, J.V., M.F. McKenna, T.J. Moore, J. Calambokidis, M.L. Deangelis, E.A. Becker, J. Barlow, K.A. Forney, P.C. Fiedler, and S.J. Chivers. 2013. Assessing the Risk of Ships Striking Large Whales in Marine Spatial Planning. Conservation Biology 27(2): 292–302.

Ribeiro, P.J. Jr., and P.J. Diggle. 2001. geoR: A Package for Geostatistical Analysis. R-NEWS: June 1(2): 15–18.

Sigrist, F., H.R. Künsch, and W.A. Stahel. 2015. Stochastic Partial Differential Equation Based Modelling of Large Space–Time Data Sets. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77(1). Wiley Online Library: 3–33.

Simard, Y., N. Roy, S. Giard, and M.Yayla. 2014. Canadian Year-Round Shipping Traffic Atlas for 2013: Volume 1, East Coast Marine Waters. Canadian Technical Report of Fisheries and Aquatic Sciences 3091: 1–327.

Sølna, K., and P. Switzer. 1996. Time Trend Estimation for a Geographic Region. Journal of the American Statistical Association 91 (434). Taylor & Francis: 577–89.

Tremblay, M.J., G.A.P. Black, and R.M. Branton. 2007. The Distribution of Common Decapod Crustaceans and Other Invertebrates Recorded in Annual Ecosystem Surveys of the Scotian Shelf 1999-2006. Canadian Technical Report of Fisheries and Aquatic Sciences 2762: 1–74.

Underwood, A.J. 1992. Beyond Baci: The Detection of Environmental Impacts on Populations in the Real, but Variable World. Journal of Experimental Marine Biology and Ecology 161: 145–178.

Vanderlaan, A.S.M., and C.T. Taggart. 2009. Efficacy of a Voluntary Area to Be Avoided to Reduce Risk of Lethal Vessel Strikes to Endangered Whales. Conservation Biology 23(6): 1467–1474.

Wikle, C.K., and N. Cressie. 1999. A Dimension-Reduced Approach to Space-Time Kalman Filtering. Biometrika. JSTOR, 815–29.

Wiley, D.N., M. Thompson, R.M. Pace, and J. Levenson. 2011. Modeling Speed Restrictions to Mitigate Lethal Collisions Between Ships and Whales in the Stellwagen Bank National Marine Sanctuary, USA. Biological Conservation 144(9): 2377–81.

Xu, K., C.K. Wikle, and N.I. Fox. 2005. A Kernel-Based Spatio-Temporal Dynamical Model for Nowcasting Weather Radar Reflectivities. Journal of the American Statistical Association 100(472). Taylor & Francis: 1133–1144.

FIGURES

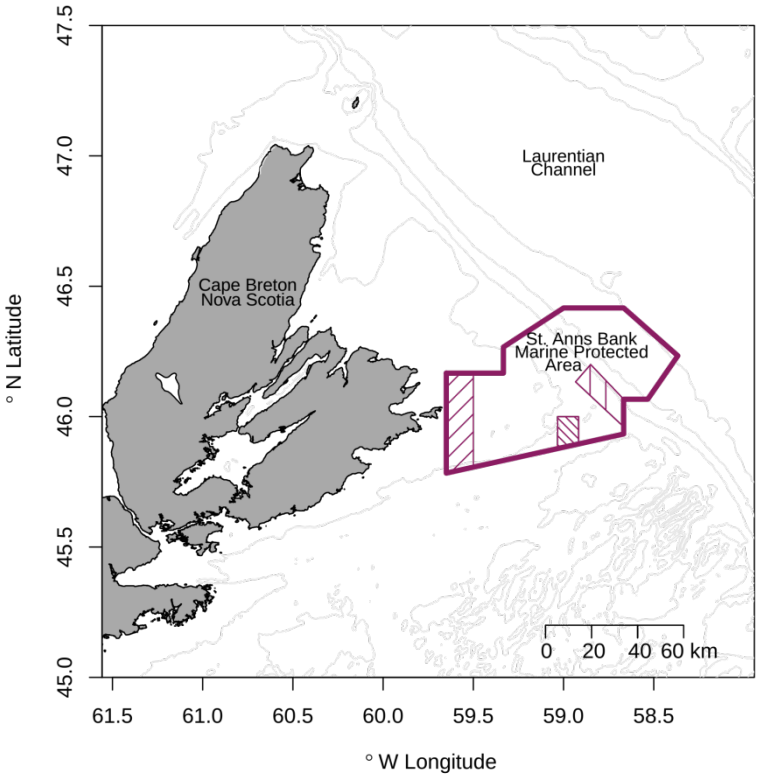

Figure 1. Bathymetric chart of the St. Anns Bank (SAB) area with the proposed Marine Protected Area (MPA, thick line) and limited fishing zones (hatched areas). See Figure 2 for geographic location in a larger map.

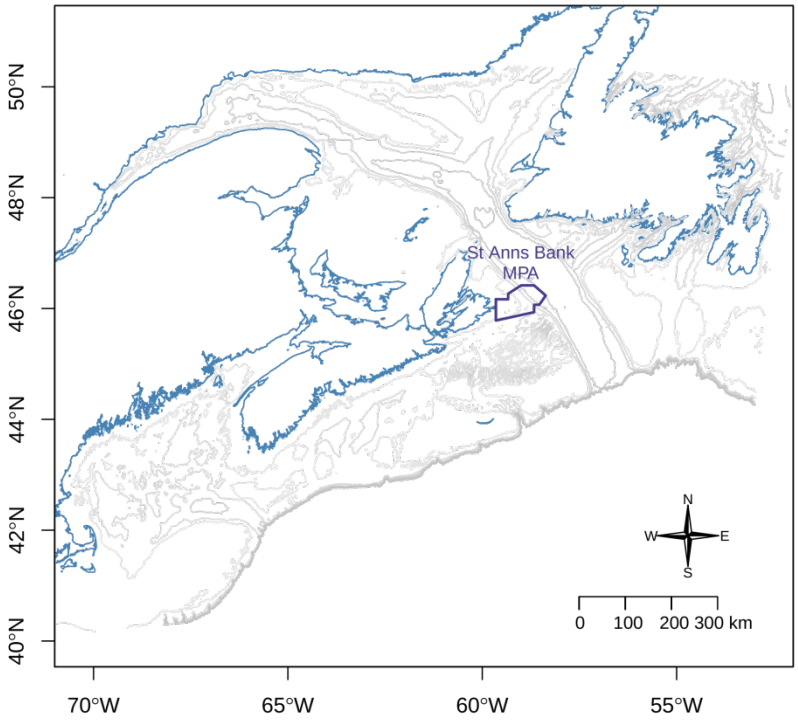

Figure 2. Map of the data extraction area 37°N to 48°N and from 48°W to 71°W and the relative location of the St. Anns Bank MPA.

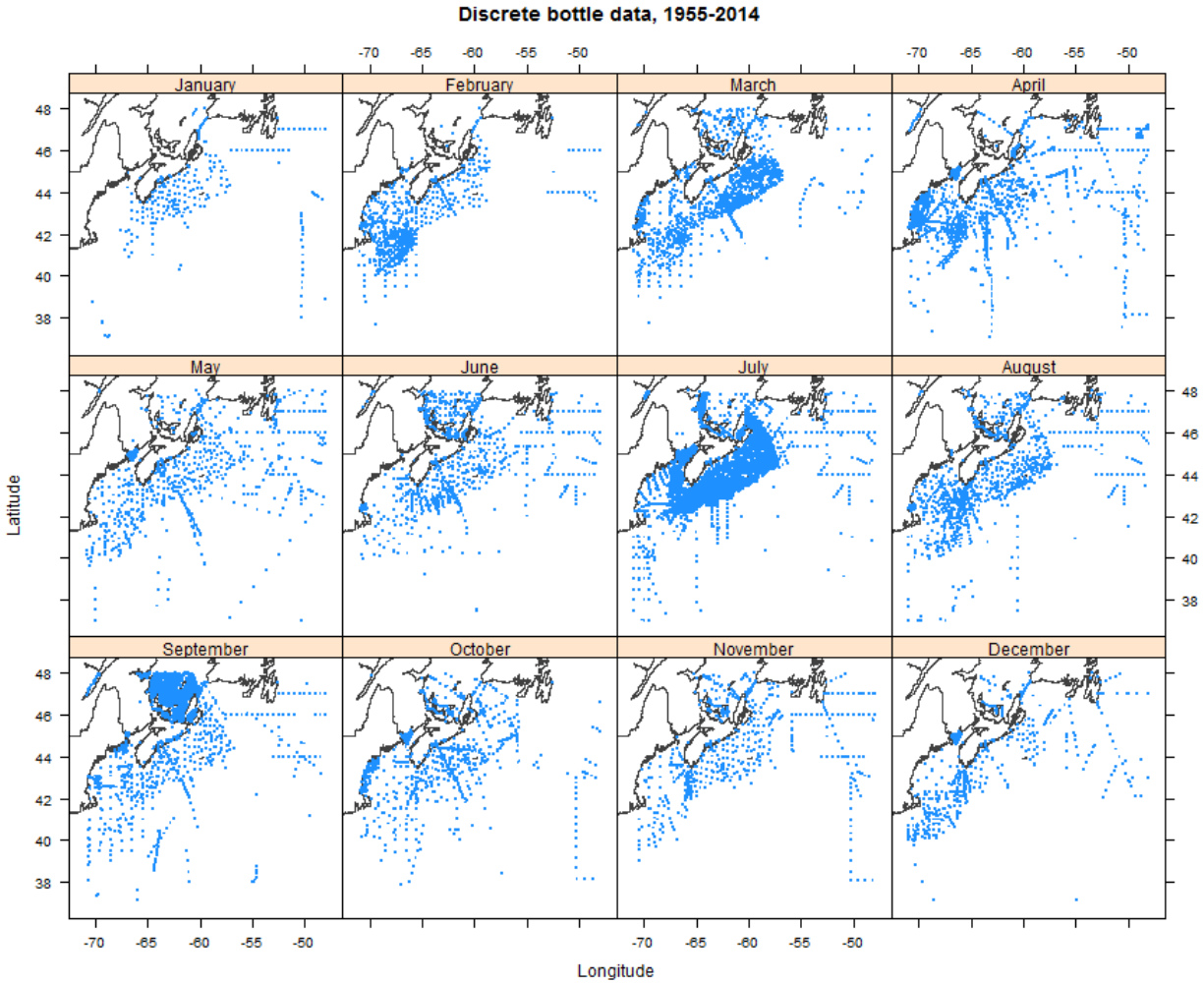

Figure 3. Monthly spatial distribution of discrete bottle data for the time period 1955-2014.

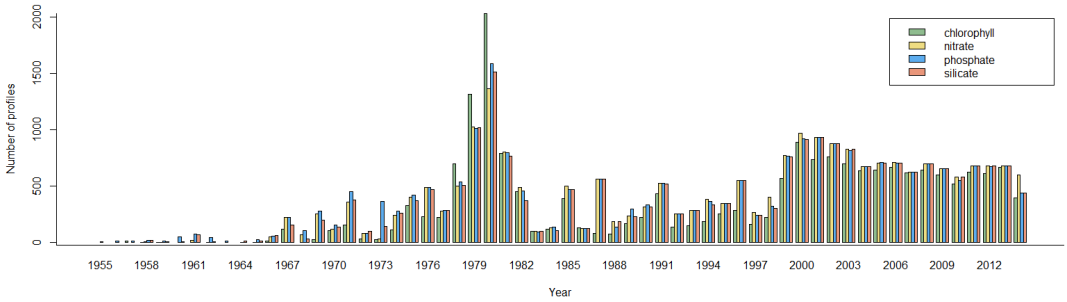

Figure 4. Number of chlorophyll and nutrient profiles extracted from the BioChem database for each year since 1955.

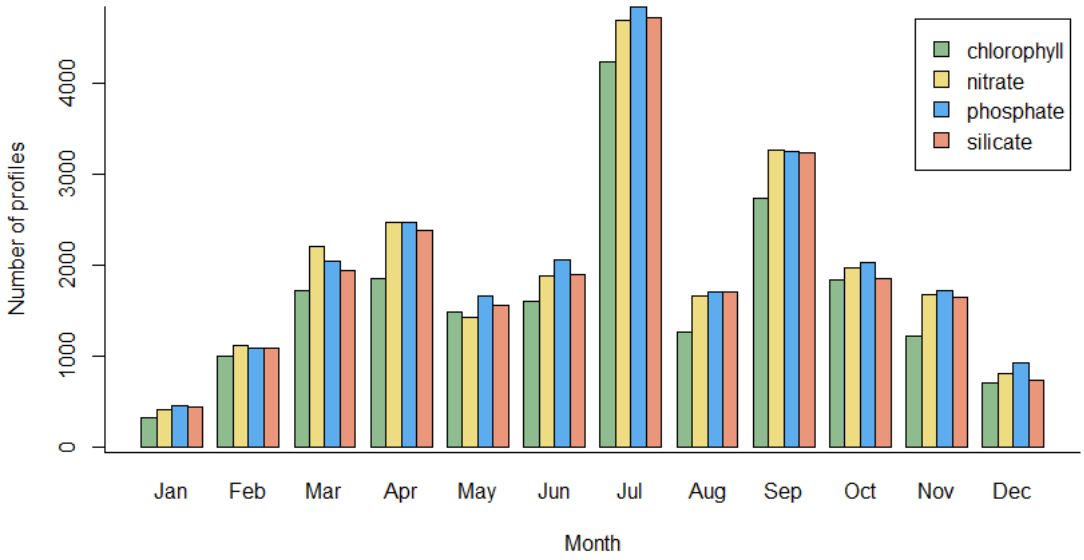

Figure 5. Number of chlorophyll and nutrient profiles extracted from the BioChem database for the time period 1955-2014, grouped monthly.

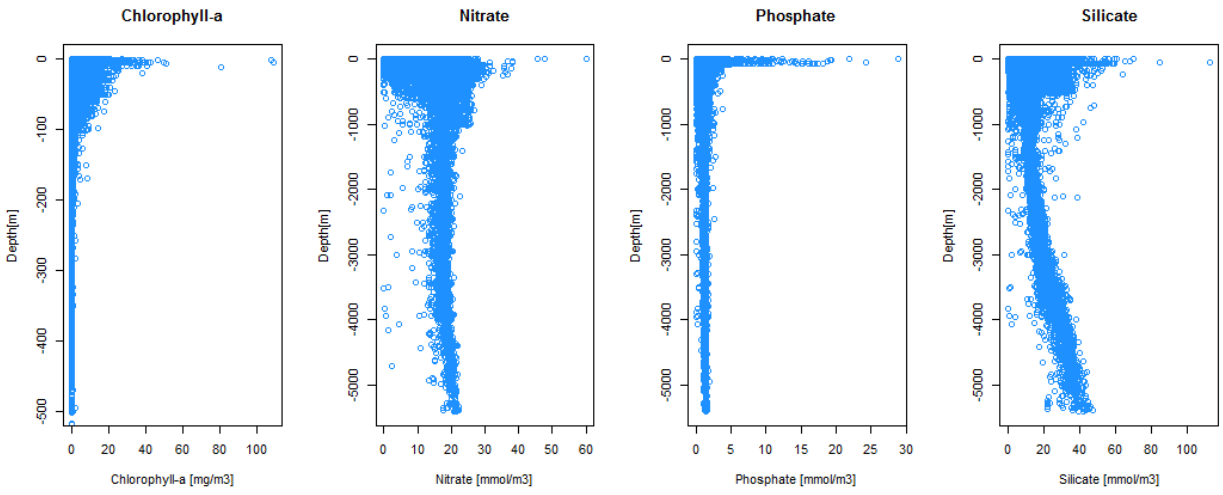

Figure 6. Depth profiles of chlorophyll-a and nutrients; all data for the time period 1955-2014.

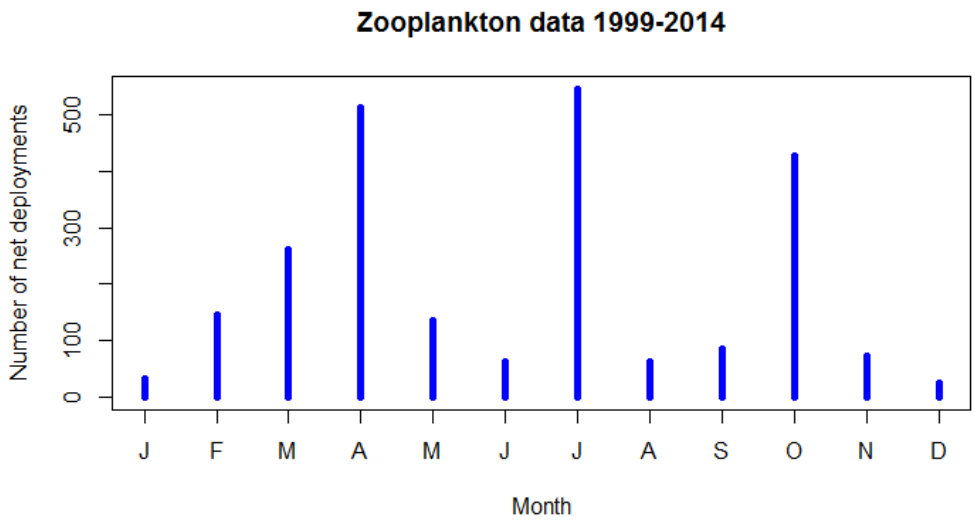

Figure 7. Total number of net deployments for each month during the time period 1999-2014.

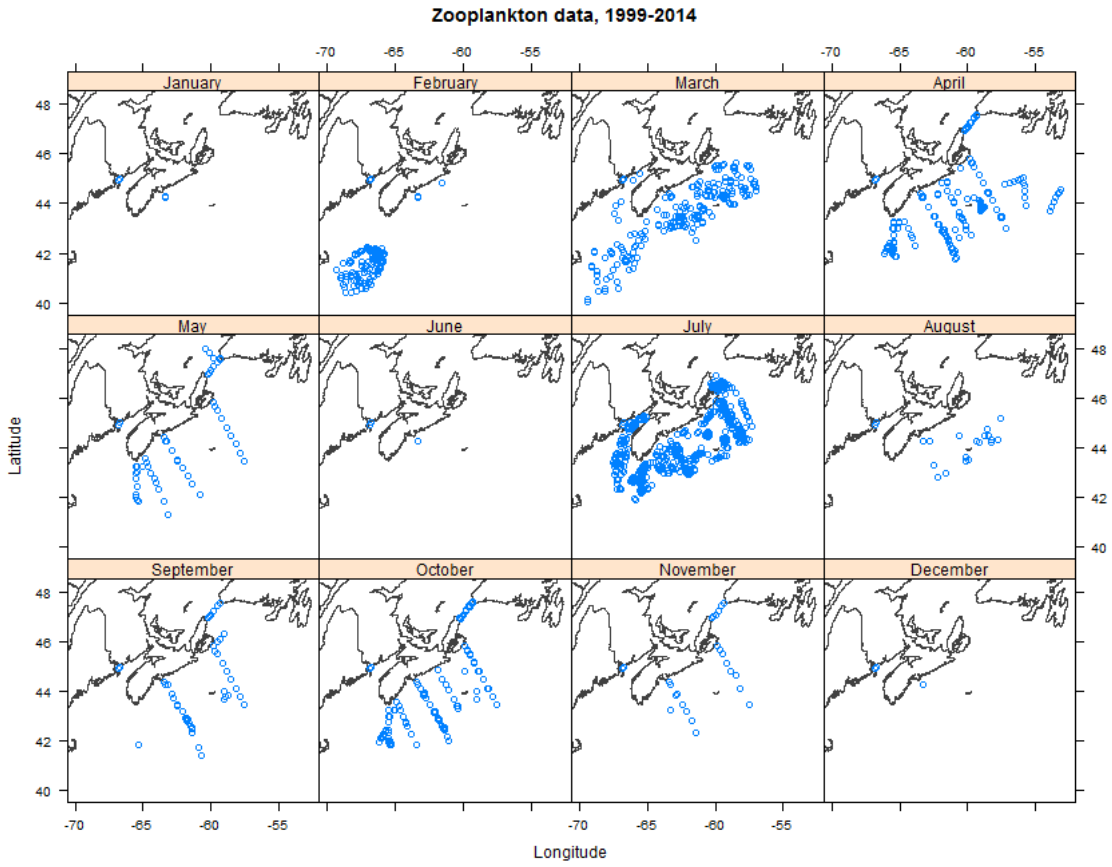

Figure 8. Total number of net deployments for each month during the time period 1999-2014.

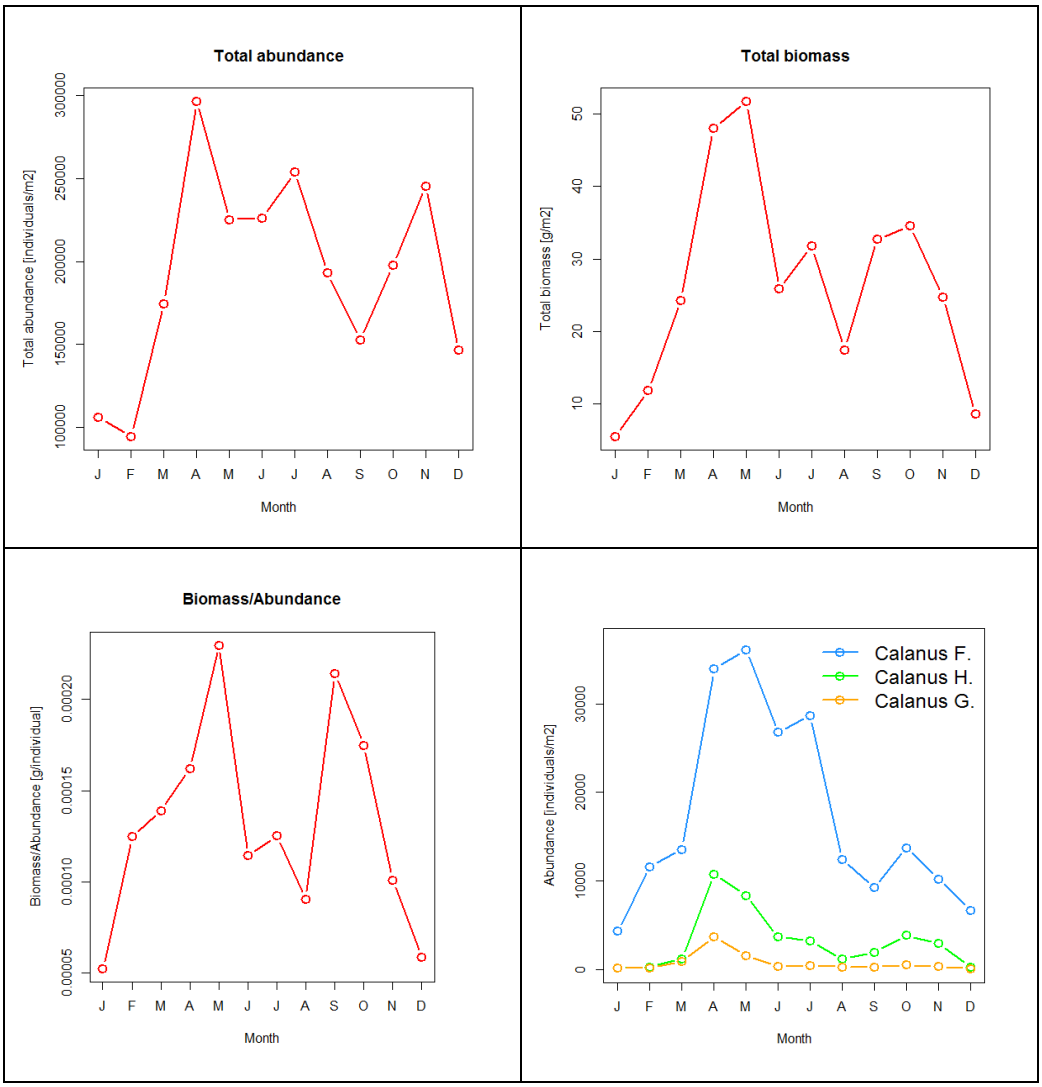

Figure 9. Monthly averages of all data from 1999 to 2014: total abundance (top left), total biomass computed from wet weight (top right), ratio of total biomass computed from wet weight to total abundance (bottom left) as a potential measure of the average weight of the individual organism, and abundance of Calanus finmarchicus, C. hyperboreus, and C. glacialis (bottom right).

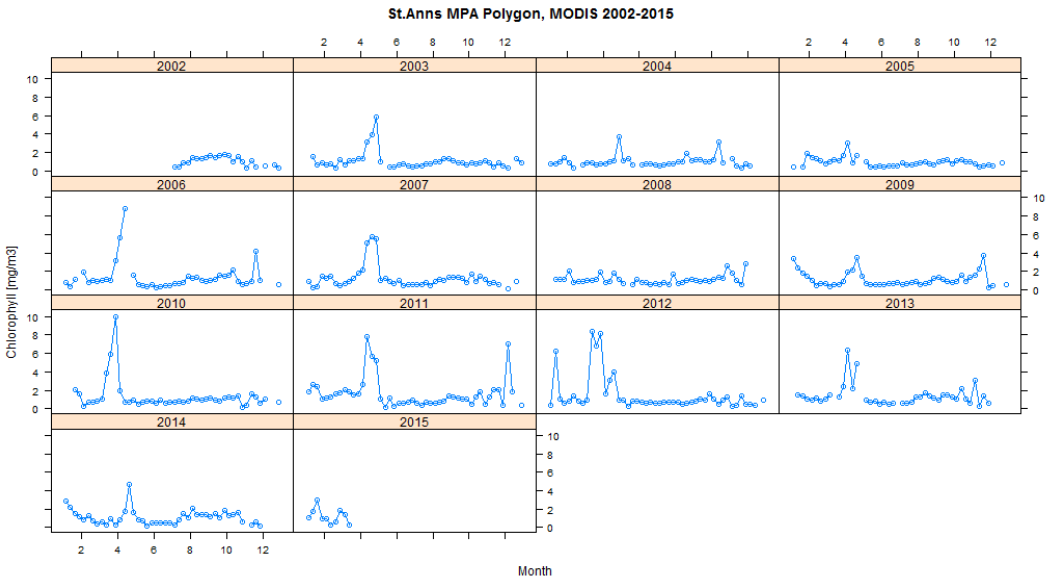

Figure 10. Chlorophyll-a concentration extracted from MODIS 8-day composite images for St. Anns Bank polygon for the time period 2002-2015.

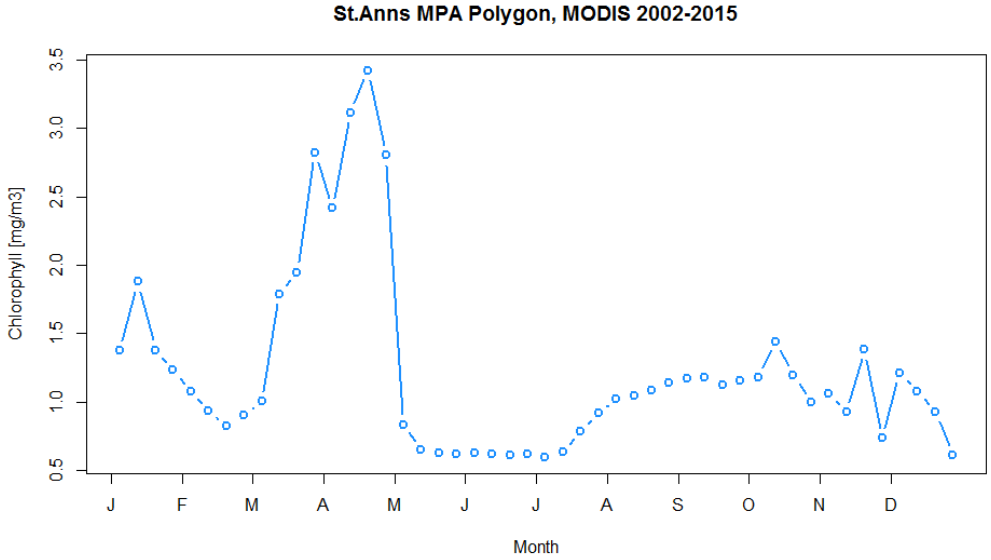

Figure 11. Average Chorophyll-a concentration computed from 8-day composite images for St. Anns Bank polygon for the time period 2002-2015.

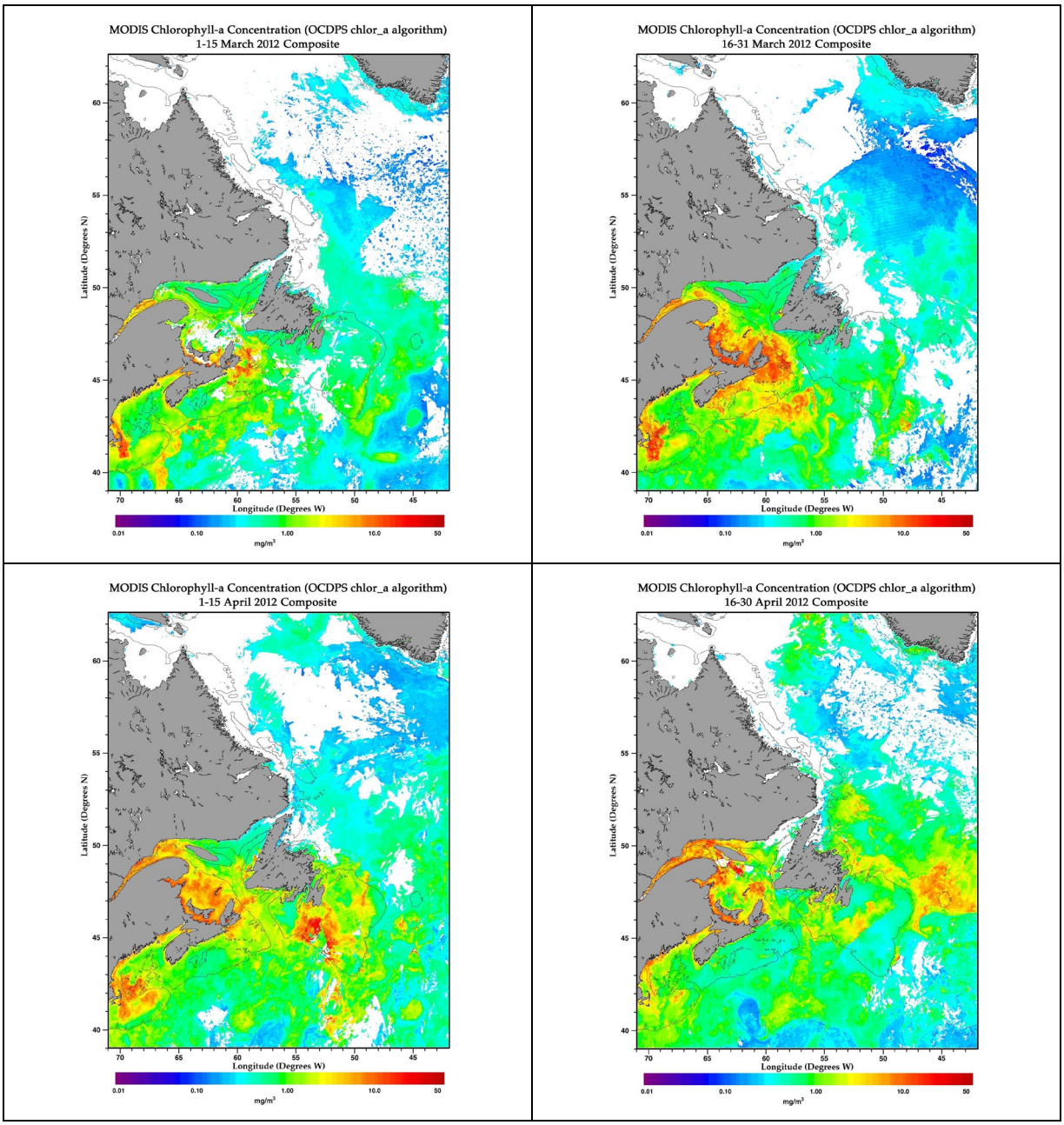

Figure 12. MODIS semi-monthly chlorophyll-a concentration showing spring bloom progression in the Northwest Atlantic in 2012. Note the intense bloom at St. Anns Bank during the last two weeks in March.

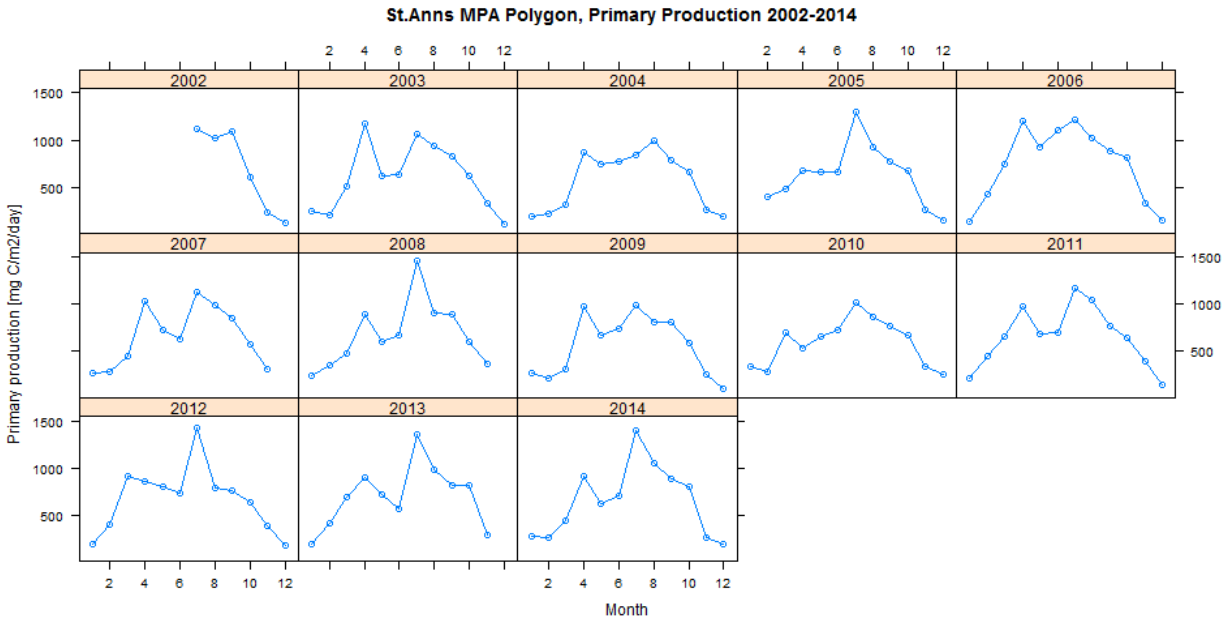

Figure 13. Annual monthly Primary Production (PP) computed from PP composite images for St. Anns Bank polygon for the time period 2002-2014.

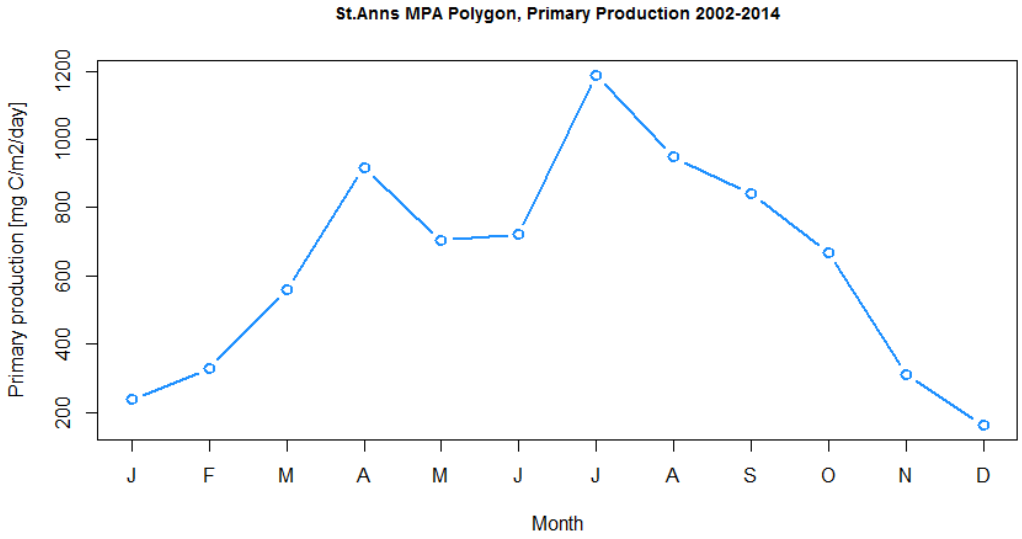

Figure 14. Average Primary Production (PP) computed from monthly composite images for St. Anns Bank polygon for the time period 2002-2014.

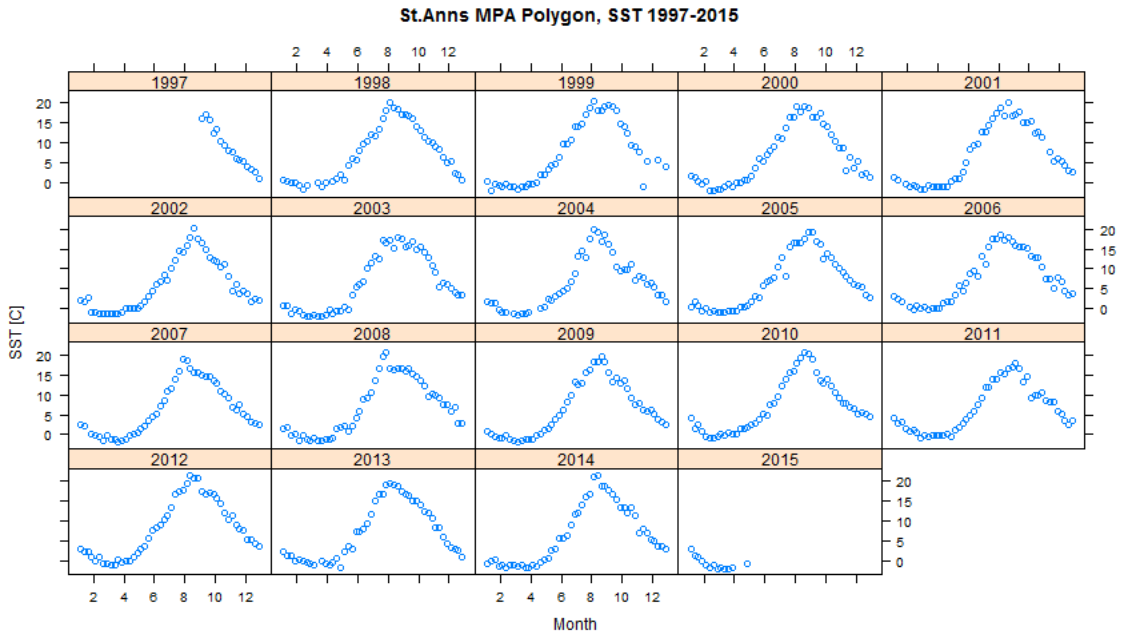

Figure 15. Sea Surface Temperature (SST) extracted from 8-day AVHRR composite images for St. Anns Bank polygon for the time period 1997-2015.

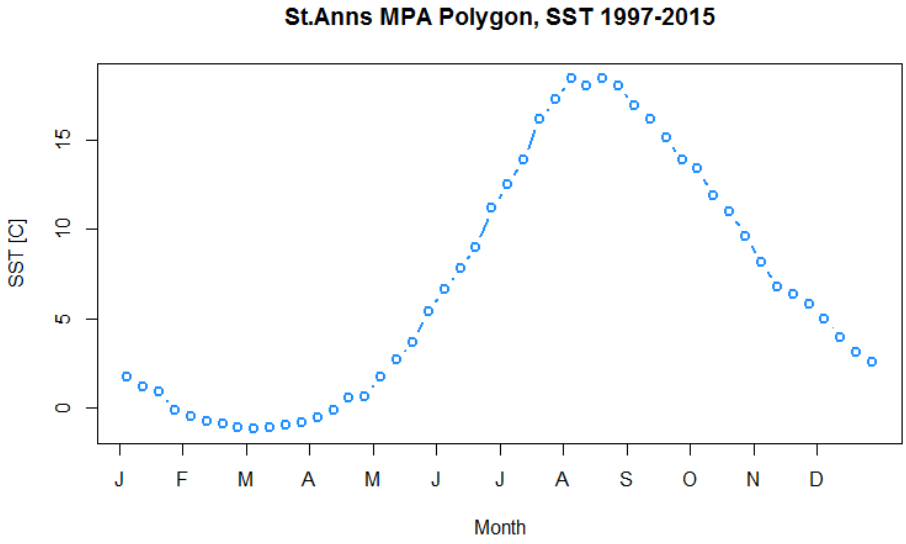

Figure 16. Average Sea Surface Temperature (SST) computed from 8-day AVHRR composite images for St. Anns Bank polygon for the time period 1997-2015.

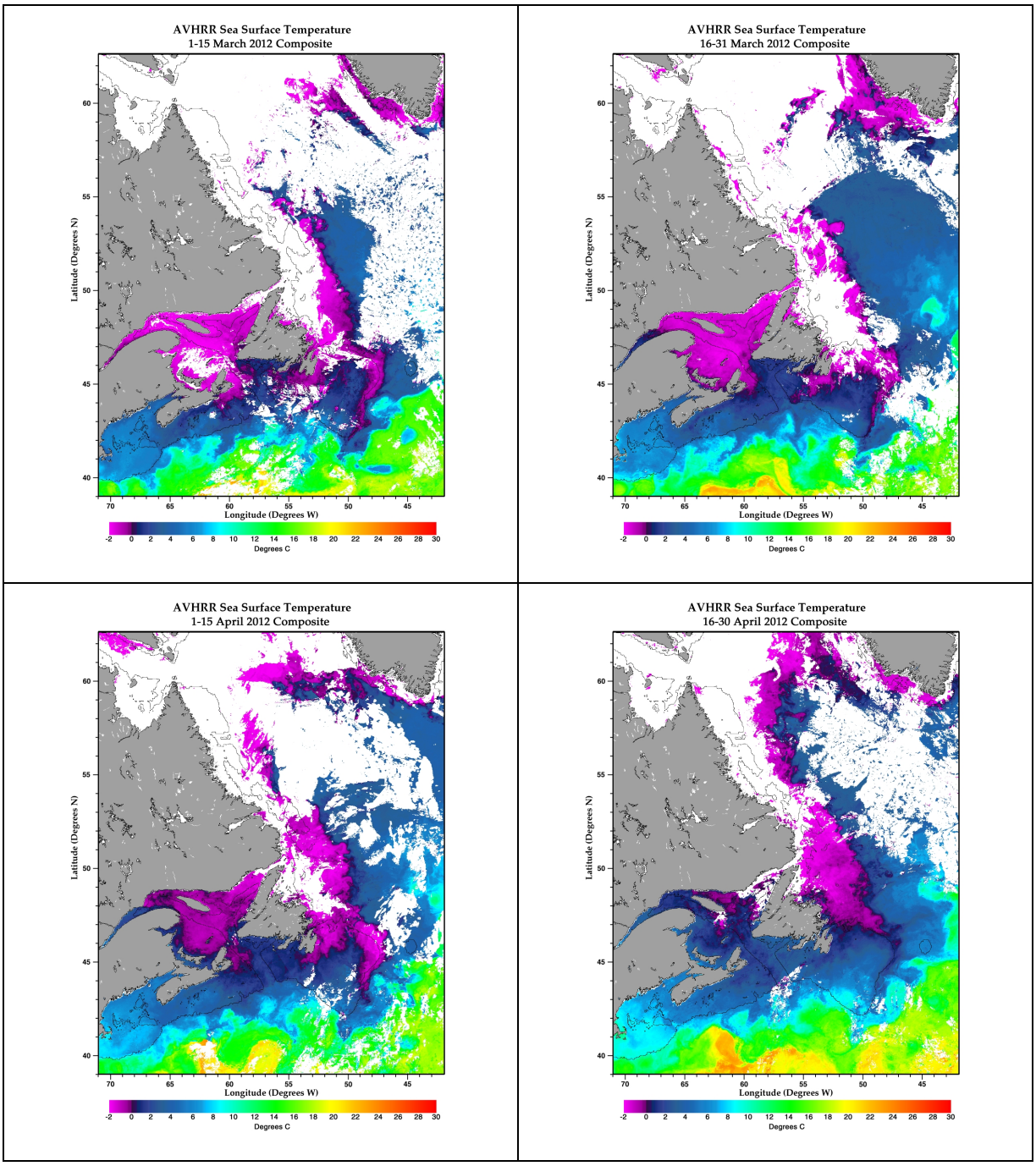

Figure 17. Bi-weekly composites from AVHRR showing SST in the Northwest Atlantic in the spring of 2012, corresponding to the intense spring bloom at St. Anns bank shown in Figure 2.

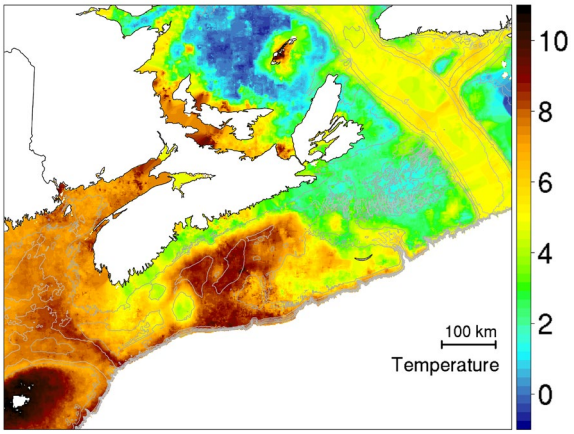

Figure 18. Average bottom temperatures computed from all available data 1950-2016.

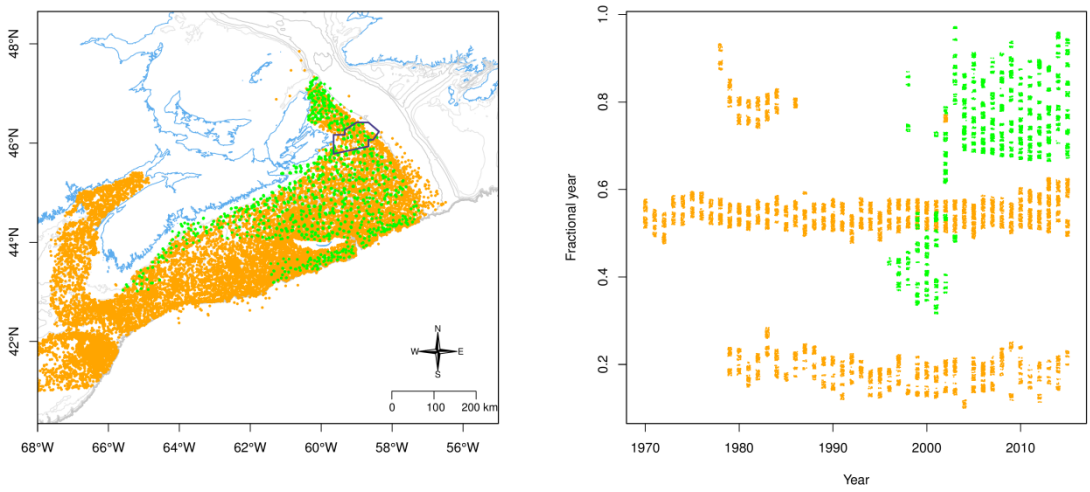

Figure 19. Left: Survey locations in the Groundfish survey (orange) and Snow Crab survey (green). Right: Timing of surveys in the Groundfish survey (orange) and Snow Crab survey (green).

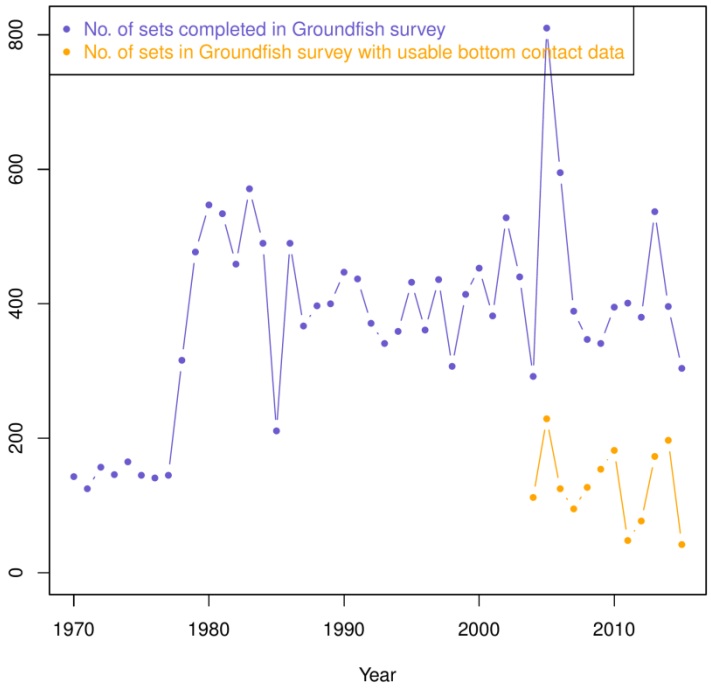

Figure 20. Number of sets in the groundfish surveys (blue) and the number of sets with usable net configuration data (orange).

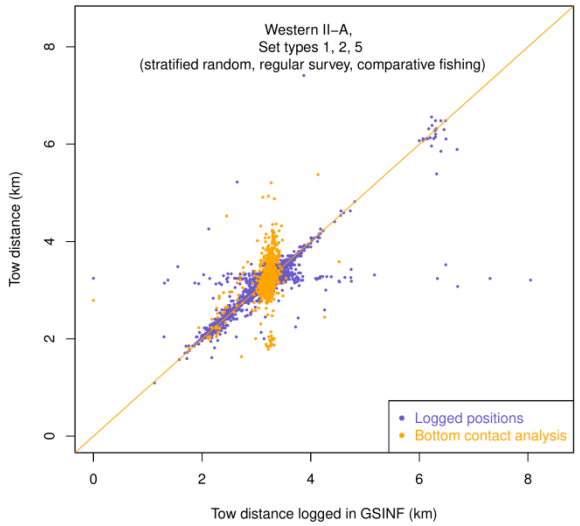

Figure 21. Towed distance comparisons in the groundfish survey.

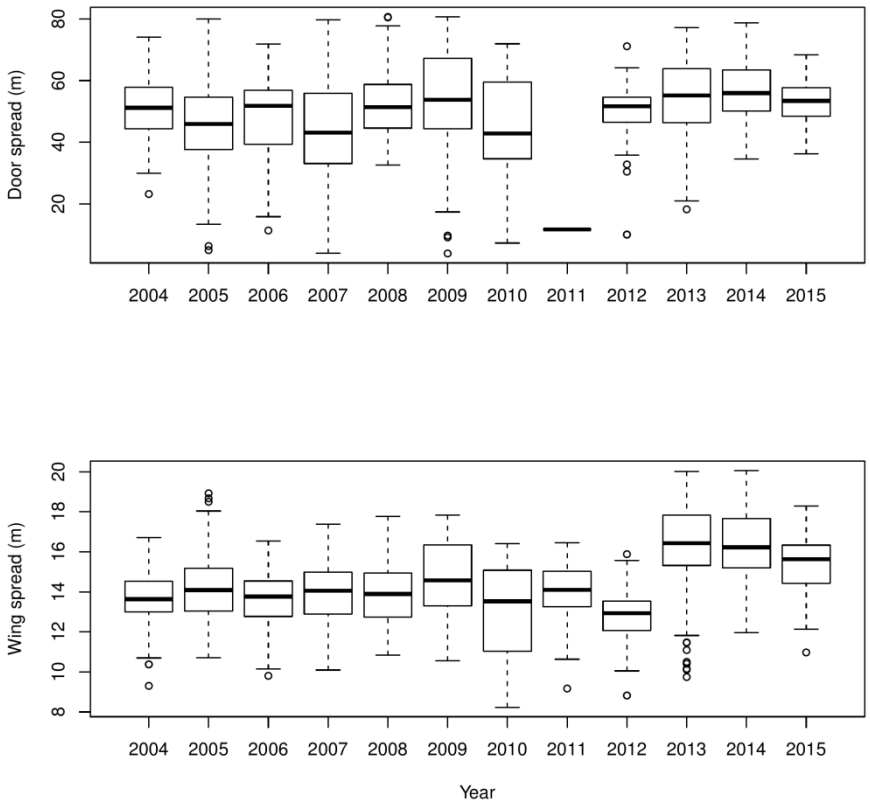

Figure 22. Net spread variations by year. Note in 2011, the doorspread sensors seem to have failed completely. Note also that wingspread has been significantly larger from 2013 to 2015.

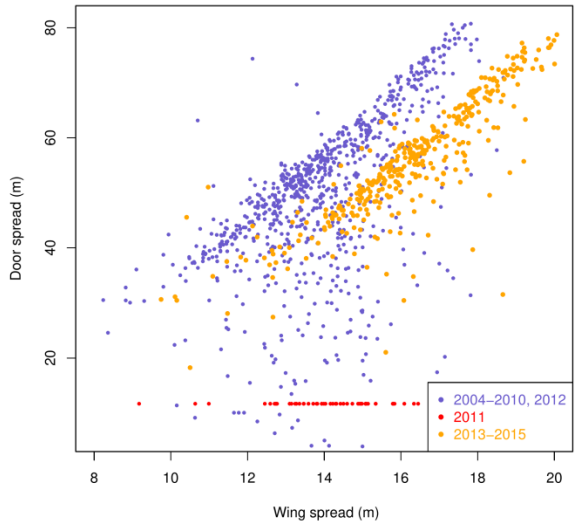

Figure 23. Net spread variations: doorspread vs wingspread. Note also that wingspread has been significantly larger from 2013 to 2015 but not doorspread. The cause of this divergence is not known.

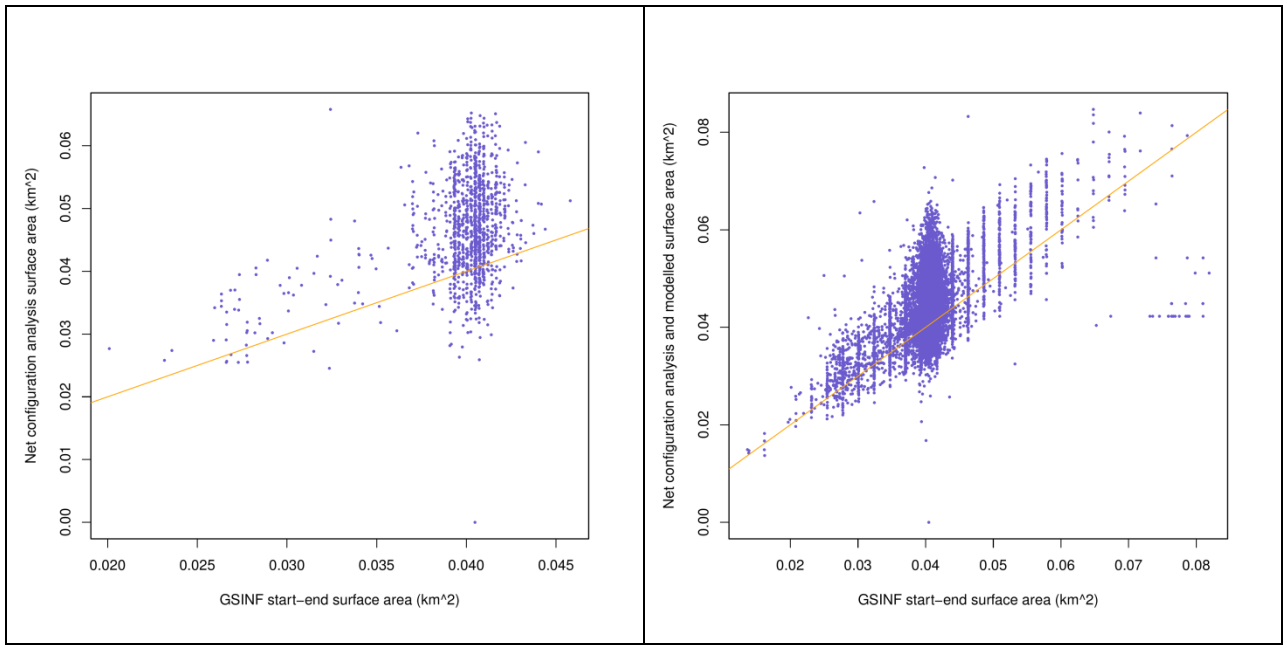

Figure 24. Left: Surface area estimates based on GSINF logged start-end positions vs computed surface area estimated from tow track and net configuration. Right: Surface area estimates based on GSINF logged start-end positions vs computed surface area estimated from tow track and net configuration, ***as well as, modeled solutions****.*

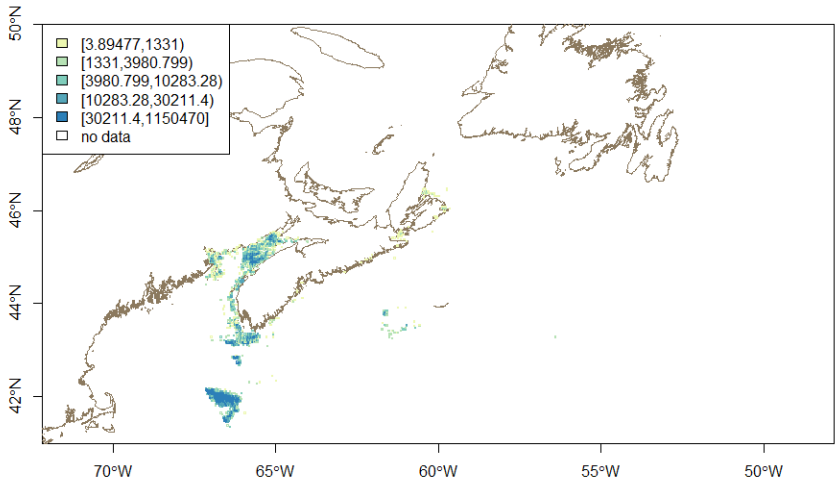

Figure 25. Commercial catch weights of Sea Scallops (Placopecten magellanicus) on Georges Bank, the Scotian Shelf, and in the Bay of Fundy.

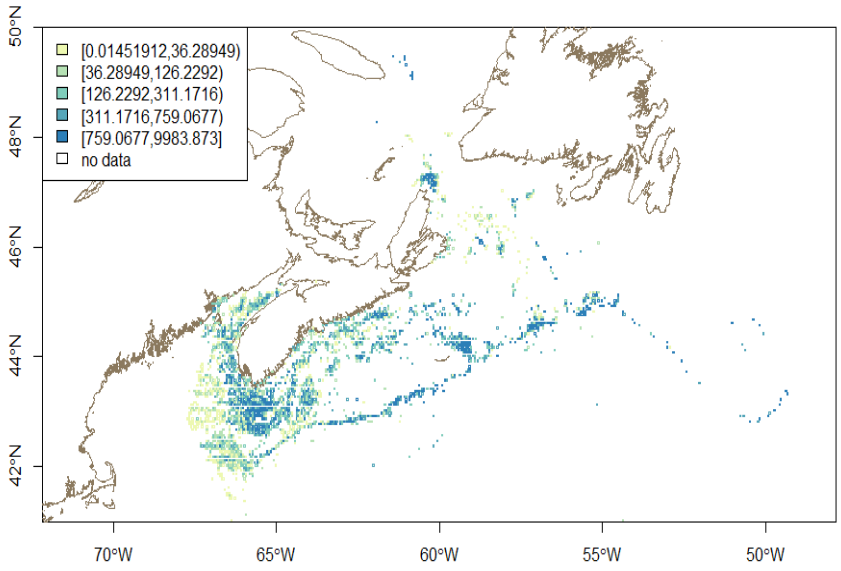

Figure 26. Commercial catch weights of Atlantic Halibut (Hippoglossus hippoglossus) on Georges Bank, the Scotian Shelf, and in the Bay of Fundy.

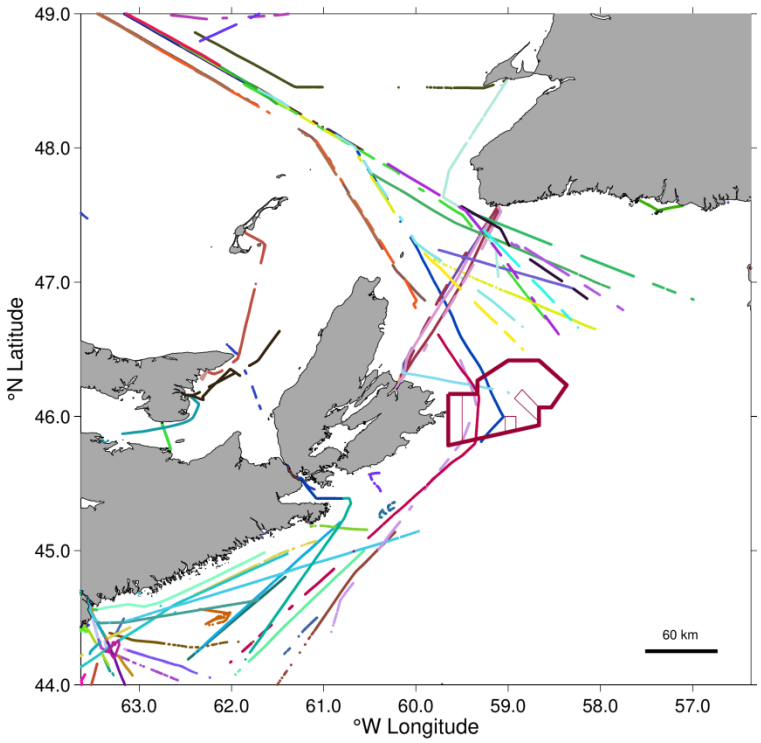

Figure 27. Automatic Identification System (AIS) data collected from the Canadian Coast Guard terrestrial network of AIS receiving stations on December 8, 2015. A total of 127 vessels were detected in the area with each colour representing a unique vessel.

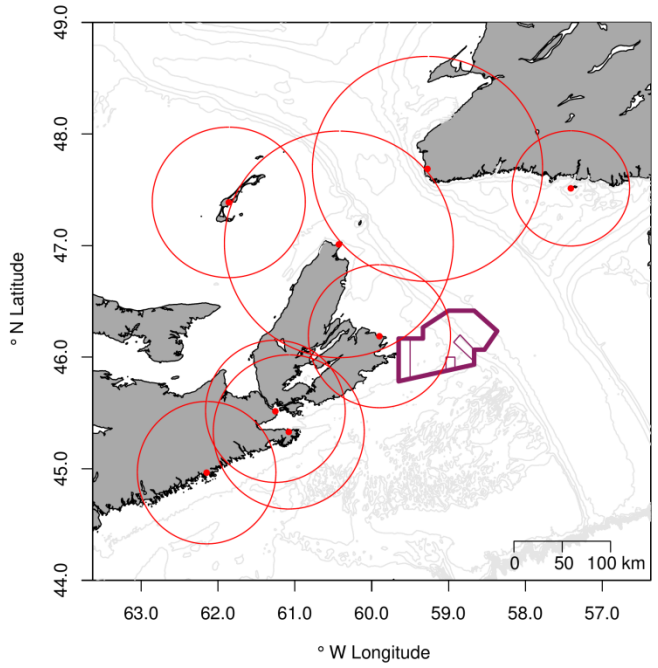

Figure 28. Bathymetic (100 m resolution) chart of the St. Anns Bank area with line of sight detection (red circles) for the terrestrial AIS receiving stations (red dots) around St. Anns Bank Area of Interest.

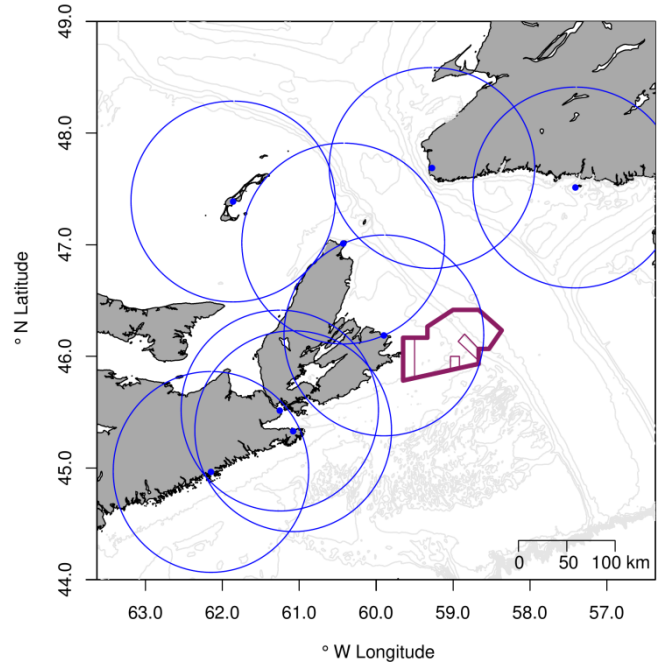

Figure 29. Bathymetic (100 m resolution) chart of the St. Anns Bank area with Simard et al. (2014) estimated vessel detection distances (blue circles) for the terrestrial AIS receiving stations (blue dots) around St. Anns Bank Area of Interest.

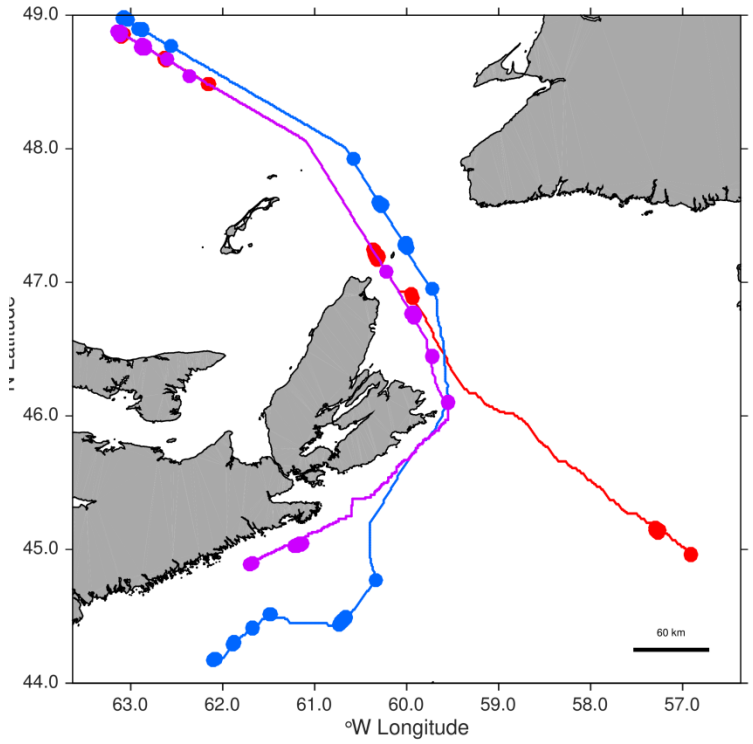

Figure 30. Hypothetical vessel positions (large filled circles) and interpolated vessel positions (lines) based on the A algorithm for three vessels transiting through the St. Anns Bank area. Each colour represents a unique vessel.*

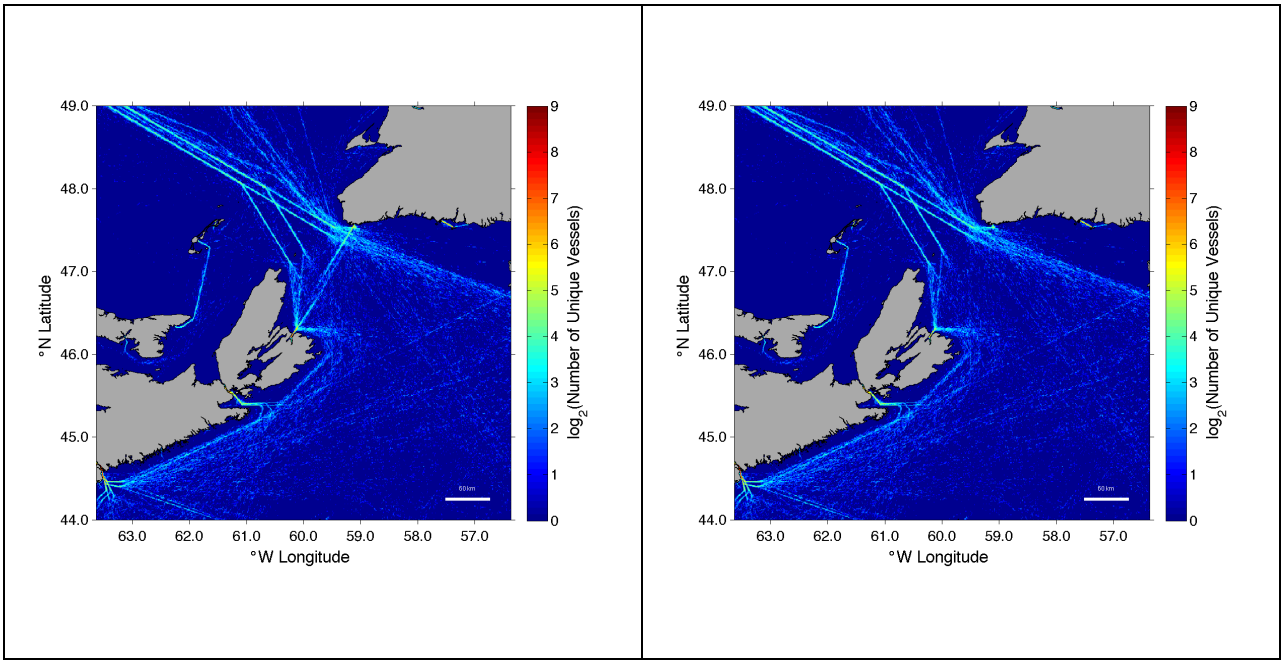

Figure 31. Vessel density maps for the first quarter of a year based on satellite AIS data from 2013-2015 for all vessels (left panel) and all vessels except of the Newfoundland ferries (right panel).

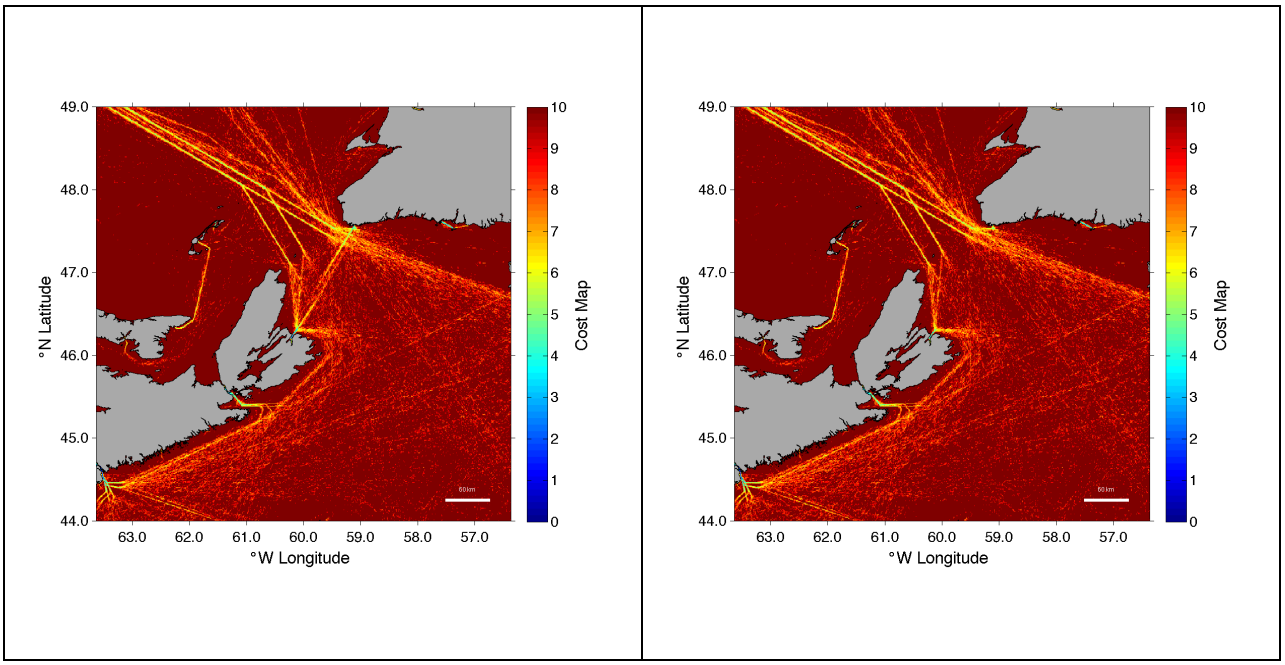

Figure 32. Cost maps developed for the A function to interpolate undetected vessel positions as vessels transit in and out of the Gulf of St. Lawrence.*

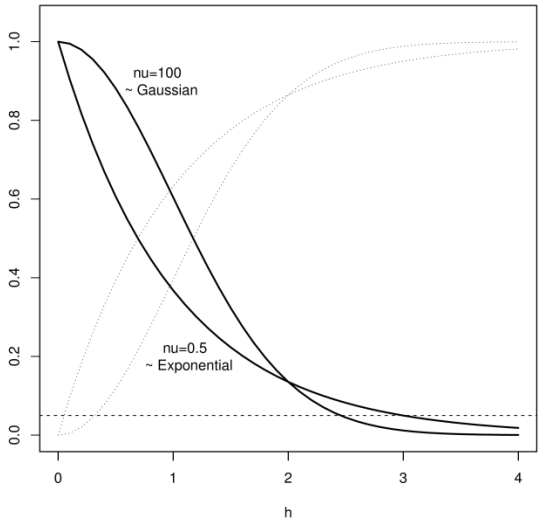

Figure 33. Matérn autocorrelation function, $\rho(h) = C(h)/C(0)$, the covariance function $C(h)$ scaled by the total variance $C(0)$, for two values of ν (dark lines). At $nu = 100$, it approaches the Gaussian curve (upper dark curve on the left side) while at $\nu = 0.5$ the curve is exponential (lower dark curve on the left side). The associated semi-variograms (scaled to unit variance) $\gamma(h)$ are shown in light stippled lines. Spatial scale is defined, in this framework, as the distance h at which the autocorrelation falls to 0.05% (dashed horizontal line) – in this example between 2.5 and 3 units, depending upon value of ν.

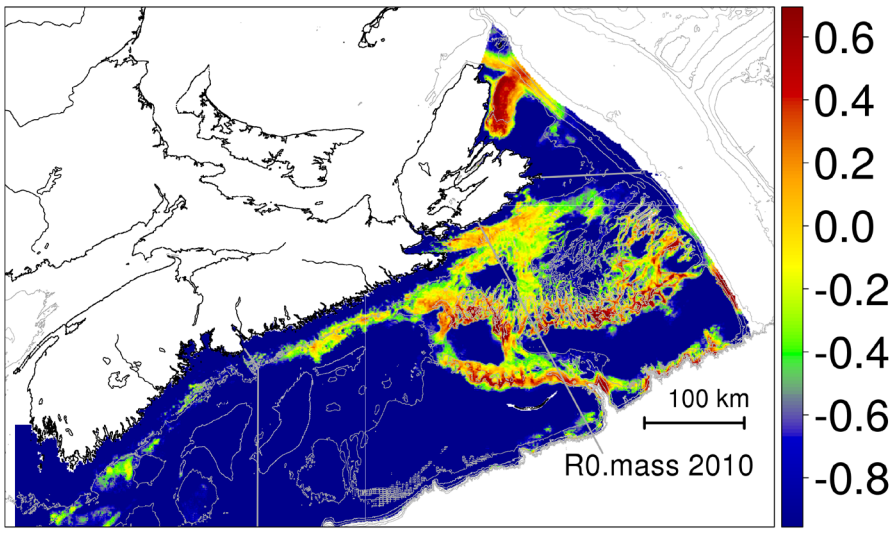

Figure 34. Predicted biomass density of Snow Crab in Maritimes Region based upon a combination of a Functional-habitat method and simple spatial interpolation.

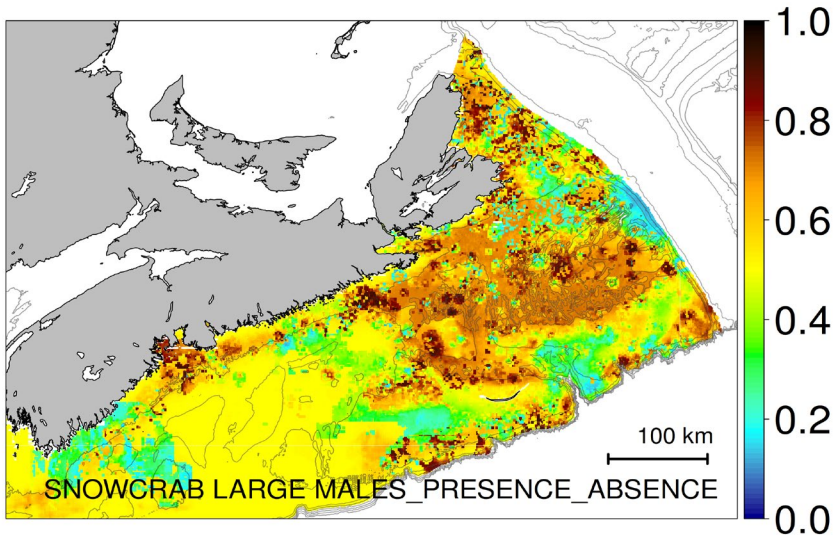

Figure 35. Functional habitat, (H_f), the predicted probability of observing Snow Crab.

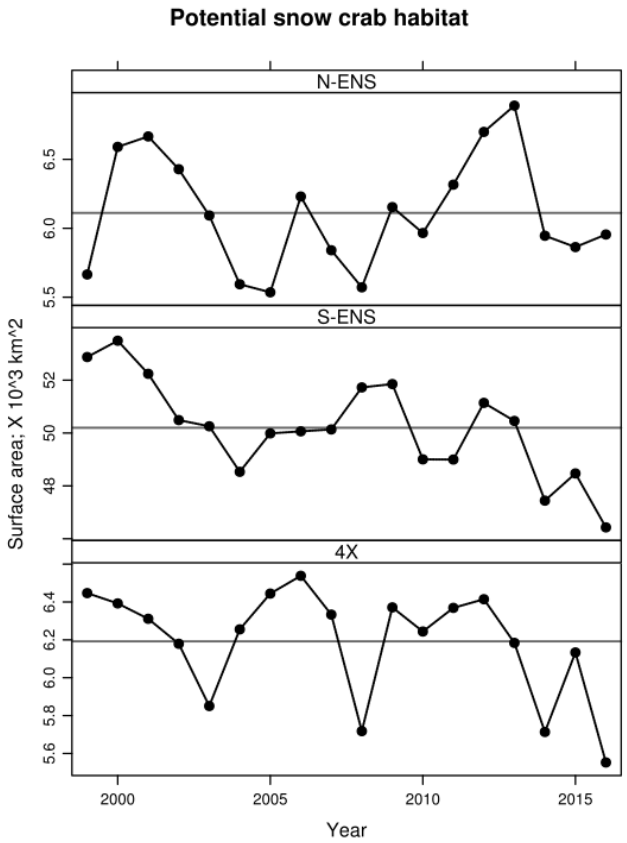

Figure 36. Surface area of potential Functional habitat (H_f) of Snow Crab in Maritimes Region. Note the large interannual variability and a decadal scale decline in the southern areas (lower two panels).

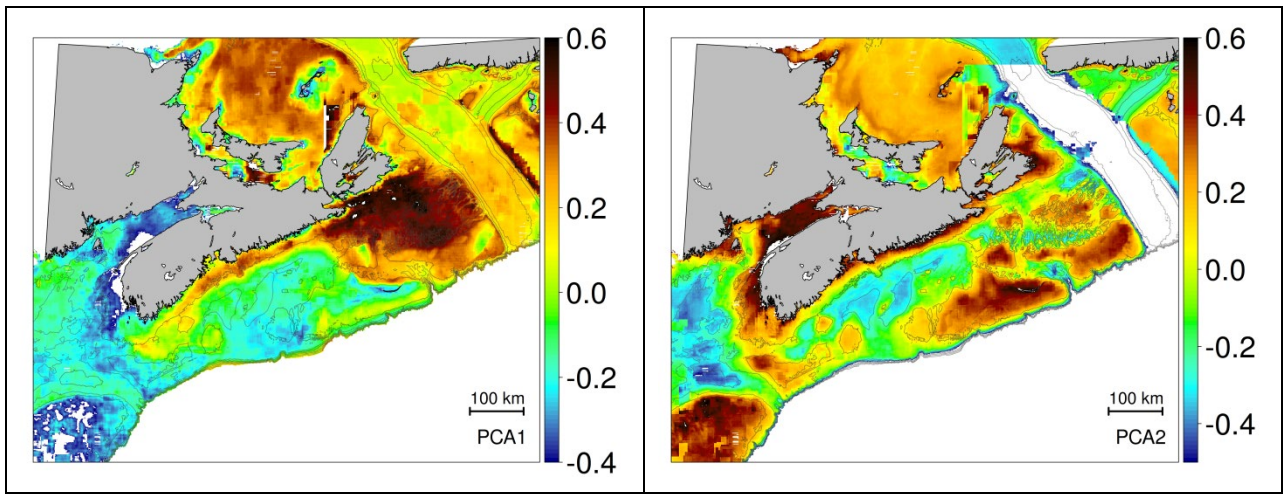

Figure 37. Integral habitat (H_i) based upon species associations in Maritimes Region. Note the first axis is primarily a temperature gradient and the second associated with depth.

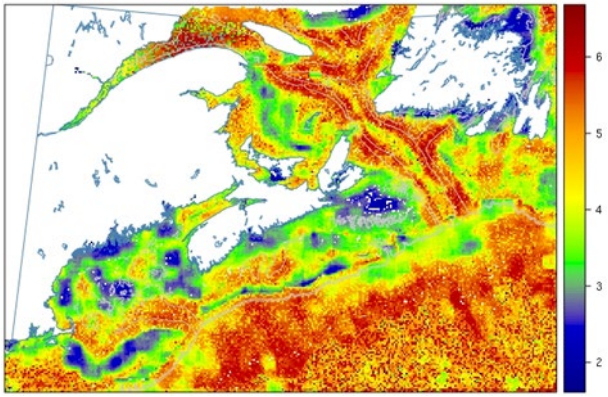

Figure 38. First estimate of log(spatial range; km) based upon depth variations.

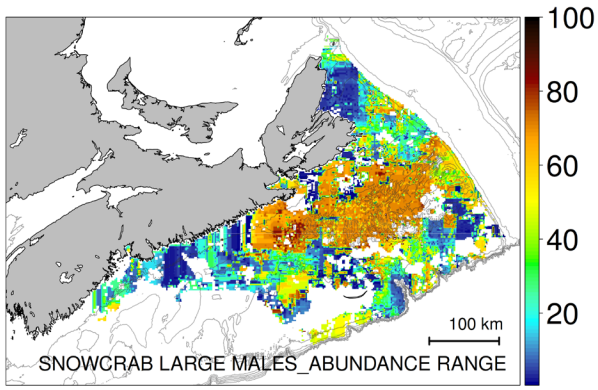

Figure 39. A first estimate of log_e(spatial range; km) of Snow Crab abundance variations.

APPENDICES

APPENDIX 1. DATA QUALITY CONTROL OF AZMP DATA

All data extraction, quality control and processing methods are documented in R scripts found in https://github.com/jae0/aegis/.

DISCRETE BOTTLE DATA

BioChem data are generally not subject to any quality control (QC). As such, substantial QC was required. The QC protocol used was based on procedures designed at DFO's Institut Maurice-Lamontagne (IML). These were in turn based on procedures developed by NOAA's National Oceanographic Data Center/World Ocean Database, as well as many of the tests proposed in the GTSPP Real-Time Quality Control Manual. The quality control procedure was as follows:

Step 1: Impossible Dates

Due to known issues with dates, database query for nutrients and chlorophyll were designed to extract records for which the sampling date is within start and end dates of the mission. Another check includes comparing HEADER_START and EVENT_START dates that should be the same. It is often found that month and day were reversed in EVENT_START field. Those records were retained and HEADER_START was used as a more reliable date.

Step 2: Quality Control Flags

A small number of records in BioChem was subject to quality control and include flags for position (POSITION_QC_CODE) and data (DATA_QC_CODE). The meaning of the codes are as follows:

- 0 = No quality control performed
- 1 = Value appears correct
- 2 = Value appears inconsistent
- 3 = Value appears doubtful
- 4 = Value appears erroneous
- 5 = Value changed as result of quality control

The QC flags were checked for parameter and inconsistent (2), doubtful (3) and erroneous (4) values were removed from the dataset.

Step 3: Depth Check

For bottle data, start and end depth at which the water samples were collected are verified to be the same. Records with different start and end depths were removed from the dataset.

Step 4: Duplicated Records

BioChem often contains duplicated records as the same data was sometimes loaded into a database twice and treated as different records. Duplicated records were removed, and the first record of each duplicate is kept in the dataset.

Step 5: Suspect Missions

Missions with suspect data were identified and removed. Those missions often show unusual data values (for example, integer numbers without decimal places with only some values out of

range), suggesting that the data for the whole mission might be compromised. The suspect mission for each parameter are following:

- Chlorophyll-a OC7908, 32G879008
- Phosphate 18HU88026
- Silicate 180167005, 31TR26870

Chlorophyll and all nutrients values were examined if they fell within expected limits for the Northwest Atlantic, which are adopted from IML quality control procedure (IML Test 2.1). The expected range of values are:

- Chlorophyll-a: 0-50 mg/m^3
- Nitrate: 0-515 mmol/m^3
- Phosphate: 0-4.5 mmol/m^3
- Silicate: 0-250 mmol/m^3

Any values outside of expected range were removed for open ocean data only. Coastal data (up to 5 km from the coast) were not filtered using expected ranges as in coastal water chlorophyll-a and nutrients concentrations can be higher.

Step 6: Profile Envelope

Data for each parameter are checked if they fall within the expected limits by depth interval, as shown in Table A1.1 (IML Test 2.4). This test does not allow zero values for silicate and phosphate in the deep water. Again, only open ocean data were subject to the profile envelope test.

Table A1.1. Expected ranges of parameters for the profile envelope test (IML Test 2.4)

Parameter	Depth Interval	Expected Range
Chlorophyll-a	0-1500 m	0-50 mg/m^3
Silicate	0-150 m	0-250 mmol/m^3
Silicate	150-900 m	0.01-250 mmol/m^3
Phosphate	0-500 m	0-4.5 mmol/m^3
Phosphate	150-1500 m	0.01-4.5 mmol/m^3
Nitrate	0-1500 m	0-515 mmol/m^3

Step 7: Impossible Profiles

This check was not implemented in the code and impossible profiles were identified by investigating unusual outliers.

Additional steps from IML QC procedure, such as checks for constant profile, excessive gradient and inversions, were not implemented in this quality control procedure. However, due to eutrophication from terrestrial sources, phosphate levels in coastal regions often exceeded the upper limits for globally and locally possible values, with the phosphate concentrations sometimes six times higher than the upper limits for the Northwest Atlantic in offshore waters.

Therefore, coastal and open ocean data were examined separately; non-coastal data were filtered using the expected limits for the Northwest Atlantic. Coastal data were defined as the ones collected less than 5 km away from the coast, where a 5 km limit was chosen as an optimal distance at which all coastal inlets are included. Buffer polygons along the coastline were created (Figure A1.1) and used for flagging the data as open ocean records (Flag 1) and coastal records (Flag 2).

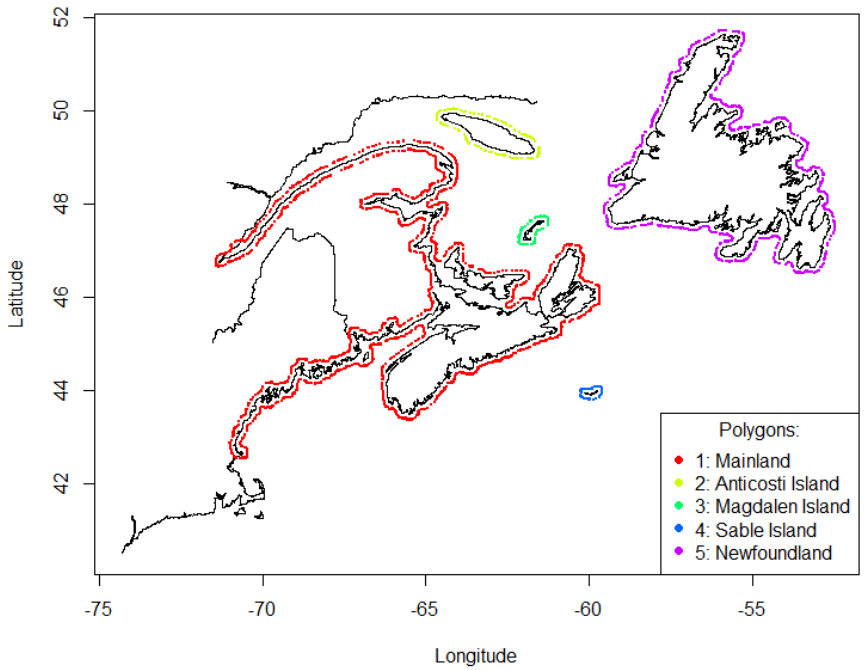

Figure A1.1. Polygons used for separation of coastal and ocean data, arbitrarily assumed to be a 5 km buffer from the coastline.

CHLOROPHYLL-A

Chlorophyll-a data are derived from four methods. The methods are listed and described in Table A1.2 and the aggregate time series associated with each method is shown in Figure A1.2. For most of the chlorophyll data, the method is not specified (unknown); Holm-Hansen fluorometric method is the standard AZMP method and is the second most frequent; Welschmeyer fluorometric method is used least frequently, often by the Quebec and Newfoundland regions.

In a number of cases, the same water sample was processed using two different methods, resulting in two sets of chlorophyll estimates for the same samples. Comparisons between these two sets of values are shown in Figure A1.3. In both cases, chlorophyll-a estimated by the Welschmeyer method are lower than the ones using the Holm-Hansen method or “unknown” method. Since there is more data mapped to the Holm-Hansen method than to the Welschmeyer method, only data derived from Chl_a and Chl-a_Holm_Hansen_sF methods were retained. No corrections were applied to correct for differences in methodology.

Table A1.2. Methods associated with chlorophyll-a records in BioChem.

Method	Description
Chl_a	Unknown method
Chl_a_Holm-Hansen_F	Holm-Hansen method; Prefiltered; Frozen before analysis (-20)
Chl_a_Holm-Hansen_sF	Holm-Hansen method; Super Frozen before analysis (-196)
Chl_a_Welschmeyer_sF	Welschmeyer method; Super Frozen before analysis (-196)

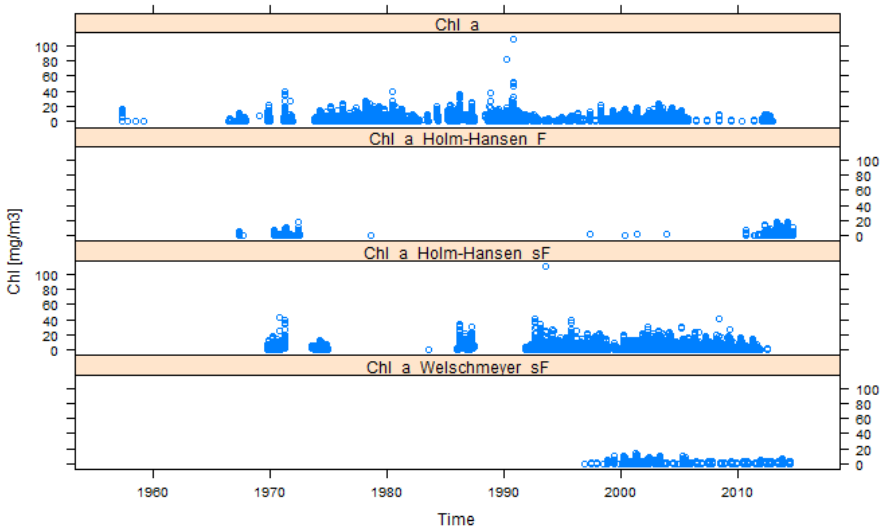

Figure A1.2. Time series of chlorophyll-a data from BioChem grouped by methods.

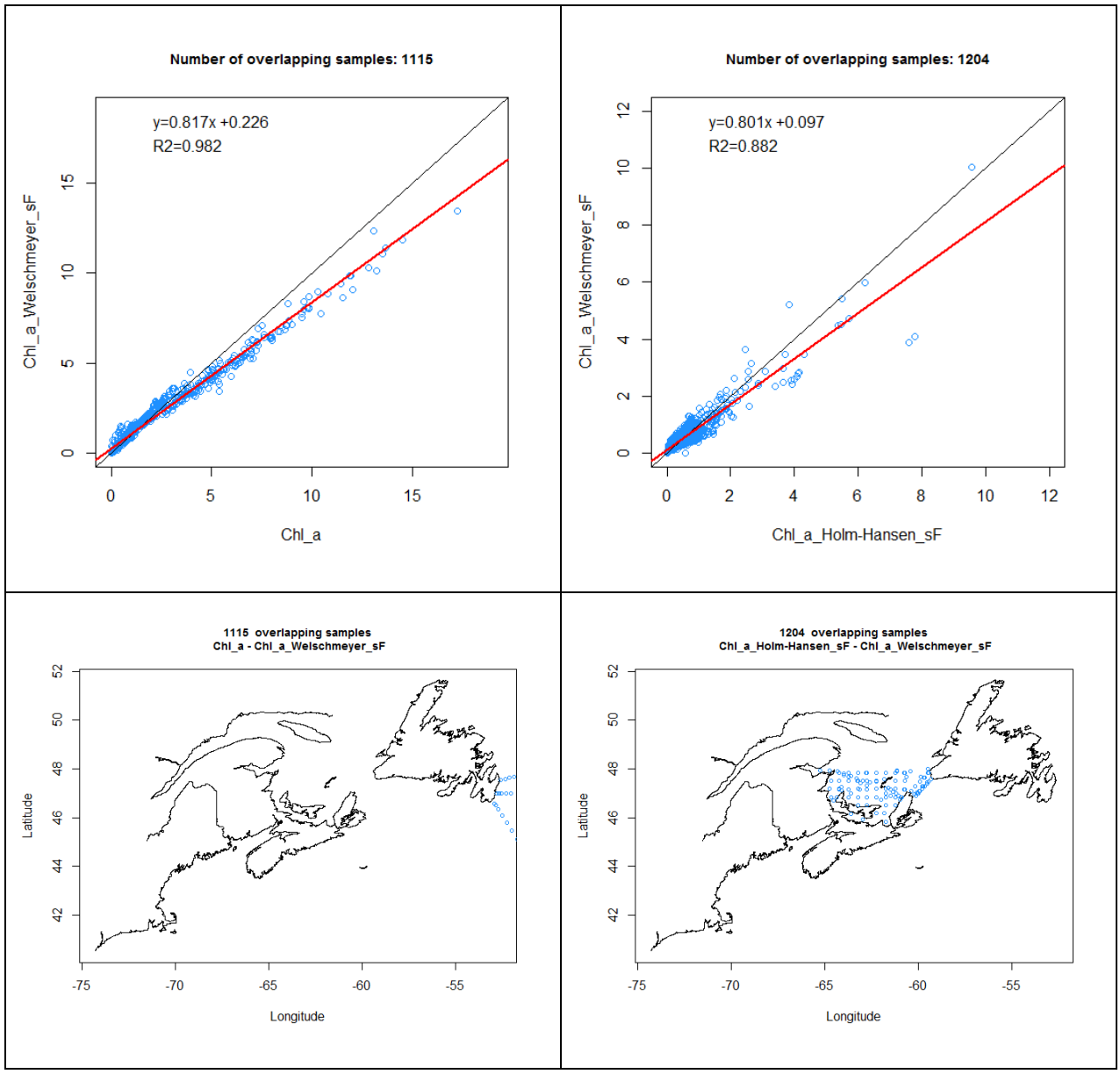

Figure A1.3. Comparison of the chlorophyll-a values collected using different methods is shown on the scatter plots on the top panel and the geographical locations of those samples is shown on the maps in the bottom panel.

NITRATE

Nitrate estimates are derived from 10 methods. The methods are listed and described in Table A1.3 and the time series of data associated with each method is shown in Figure A1.4. Most of the methods measure nitrate and nitrite together. We also included data for nitrate only since in most seawater the concentration of nitrite is small compared to that of nitrate.

Table A1.3. Methods associated with nitrate records in BioChem.

Method	Description
NO2NO3_0	Nitrate+Nitrite / Unknown method
NO2NO3_Alp_F	Nitrate+Nitrite / Alpchem / Frozen
NO2NO3_Alp_SF	Nitrate+Nitrite / Alpchem / SuperFrozen
NO2NO3_S&P1968	Nitrate+Nitrite/ S&P(1968) / filtered and frozen
NO2NO3_Tech_F	Nitrate+Nitrite / Technicon / Frozen
NO2NO3_Tech_Fsh	Nitrate + Nitrite / Technicon / Fresh / Strain / Unfiltered
NO2NO3_Tech_SF	Nitrate+Nitrite / Technicon / SuperFrozen
NO2NO3_Tech2_F	Nitrate+Nitrite / Technicon2 / Frozen
NO3_Tech_F	Nitrate / Technicon / Frozen, corrected for NO2
NO3_Tech_SF	Nitrate / Technicon / SuperFrozen

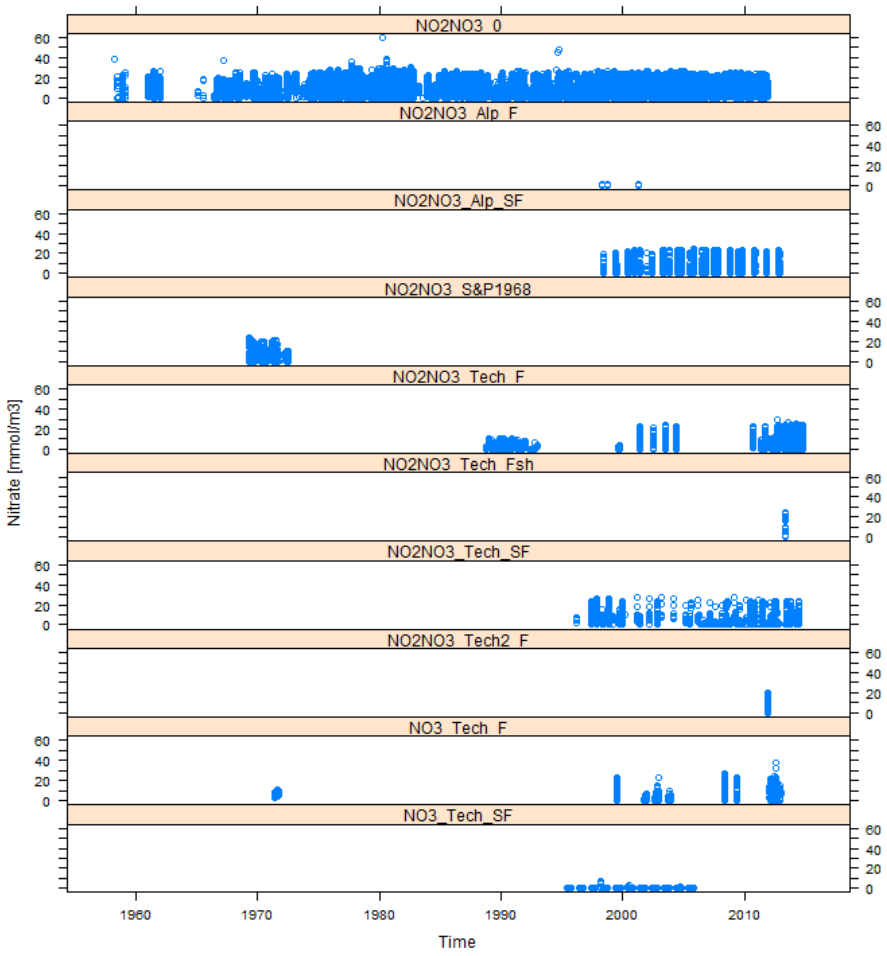

Figure A1.4. Time series of nitrate data from BioChem grouped by methods.

PHOSPHATE

Phosphate data available in BioChem are mapped to 7 methods. The methods are listed and described in Table A1.4 and the time series of data associated with each method is shown in Figure A1.5.

Table A1.4. Methods associated with phosphate records in BioChem.

Method	Description
PO4_0	Phosphate / Unknown method
PO4_Alp_SF	Phosphate / Alpchem / SuperFrozen / Filtered
PO4_Tech_2	Phosphate / Murphy and Riley / filtered and frozen
PO4_Tech_F	Phosphate / Technicon / Frozen / Unfiltered
PO4_Tech_Fsh	Phosphate / Technicon / Fresh / Strain / Unfiltered
PO4_Tech_SF	Phosphate / Technicon / SuperFrozen / Filtered
PO4_Tech2_F	Phosphate / Technicon2 / Frozen / Unfiltered

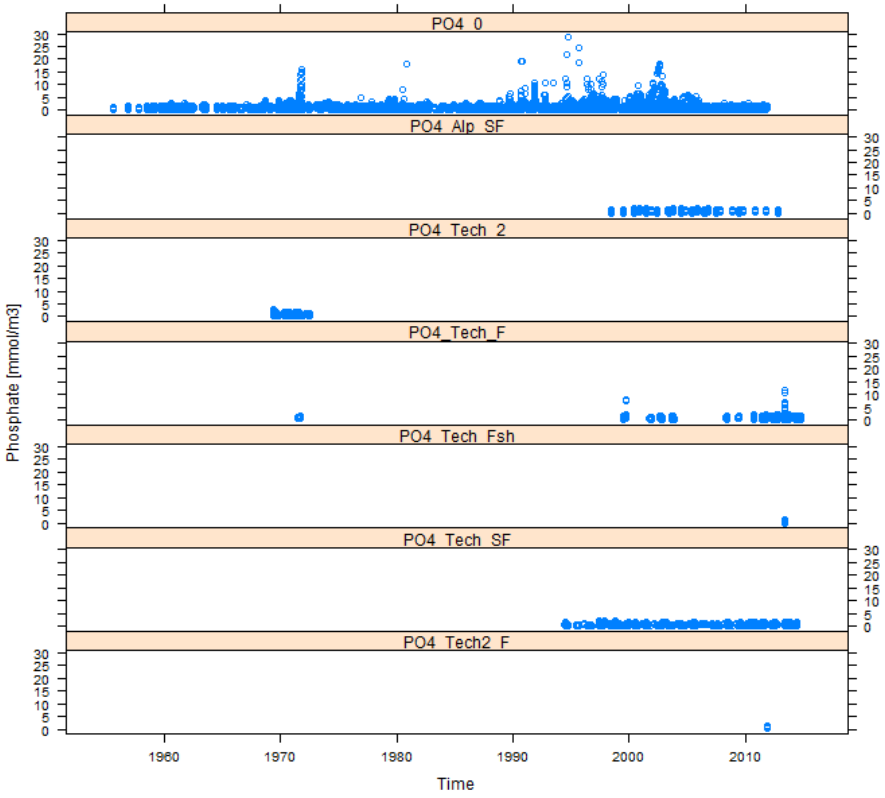

Figure A1.5. Time series of phosphate data from BioChem grouped by methods.

SILICATE

Silicate data available in BioChem are mapped to 8 methods (Table A1.5). Their time series of are shown on Figure A1.6.

Table A1.5. Methods associated with silicate records in BioChem.

Method	Description
SiO4_0	Silicate, Unknown methods and handling
SiO4_1	Silicate / Mullin and Riley / filtered and frozen
SiO4_Alp_F	Silicate / Alpchem / Frozen / Unfiltered
SiO4_Alp_SF	Silicate / Alpchem / SuperFrozen / Filtered
SiO4_Tech_F	Silicate / Technicon / Frozen / Strain / Unfiltered
SiO4_Tech_Fsh	Silicate / Technicon /Fresh /Strain / Unfiltered
SiO4_Tech_SF	Silicate / Technicon / SuperFrozen / Filtered
SiO4_Tech2_F	Silicate / Technicon2 / Frozen

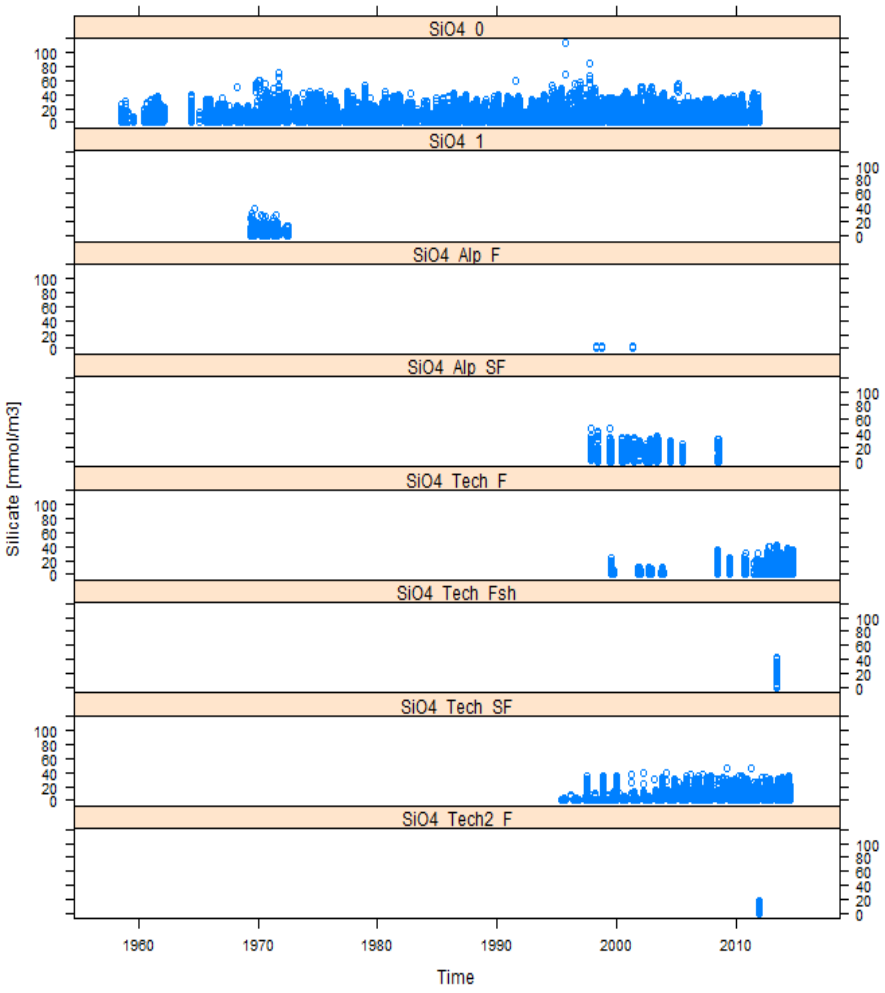

Figure A1.6. Time series of silicate data from BioChem grouped by methods.

ZOOPLANKTON

Zooplankton data was extracted from DFO's BioChem database (Devine et al. 2014) from 1914 to 2014. They were comprised of 687 missions and 53,787 samples, using 13 different kinds of nets, 35 different mesh sizes ranging from 20 microns to 4.23 mm, and various net deployment and sample processing protocols. To ensure data consistency and comparability, only samples collected and analyzed using Atlantic Zone Monitoring Program (AZMP) protocol (Mitchell et al. 2002) were retained for study, reducing the time scope to 1999-2014.

As AZMP samples are not always properly flagged in the BioChem database, a list of missions that followed the AZMP protocol were provided by Ocean and Ecosystem Science Division (OESD) and Ocean Data and Information Services (ODIS). The relevant missions include AZMP spring and fall cruises, summer and winter groundfish survey missions, bi-weekly sampling at fixed stations (Halifax Station 2 and Prince 5) and samples collected on the Scotian Shelf during Labrador Sea missions.

AZMP protocol samples zooplankton with Ring nets (0.75 m diameter, mesh size of 202 um) deployed as vertical tows from either near bottom or 1000 m (whichever is shallower) to the surface. The sample analysis includes estimation of abundance, species composition and

biomass in terms of wet and dry weight in two size fractions; one for organisms ranging from 0.2 mm to 10 mm in size and the other for all organisms larger than 10 mm. The protocol is as follows:

- Organisms larger than 10 mm are manually separated from the sample, identified and counted. Wet weight is determined and reported for each individual species. In addition total wet weight for all large organisms is reported as the sum of all individual wet weights.
- Captured organisms smaller than 10 mm (0.2 - 10 mm fraction) are identified and counted. Dry and wet weight is determined and reported for the whole sample containing all organisms in that size fraction.
- Total wet weight is reported for all captured organisms as the sum of wet weights of large and small organisms.
- Developmental stages are identified for *Calanus finmarchicus*, *C. glacialis* and *C. hyperboreus*.

Since the data hosted in BioChem is generally not subject to quality control and can contain erroneous records, substantial quality control was conducted to ensure correct representation of actual measurements. The quality control included verification of the following fields:

- time stamps: comparing mission dates with header dates and event dates start and end depths of the nets that cannot be equal or close together;
- volume of the samples: all records with volumes 0, or NA were removed;
- split fraction of the sample: all the records with split fraction NA, 0 or > 1 were removed;
- minimum and maximum sieve for dry weight records: records NA sieve were removed; and
- repeated records.

Finally, the numerical and biomass density for each species per unit surface area of a tow was computed as follows:

$$\text{abundance} = \text{counts} \times \frac{abs(\text{depth}_{\text{start}} - \text{depth}_{\text{end}})}{\text{split fraction} \times \text{volume}} \quad [\text{individuals}/\text{m}^2]$$

$$\text{biomass} = \text{weight} \times \frac{abs(\text{depth}_{\text{start}} - \text{depth}_{\text{end}})}{\text{split fraction} \times \text{volume}} \quad [\text{g}/\text{m}^2],$$

where counts refer to number of organisms encountered in the sample, $\text{depth}_{\text{start}}$ and $\text{depth}_{\text{end}}$ to the start and end depth of the net deployment, split fraction to the fraction of the sample analyzed and volume to the sample volume. The final filtered dataset includes 126 missions in the time period 1999 to 2014, with 2,367 net deployments and more than 400 taxonomic species.

APPENDIX 2. MATÉRN FUNCTION

The Matérn *correlation* function (Figure 33) is parameterised in a number of ways, and nomenclature of variables have also been inconsistent and can potentially cause confusion. The geoR library (Diggle and Ribeiro 2007; Ribeiro and Diggle 2001) defines it as:

$$\rho(x) = \frac{1}{2^{\kappa-1}\Gamma(\kappa)}\left(\frac{x}{\phi}\right)^{\kappa} K_{\kappa}\left(\frac{x}{\phi}\right),$$

where, $\phi > 0$ is the "range parameter"; $\kappa > 0$ is the shape (smoothness) parameter; $K_{\kappa}(\cdot)$ is the modified Bessel function of the third kind of order κ; and $\Gamma(\cdot)$ is the Gamma function. It is also related to the fractal dimension of the surface complexity (Constantine and Hall 1994).

spBayes's (Finley et al. 2007) parameterization just a little different as well as is the nomenclature:

$$\rho(x) = \frac{1}{2^{\nu-1}\ \Gamma(\nu)}(\phi x)^{\nu} K_{\nu}(\phi x).$$

INLA's (Lindgren and Rue 2015) parameterization is the same as spBayes', but nomenclature is slightly different:

$$\rho(x) = \frac{1}{2^{\lambda-1}\ \Gamma(\lambda)}(\kappa x)^{\lambda} K_{\lambda}(\kappa x).$$

Thus, the following identities are important to interpret the discussions and outputs from these authors:

$$\begin{aligned} \kappa_{geoR} &\equiv \nu_{spBayes} \equiv \lambda_{INLA} \\ 1/\phi_{geoR} &\equiv \phi_{spBayes} \equiv \kappa_{INLA}. \end{aligned}$$

Additionally, INLA defines $range_{13\%} = \sqrt{8\lambda}/\kappa$ (see rationale below).

The Matérn *covariance* function, $\gamma(x)$ is the correlation function $\rho(x)$ scaled by the variance. An offset of τ^2 the variability at local scales associated with sampling error smaller than the unit of measurement (sometimes called "nugget" variance) is also modeled:

$$\gamma(x) = \tau^2 + \sigma^2(1 - \rho(x)).$$

When $\nu_{spBayes} = 1/2$, this become the exponential model. When $\nu_{spBayes} \to \infty$, the model becomes the Gaussian model. The curves become only incrementally different once $\nu_{spBayes} > 2$ and so Finley et al. (2007) suggests limiting the prior to the interval (0,2) as a pragmatic solution to speeding up MCMC convergence. We have used the interval $\nu_{spBayes} = (0{,}5)$, just to be sure. Further, the effective range of spatial dependence (distance at which the correlation drops to 0.05 is given by $-ln(0.05)/\phi_{spBayes}$ (Finley, Banerjee, and Carlin 2007). As such, they recommend a uniform prior in the interval of $(\phi_{spBayes}, -ln(0.05))$.

The spatial scale we define as exactly this practical or effective spatial range: the distance at which spatial dependence drops asymptotically to $\rho(x) \to 0.05$.

One of the reasons for the popularity of the Matérn covariance is that it is a stationary solution to the SPDE:

$$(\kappa^2 - \Delta)^{\alpha/2}(\tau\xi(s)) = W(s),$$

where $W(s)$ is Gaussian spatial white noise and τ controls the variance and κ is the scale parameter. INLA defines the effective range at $\sqrt{8}/\phi_{spBayes}$, the approximate distance where correlation falls to 0.13 (vs 0.05):

$$\begin{cases} \lambda & = \alpha - d/2 \\ \sigma^2 & = \dfrac{\Gamma(\lambda)}{\Gamma(\alpha)(4\pi)^{d/2}\kappa^{2\lambda}\tau^2} \end{cases},$$

which for d=2 dimensions:

$$\begin{cases} \lambda & = \alpha - 1 \\ \sigma^2 & = \dfrac{\Gamma(\lambda)}{\Gamma(\alpha)(4\pi)\kappa^{2\lambda}\tau^2} \end{cases},$$

and $\alpha = 2$ by default and so $\lambda = 1$ such that:

$$\begin{cases} range & = \sqrt{8}/\kappa \\ \sigma^2 & = \dfrac{1}{4\pi\kappa^2\tau^2} \end{cases},$$

$$\text{range} = \sqrt{8}/\kappa.$$

APPENDIX 3. ADVECTION-DIFFUSION STOCHASTIC PARTIAL DIFFERENTIAL EQUATION (SPDE)

Following the development of Sigrist et al. (2015), we begin with the general regression model

$$Y(\mathbf{s},t) = \mu(\mathbf{s},t) + e(\mathbf{s},t),$$

where, $\mu(\mathbf{s},t) = x(\mathbf{s},t)^T\beta(\mathbf{s},t)$ is the mean process and $e(\mathbf{s},t)$ the residual error. This error process can be separated into a spatiotemporally structured component ω and an unstructured component ε: $e(\mathbf{s},t) = \omega(\mathbf{s},t) + \varepsilon(\mathbf{s},t)$. The *unstructured* error is usually assumed to be a white error process: $\varepsilon(\mathbf{s},t) \sim N(0,\sigma_\varepsilon^2)$.

The structured error $\omega(\mathbf{s},t)$ can be defined in terms of the following advection-diffusion SPDE:

$$\frac{\partial}{\partial t}\omega(\mathbf{s},t) = -\mathbf{u}^T\nabla\omega(\mathbf{s},t) + \nabla\cdot\Sigma\nabla\omega(\mathbf{s},t) - \zeta\omega(\mathbf{s},t) + \epsilon(\mathbf{s},t),$$

where, $\mathbf{s} = (x,y)^T \in \Re^2$: $\mathbf{u} = (u_x,u_y)^T$ parameterizes the drift velocity (advection); $\nabla = (\frac{\partial}{\partial x},\frac{\partial}{\partial y})^T$ is the gradient operator; $\nabla\cdot\mathbf{F} = (\frac{\partial F_x}{\partial x},\frac{\partial F_y}{\partial y})^T$ is the divergence operator for a vector field $\mathbf{F} = (F_x,F_y)^T$; $\Sigma^{-1} = \frac{1}{\phi_d^2}\begin{pmatrix}\cos\alpha & sin\alpha\\ -\gamma\cdot\sin\alpha & \gamma\cdot\cos\alpha\end{pmatrix}^T\begin{pmatrix}\cos\alpha & sin\alpha\\ -\gamma\cdot\sin\alpha & \gamma\cdot\cos\alpha\end{pmatrix}$ parameterizes the anisotropy in diffusion via ($\gamma > 0$, $\alpha \in [0,\pi/2]$) with $\phi_d > 0$ parameterizing the diffusion range; $\zeta > 0$ parameterizing local damping; and $\epsilon(\mathbf{s},t)$ parameterizing a Gaussian random field that accounts for source-sink processes with white noise in time and Matérn spatial covariance (aka, "innovation").

If $\epsilon(\mathbf{s},t)$ is stationary and has the form of a Whittle spatial covariance function, then it has the following spectrum:

$$\widetilde{f}(\mathbf{k}) = \frac{\sigma^2}{(2\pi)^2}(\mathbf{k}^T\mathbf{k} + \frac{1}{\phi_s^2})^{-(\nu+1)},$$

where $\mathbf{k}$ are the spatial wave numbers. The spectrum of the $\omega(\mathbf{s},t)$ process is then:

$$f(w,\mathbf{k}) = \widetilde{f}(\mathbf{k})\frac{1}{(2\pi)}((\mathbf{k}^T\Sigma\mathbf{k}+\zeta)^2 + (w+\mathbf{u}^T\mathbf{k})^2)^{-1},$$

where w is temporal frequency. The covariance function can be recovered as:

$$\begin{aligned} C(\mathbf{s},t) &= \int f(w,k)\ exp(i\cdot tw)\ exp(i\cdot\mathbf{s}'\mathbf{k})\ d\mathbf{k}dw \\ &= \int \widetilde{f}(\mathbf{k})\frac{exp(-i\cdot\mathbf{u}^T\mathbf{k}t - (\mathbf{k}^T\Sigma\mathbf{k}+\zeta)|t|)}{2(\mathbf{k}^T\Sigma\mathbf{k}+\zeta)}\ exp(i\cdot\mathbf{s}'\mathbf{k})\ d\mathbf{k}. \end{aligned}$$

Upon discretization:

$$\begin{aligned} \omega(\mathbf{s},t) &= \boldsymbol{\phi}(\mathbf{s})^T\alpha(t) \\ &\approx \sum_{j=1}^{K}\phi_j(\mathbf{s})\alpha_j(t) \\ &\approx \sum_{j=1}^{4}\phi_j^{(cos)}(\mathbf{s}_l)\alpha_j^{(cos)}(t) + \sum_{j=5}^{K/2+2}(\phi_j^{(cos)}(\mathbf{s}_l)\alpha_j^{(cos)}(t) + \phi_j^{(sin)}(\mathbf{s}_l)\alpha_j^{(sin)}(t)), \end{aligned}$$

where $\mathbf{k} = (k_{x,j}, k_{y,j})^T$ is the spatial wavenumber of the j component of K Fourier components and $\phi_j(\mathbf{s}) = \exp(\iota \cdot \mathbf{k}_j^T \mathbf{s}) = \{\cos(\mathbf{k}_j^T \mathbf{s}), \sin(\mathbf{k}_j^T \mathbf{s})\}$ is the spatial function. And, for each time t and location s_l, $l = 1, \ldots, n^2$ in the SPDE:

$$\begin{aligned} \mathbf{u}^T \nabla \phi_j^{(cos)}(\mathbf{s}_l) &= -\mathbf{u}^T \mathbf{k}_j \phi_j^{(sin)}(\mathbf{s}_l) \\ \mathbf{u}^T \nabla \phi_j^{(sin)}(\mathbf{s}_l) &= -\mathbf{u}^T \mathbf{k}_j \phi_j^{(cos)}(\mathbf{s}_l), \end{aligned}$$

and,

$$\begin{aligned} \nabla \cdot \Sigma \nabla \phi_j^{(cos)}(\mathbf{s}_l) &= -\mathbf{k}_j^T \Sigma \mathbf{k} \phi_j^{(cos)}(\mathbf{s}_l) \\ \nabla \cdot \Sigma \nabla \phi_j^{(sin)}(\mathbf{s}_l) &= -\mathbf{k}_j^T \Sigma \mathbf{k} \phi_j^{(sin)}(\mathbf{s}_l). \end{aligned}$$

The full non-separable, advection-diffusion model specification is thus:

$$\begin{aligned} Y(\mathbf{s},t) &= x(\mathbf{s},t)^T \boldsymbol{\beta} + \omega(\mathbf{s},t) + \varepsilon(\mathbf{s},t) \\ \omega(\mathbf{s},t) &= \boldsymbol{\Phi}\boldsymbol{\alpha}(\mathbf{s},t) \quad \{\text{advection} - \text{diffusion model}\} \\ \alpha(\mathbf{s},t) &= \mathbf{G}\boldsymbol{\alpha}(\mathbf{s},t-1) + \hat{\varepsilon}\,(\mathbf{s},t) \quad \{\text{transition model}\} \\ \varepsilon(\mathbf{s},t) &\sim N(\mathbf{0}, \tau^2 \mathbf{1}) \quad \{\text{unstructured error}\} \\ \hat{\varepsilon}\,(\mathbf{s},t) &\sim N(\mathbf{0}, \widehat{\mathbf{Q}}) \quad \{\text{innovation}\}, \end{aligned}$$

where, $\boldsymbol{\alpha}(\mathbf{s},t)$ are Fourier coefficients; $\boldsymbol{\Phi} = [\boldsymbol{\phi}(\mathbf{s}_1), \ldots, \boldsymbol{\phi}(\mathbf{s}_N)]^T$ is a matrix of spatial basis functions; $\mathbf{G}$ is the transition (propagator) matrix; and $\widehat{\mathbf{Q}}$ is the innovation covariance matrix (residual errors). The Fourier functions are:

$$\boldsymbol{\phi}(\mathbf{s}_l) = \left(\cos(\mathbf{k}_1^T \mathbf{s}_l), \ldots, \cos(\mathbf{k}_4^T \mathbf{s}_l), \cos\left(\mathbf{k}_5^T \mathbf{s}_l\right), \sin\left(\mathbf{k}_5^T \mathbf{s}_l\right) \ldots, \cos\left(\mathbf{k}_{\frac{K}{2}+2}^T \mathbf{s}_l\right), \sin\left(\mathbf{k}_{\frac{K}{2}+2}^T \mathbf{s}_l\right)\right)$$

APPENDIX 4. LOGISTIC MODEL

The derivation of the logistic model presented below is more phenomenological than mechanistic and attributed to Lotka (1925). It is our opinion that this phenomenological interpretation renders it useful as a general characterisation of system state. The argument is that the rate of change in time t of any state Y (in our context: abundance, species richness, "habitat", size) can be expected to be in some way, a function of itself:

$$dY/dt = g(Y).$$

It is expected that when $Y = 0$, dY/dt will also be zero and so represents an algebraic root of g. A Taylor series expansion of g near this root $Y = 0$ gives:

$$\begin{aligned} dY/dt &= g'(0)Y + g''(0)Y^2/2 + \text{higher order terms} \ldots; \\ &\approx Y[g'(0) + g''(0)Y/2]. \end{aligned}$$

With the identities $g'(0) = r$ and $g''(0) = -2r/K$, the standard form of the logistic equation is obtained:

$$dY/dt \approx rY(1 - Y/K).$$

The intrinsic rate of increase, r, is, therefore, some abstract and aggregate function that describes the net increase or decrease of the system state Y when Y is small. In biological systems, this means a maximum rate of growth, recruitment, mortality, movement, climatic change, extinction, speciation, etc. as the rate is expected to decline as it approaches some upper limit K of the magnitude of the system state Y.

With normalization by K such that $y = Y/K$:

$$dy/dt \approx ry(1 - y).$$

Many variations of this basic model are known, mostly different ways of adjusting the shape of the curve and/or the location of the inflection point (i.e., $K/2$). A flexible family is the gamma-logistic:

$$dy/dt = ry^{\alpha}(1 - y)^{\beta} + f(\ldots).$$

The additional parameters exponentiate different components which ultimately amounts to adding higher order and even fractional order terms in the Taylor series) and also adding additional terms $f(\ldots)$ that are external to the dynamic that r and K govern, such as fishing, advection, diffusion, noise. For the purposes of this discussion, we will use only the basic model and estimate the parameters of interest, $\theta = \{r, K\}$, but other parameterizations are possible, such as $\theta = \{r, K, \alpha, \beta\}$. The intent is to explore the utility of this more flexible formulation in conjunction with the basic model.

In discrete form, where $\Delta t = 1$ year and after a Euler discretization, the basic model becomes:

$$y_t \approx r y_{t-1}(1 - y_{t-1}).$$

This, we call the "basic" form of the discrete logistic model.