Analysis of Automatic Identification System (AIS) Data for the Eastern Shore Islands, Nova Scotia, 2017-2018

Christine M. Konrad

Fisheries and Oceans Canada
Aquatic Ecosystems Branch, Maritimes Region
Bedford Institute of Oceanography
1 Challenger Drive, P.O. Box 1006
Dartmouth, Nova Scotia
B2Y 4A2

2020

Canadian Manuscript Report of Fisheries and Aquatic Sciences 3172

Fisheries and Oceans Canada Pêches et Océans Canada

Canada

Canadian Manuscript Report of Fisheries and Aquatic Sciences

Manuscript reports contain scientific and technical information that contributes to existing knowledge but which deals with national or regional problems. Distribution is restricted to institutions or individuals located in particular regions of Canada. However, no restriction is placed on subject matter, and the series reflects the broad interests and policies of Fisheries and Oceans Canada, namely, fisheries and aquatic sciences.

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

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

Numbers 1-900 in this series were issued as Manuscript Reports (Biological Series) of the Biological Board of Canada, and subsequent to 1937 when the name of the Board was changed by Act of Parliament, as Manuscript Reports (Biological Series) of the Fisheries Research Board of Canada. Numbers 1426 - 1550 were issued as Department of Fisheries and Environment, Fisheries and Marine Service Manuscript Reports. The current series name was changed with report number 1551.

Rapport manuscrit canadien des sciences halieutiques et aquatiques

Les rapports manuscrits contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui traitent de problèmes nationaux ou régionaux. La distribution en est limitée aux organismes et aux personnes de régions particulières du Canada. II n'y a aucune restriction quant au sujet; de fait, la série reflète la vaste gamme des intérêts et des politiques de Pêches et Océans Canada, c'est-à-dire les sciences halieutiques et aquatiques.

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

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

Les numéros 1 à 900 de cette série ont été publiés à titre de Manuscrits (série biologique) de l'Office de biologie du Canada, et après le changement de la désignation de cet organisme par décret du Parlement, en 1937, ont été classés comme Manuscrits (série biologique) de l'Office des recherches sur les pêcheries du Canada. Les numéros 901 à 1425 ont été publiés à titre de Rapports manuscrits de l'Office des recherches sur les pêcheries du Canada. Les numéros 1426 à 1550 sont parus à titre de Rapports manuscrits du Service des pêches et de la mer, ministère des Pêches et de l'Environnement. Le nom actuel de la série a été établi lors de la parution du numéro 1551.

Canadian Manuscript Report of
Fisheries and Aquatic Sciences 3172

2020

ANALYSIS OF AUTOMATIC IDENTIFICATION SYSTEM (AIS) DATA FOR THE EASTERN SHORE ISLANDS, NOVA SCOTIA, 2017-2018

by

Christine M. Konrad

Fisheries and Oceans Canada
Aquatic Ecosystems Branch, Maritimes Region
Bedford Institute of Oceanography
1 Challenger Drive, P.O. Box 1006
Dartmouth, Nova Scotia
B2Y 4A2

© Her Majesty the Queen in Right of Canada 2020

Cat. No. Fs97-4/3172E-PDF ISBN 978-0-660-29771-2 ISSN 1488-5387

Correct citation for this publication:

Konrad, C.M. 2020. Analysis of Automatic Identification System (AIS) Data for the Eastern Shore Islands, Nova Scotia, 2017-2018. Can. Manuscr. Rep. Fish. Aquat. Sci. 3172: vi + 39 p.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Konrad, C.M. 2020. Analysis of Automatic Identification System (AIS) Data for the Eastern Shore Islands, Nova Scotia, 2017-2018. Can. Manuscr. Rep. Fish. Aquat. Sci. 3172: vi + 39 p.

The Automatic Identification System (AIS) provides information on vessel characteristics and movements and can be used to characterize human activities, such as fishing and shipping. Here I use AIS data to characterize vessel traffic in an Area of Interest (AOI) being considered for a Marine Protected Area, in the Eastern Shore Islands, Nova Scotia. Data quality can limit the robustness of AIS-based analyses , thus, I carried out and thoroughly documented a systematic quality control protocol and I quantified error rates. The most common vessel types in the AOI were fishing vessels, cargo vessels, pleasure craft, tankers and passenger vessels, each of which used the area differently. Vessel types that are required to carry AIS (e.g. tankers and cargo vessels) had much lower error rates in vessel type than vessels for which AIS is voluntary (e.g. pleasure craft and fishing vessels).

RÉSUMÉ

Konrad, C.M. 2020. Analysis of Automatic Identification System (AIS) Data for the Eastern Shore Islands, Nova Scotia, 2017-2018. Can. Manuscr. Rep. Fish. Aquat. Sci. 3172: vi + 39 p.

Le Système d’identification automatique (SIA) fournit des renseignements sur les caractéristiques et déplacements des navires, et il peut servir à caractériser les activités anthropiques comme la pêche et la navigation. Dans le présent document, j’utilise les données du SIA pour caractériser le trafic maritime dans une zone d’intérêt à l’étude pour la désignation de zone de protection marine, à savoir les Îles de la côte Est, en Nouvelle-Écosse. Comme la qualité des données peut limiter la robustesse des analyses fondées sur le SIA, j’ai mené et dûment consigné un protocole systématique du contrôle de la qualité, en plus de quantifier les taux d’erreur. Les types de navires les plus courants dans la zone d’intérêt étaient des navires de pêche, des navires de charge, des embarcations de plaisance, des navires-citernes et des navires à passagers, chacun utilisant la zone d’une différente façon. Les types de navires pour lesquels le SIA est obligatoire (p. ex. les navires-citernes et les navires de charge) avaient des taux d’erreur beaucoup plus faibles que les types de navires pour lesquels le SIA est facultatif (p. ex. embarcations de plaisance et navires de pêche).

1.0 INTRODUCTION

1.1. CONTEXT

The process of establishing Oceans Act Marine Protected Areas (MPAs) requires Fisheries and Oceans Canada (DFO) to conduct overviews of key human uses of the area as a component of the information-gathering process. Examining vessel traffic can be an informative part of this overview, as vessel traffic reflects several aspects of human use in an area, such as fishing, shipping and recreation. The analyses of vessel traffic reported here were conducted to inform the design and management of a potential future MPA, which is proposed to be located along the eastern coast of Nova Scotia (Figure 1). This area is referred to as the Eastern Shore Islands Area of Interest (AOI).

AIS data were collected and vessel densities mapped for a broader study area and vessel statistics were calculated for the approximated area of interest, which was used as an approximation of the Eastern Shore Islands Area of Interest (Figure 1).

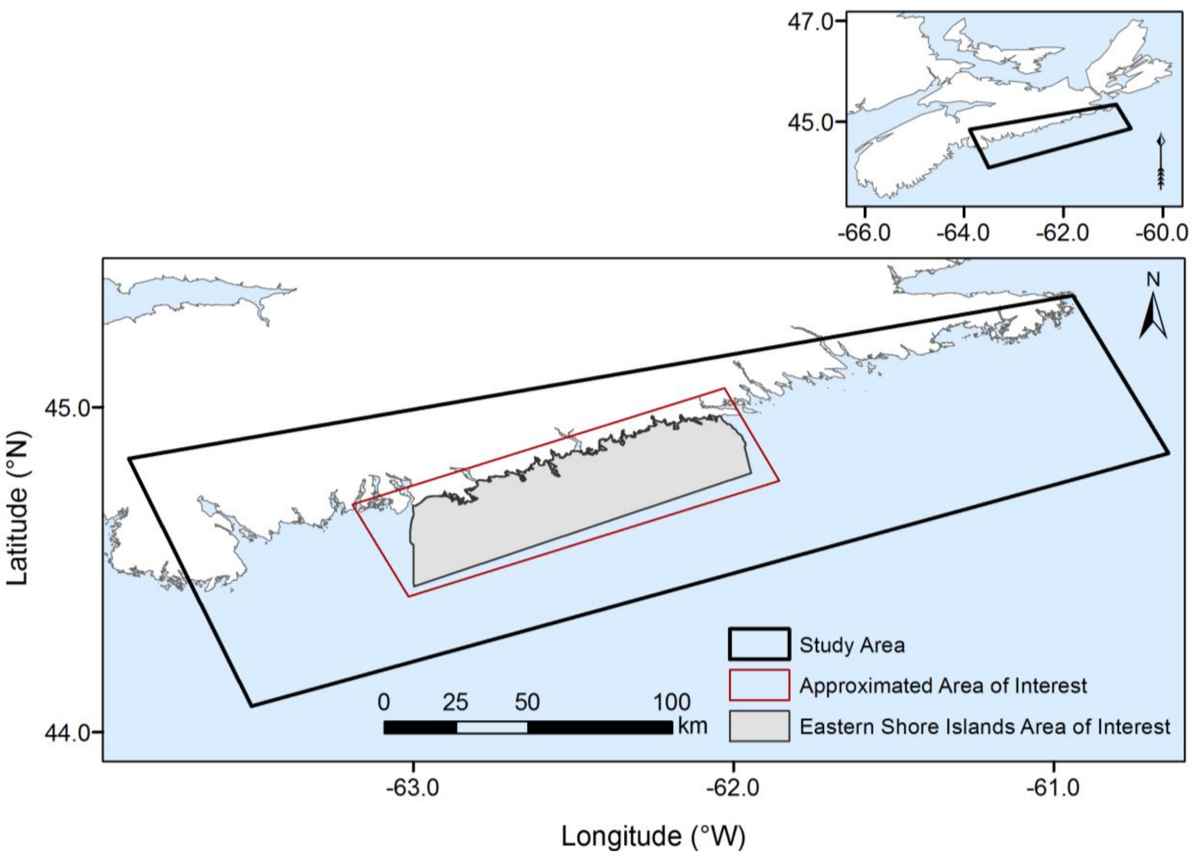

Figure 1. Areas located off of the eastern coast of Nova Scotia used for the AIS analyses. See main text for details of how each area was used in the AIS data collection and analysis. Note that the boundaries delineate the candidate area for assessment and planning purposes, and are not indicative of the boundaries for the potential future marine protected area.

1.2. AUTOMATIC IDENTIFICATION SYSTEM (AIS)

The Automatic Identification System (AIS) is an automated system for vessel tracking, with the primary aim to support safety of life at sea. AIS relies on ship-borne VHF transponders or transceivers that transmit information to receivers or transceivers on shore stations, satellites and other ships. The information transmitted is divided up between dynamic messages with details on what the vessel is doing (e.g. speed, location, and direction of travel) and static messages with details on the vessel itself (e.g. vessel name, dimension and type). Dynamic messages are transmitted as often as every two to ten seconds, while static messages are transmitted every six minutes. Information from these two messages types can be linked, because both include reference to the vessel's Maritime Mobile Service Identity (MMSI), a nine-digit vessel identifier.

As a requirement of the Convention for the Safety of Life at Sea (SOLAS; 1974), AIS is mandatory for all vessels of 300 gross tonnage (GT) or more on international voyages, and cargo ships of 500 GT or more not on international voyages, and passenger vessels irrespective of size. In Canada, the Navigation Safety Regulations (Section 65) that came into force on May 10, 2005, further specifies that AIS is mandatory for *all* vessels of 500 GT or more not on international voyages, with the exception of fishing vessels, and that fishing vessels on international voyages are exempt from the requirement to carry AIS, regardless of their size (Government of Canada 2005). These regulations also imply that the SOLAS "passenger vessel" requirement applies to ships of 150 GT or more carrying more than 12 passengers and engaged on an international voyage. In addition to the vessels for which AIS is a requirement, other vessels, such as fishing vessels and pleasure craft, may voluntarily carry AIS. Beyond its primary role for safety at sea, AIS can and has been used for a variety of purposes, including management and monitoring of human activities in important marine areas, such as MPAs (McCauley et al. 2016).

1.3. PREVIOUS WORK

Previous studies have assessed shipping traffic in areas off the east coast of Canada, using data from AIS and other systems. Using data from the Canadian Coast Guard (CCG)'s Eastern Canada Vessel Traffic Services Zone (ECAREG) system, from the year 2000, Breeze and Horsman (2005) mapped inbound commercial shipping traffic on the Scotian Shelf at a spatial resolution of four-arcminutes. The ECAREG system is mandatory for all commercial vessels over 500 gross registered tons transiting within Canada's 12-nautical-mile territorial sea. Subsequently, Koropatnick et al. (2012) used Long Range Identification and Tracking (LRIT) data from March 2010 to February 2011 to map vessel density in Atlantic Canadian waters across a broader region and at a finer resolution (two arcminutes) than the study by Breeze and Horsman (2005). LRIT is a satellite-based system and it has a broader reach than the ECAREG system in that it is mandatory for all commercial vessels subject to the International Convention for the Safety of Life at Sea (SOLAS; 1974). The LRIT dataset used by Koropatnick et al. (2012), which was provided by the Maritime Security Branch of the CCG, included positions every six hours for Canadian-flagged vessels, vessels within 1000 nautical miles (NM) of a Canadian coast, and vessels bound for a Canadian port. Using this LRIT data, Koropatnick et al. (2012) were able to track vessels in two ecologically important marine areas: the Gully MPA and the Roseway Basin Area to be Avoided.

Also mapping vessel density broadly in Atlantic Canadian waters, Simard et al. (2014) created an atlas of shipping traffic for the east coast of Canada, using AIS data that was collected by the CCG's network of terrestrial AIS receivers in 2013. AIS data has much greater temporal resolution than LRIT data; by default, AIS messages are transmitted every few seconds to every 30 seconds, allowing for high resolution of vessel movements. Combining this higher reporting rate with a finer spatial resolution (1 km by 1 km grid), Simard et al. (2014) produced their maps that provide a more detailed overview of vessel density in Atlantic Canadian waters. In addition, Simard et al. (2014) mapped vessel traffic according to month, vessel type, length and speed.

In addition to providing broad, regional overviews of vessel traffic, AIS data can also be used to assess vessel traffic in relation to specific areas of interest. For example. Vanderlaan and Taggart (2009) used AIS data to evaluate vessel-operator compliance with a voluntary area to be avoided that was aimed at reducing the risk of lethal vessel strikes to right whales.

1.4. AIS DATA RELIABILITY

AIS provides a wealth of information, but this information often contains errors or missing information, due to system issues (e.g. corruption of messages or interference between messages during transmission), human error, or deliberate misrepresentation (Harati-Mokhtari et al. 2007, McCauley et al. 2016). Simard et al. (2014) acknowledged the presence of data gaps and errors in the AIS data they used that they were not able to fully address, even with much effort directed at manual data validation. Human errors can be particularly prevalent in data fields that are manually entered, such as vessel type and dimensions (Harati-Mokhtari et al. 2007, Shelmerdine, 2015). As a result, quality control is of particular importance when these data fields are being used in an analysis. Yet, in AIS-based studies, rates of errors are not always well quantified, nor the resources used to validate data well described (e.g. Simard et al. 2014, Shelmerdine, 2015).

1.5. STUDY OBJECTIVES

The above-mentioned previous studies are informative for assessing broad vessel traffic patterns in the region, and demonstrate the utility of vessel tracking systems for assessing human use of important marine areas. Here I applied similar methods to characterize vessel traffic along the Eastern Shore of Nova Scotia, with a fine-scale resolution and particular emphasis on the Eastern Shore Islands AOI. Another objective of this study was to develop and describe a systematic protocol for thorough and efficient quality control of manually entered AIS vessel information, such as vessel identity and vessel type. Lastly, the study aimed to use the validated data to quantify rates and types of errors in manually entered AIS vessel characteristics, specifically vessel type and dimensions.

2.0 METHODS

2.1. RAW DATA SUMMARY AND PROCESSING

AIS data streamed from the CCG's terrestrial AIS receiver network, from 1 January 2017 to 11 April 2018, were used in this analysis, with 36 days[1] missing from the dataset, due to inconsistencies in the streamed data that caused incompatibilities with the decoding process. Raw AIS messages were decoded using a custom R script.

Data from AIS dynamic messages were restricted to records within the study area (Figure 2). For an unknown reason, no AIS information was received from an AIS receiving station in a key location (Ecum Secum; Figure 2) from January 1 to April 11, 2017, leading to poor coverage of the area of interest for these dates. To compensate for this period of poor coverage, AIS information for 1 January 2018 to 11 April 2018 was included, so that all times of the year could be analysed.

For each unique MMSI identified in the study area from January 2017 to December 2017, the corresponding static data were extracted for the days that the vessel was observed in the study area. A database of vessel static information was generated, containing all unique combinations of vessel identifiers from the static data, namely: MMSI, International Maritime Organization (IMO) number (a seven-digit unique vessel identifier), call sign (an alpha numeric identifier used for radio communications), and vessel name. The accompanying vessel type and vessel dimensions from the AIS static messages were also included in the database.

For custom R scripts used to restrict the AIS data to records within the study area and to generate a database of vessel static information see: https://github.com/cmkonrad/AIS-processing

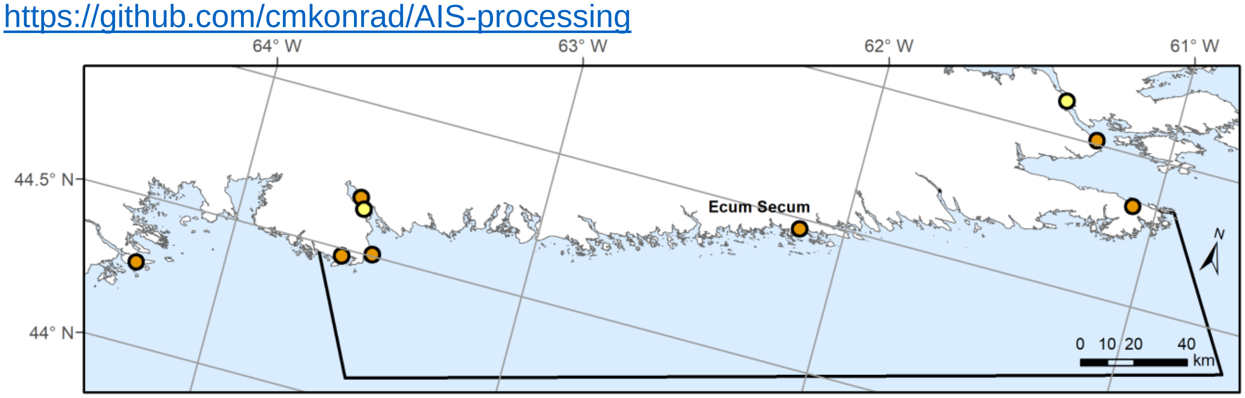

Figure 2. Automatic Identification System (AIS) terrestrial receiver station locations in and around the study area (black line). Data from Canadian Coast Guard receivers (orange points) were used in all analysis. Data from receiver stations operated by Dalhousie University (yellow points) were only used for the analysis of AIS data errors.

[1] Missing dates: 03-29, 06-07, 07-19, 08-09, 08-26, 08-27, 08-28, 08-29, 08-30, 08-31, 09-06, 09-13, 09-27, 10-21, 10-22, 10-31, 11-01, 11-11, 11-12, 11-13, 11-15, 11-17, 11-18, 11-23, 11-26, 11-27, 11-28, 11-29, 11-30, 12-01, 12-02, 12-03, 12-06, 12-08, 12-09, 12-11.

2.2. QUALITY CONTROL OF AIS DATA

2.2.1. Initial manual screening

For all data (dynamic messages and the database of vessel static information), records of “vessels” with reported MMSIs that were three digits or fewer were excluded, because, in many cases, these appeared to be stand-in MMSIs used for fishing buoys, not true vessels (to be strictly valid, MMSIs should be nine digits). Additionally, all vessel records with a reported speed over ground (SOG) of less than three knots were removed, to exclude drifting buoys or moored vessels. This SOG cut-off was selected based on visual examination of a histogram of SOG values (Figure 3), which revealed a clear valley at three knots, separating a peak at 0 knots (assumed to comprise moored ships and drifting buoys) from the rest of the distribution, which was centred at roughly 10 knots (assumed to comprise legitimate vessels underway). However, we acknowledge that this likely also removed some vessels that were actively fishing at slow speed.

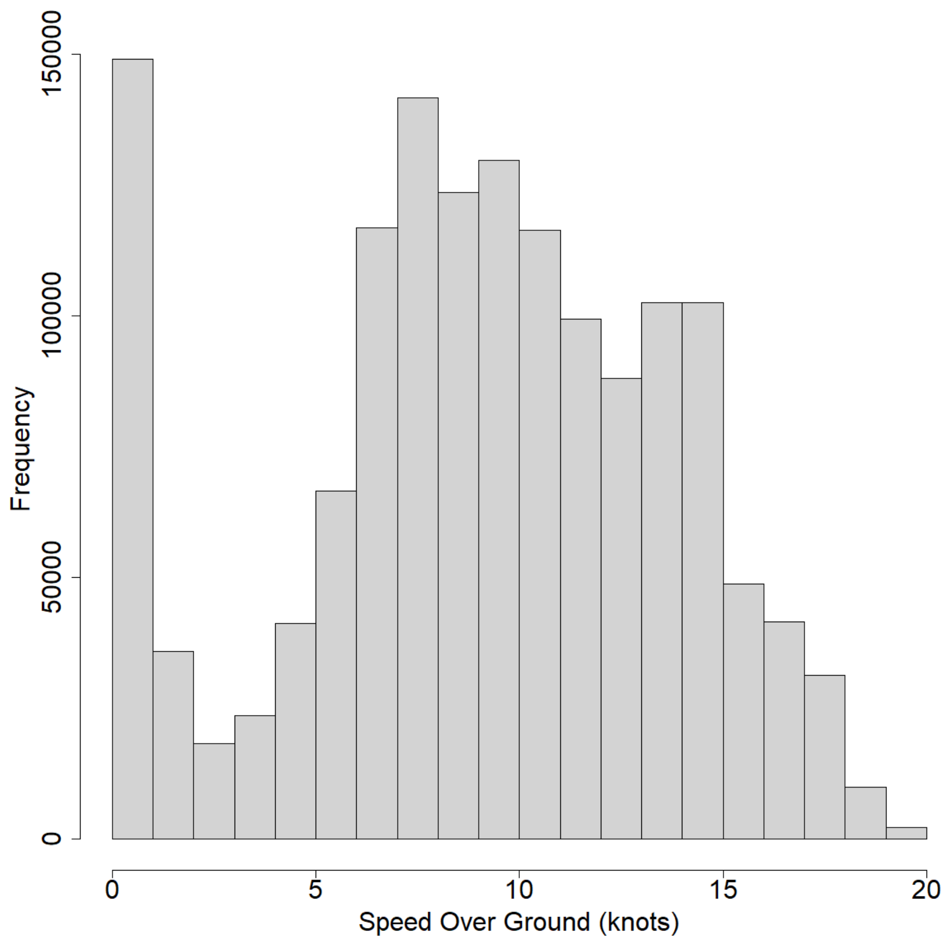

Figure 3. Frequency of speed over ground values observed in the Eastern Shore Islands Area of Interest in 2017. Speeds > 20 knots are not shown (0.64% of observations).

In the database of vessel static information, any records that could be identified, based on their name, as non-vessels were removed (e.g. bridges, buoys or maritime electronics companies). IMO numbers with two extra trailing zeros were corrected by removing the extra zeros. I then identified instances of multiple records for the same ship (e.g. rows with the same MMSI). If there were conflicts of information between the records (e.g., a value was missing from one row), I consolidated the records. If a field's value differed between the records but one value was clearly invalid (e.g. IMO < 7 digits, call sign is "CH.16") while the other was valid, the row with invalid information was removed. Duplicate records with discrepancies that could not be immediately resolved were retained for further steps of quality control.

2.2.2. Validation of vessel characteristics using online databases

For AIS data from January 2017 to December 2017, vessel type was validated by matching vessels in the database of vessel static information to records of those vessels in reliable online databases. Matches were identified based on one or more of four vessel identifiers: IMO number, MMSI, vessel name and call sign. These identifiers differ in their reliability and stability, and in how ubiquitous they are across vessel types. IMO numbers are issued by IHS Markit (previously Lloyd's Register) and are designed to remain linked to the hull for its lifetime, regardless of changes of names, flags, or owners. They are mandatory for cargo vessels of at least 300 GT and passenger vessels of at least 100 GT. Many vessels which fall outside the mandatory requirements still have IMO numbers allocated by IHS Markit, including fishing vessels and commercial yachts. In contrast, MMSIs can change, particularly when a vessel changes flag countries, because the first three digits are assigned regionally (e.g. 316 for Canada). Vessel names can also change for a given vessel, especially when ownership changes, and they are not always unique. Call signs are assigned by national licensing authorities, but leisure craft may not always be assigned call signs.

Three different resources for online validation of the database of vessel static information were used, in order of their perceived relative reliability and ease of use: 1) Maritime Portal's Sea-web database; 2) International Telecommunication Union's Ship Station list; and 3) google search (including marinetraffic.com and fleetmon.com websites). For a comparison of these resources see Appendix 1. After the validation process using these resources (described below) I ensured there were no duplicate records (i.e., records with the same MMSI).

First, all unique records were searched for matches in Maritime Portal's Sea-web Ships database (maritime.ihs.com), operated by IHS Markit, the body responsible for issuing all IMO numbers. As such, the database should contain all vessels that are registered with IMO numbers. IMO numbers were first used to search the database. If no match was found, the following vessel identifiers were used, in the given order, until a match was found: MMSI, vessel name, and call sign. Searching by vessel name included searching for matches with past names of vessels in the Sea-web Ships database. If inconsistencies existed between the vessel identifier information reported by AIS and by the Sea-web Ships database, the inconsistencies were manually inspected to assess whether the match was correct and to reconcile the information where possible (see

Appendix 2 for step-by-step procedure). Vessel records validated using Sea-web Ships were considered to have validated IMO numbers, MMSIs, call signs, vessel names, vessel types and dimensions.

For vessels with no verifiable match in the Sea-web Ship database, MMSIs were searched on the International Telecommunication Union (ITU)'s Ship Station list (www.itu.int/mmsapp). The information available on the database is that which has been notified to the ITU Radiocommunication Bureau by the Administrations of ITU Member States, and it is updated daily (see Appendix 1 for more information on this database). See Appendix 2 for the step-by-step procedure followed for verification. Vessel records validated using ITU's Ship Station list were considered to have validated MMSIs, call signs, vessel names and vessel type. This database did not provide vessel dimensions.

For vessels that remained unmatched after this search, vessel type was deduced based on vessel name if possible. This meant that vessels with "warship" in their name were assumed to be military vessels and that vessels with names beginning with "CG" were assumed to be coast guard vessel and thus classified as search and rescue vessels. Otherwise, the MMSIs for these vessels were searched on Google.ca and vessel type was recorded as verified if an image of the vessel was available (e.g. on marinetraffic.com or fleetmon.com) that allowed the vessel type to be visually confirmed. Any remaining unmatched vessels were assigned to the vessel category "Other/Unknown." For classifications of vessel types used in this analysis, see Table 1. A category for drilling rigs and ships was introduced because this vessel type was of interest but there was no AIS vessel type classification specifically for these vessel types. Vessel records validated using these remaining methods were considered to have validated vessel type only.

Using additional information on vessel sub-types from the Sea-web Ship database, I also identified any tankers of the following sub-types: liquefied natural gas (LNG) tanker, liquefied petroleum gas (LPG) tanker and combination LNG/LPG gas tanker. These sub-types were assessed because of particular interest in characterizing the use, if any, of the Eastern Shore Islands AOI by these vessels.

2.2.3. Automation of data mining

Manually searching for ship matches and exporting data can be tedious and time consuming. This process can be improved by use of web crawlers to automatically navigate websites to the desired pages and web scrapers to automatically collect data from those pages. I developed an iMacro (using software downloaded at imacros.net) to automate searching on the Sea-web Ships database for vessel matches using MMSI, vessel name and call sign. I also wrote a Python script to automate searching for vessel matches on the ITU Ship Station list for vessel matches using MMSI. For further details see Appendix 3.

Table 1. Vessel type classifications (based on AIS categories, with modification)

Code	AIS categories	Vessel type classification
30	30	Fishing (also includes fish carrier and fish farm support vessel)
33	33	Dredging or underwater ops (buoy tending, pipe burying vessel, ice breaking and research)
34	34	Diving ops
35	35	Military ops (also includes naval training ships)
37	36-37	Pleasure Craft (motorized pleasure craft and sailing vessels)
50	50	Pilot Vessel
51	51	Search and Rescue vessel
52	31, 32, 52, 57	Tug/Towing (AIS types 31, 32, 52 and 57; also includes “salvage ship”)
53	53	Port Tender (attending and off-shore supply vessels, and similar support craft)
55	55	Law Enforcement (patrol vessels and fishery patrol vessels)
60	40-49, 60-69	Passenger (passenger ferries, high-speed craft and “Passenger/Ro-Ro Ship”)
70	70-79	Cargo (also includes “heavy load carrier – semi-submersible”)
80	80-89	Tanker (also includes “FPSO, Oil”)
90	90-99	Other Type (cable layer, exhibition vessel, training ship, escort ship, launch, and any ships for which vessel type could not be verified)
100	n/a	Drilling (rigs and ships)

2.3. ANALYSIS OF AIS DATA ERRORS

The verified vessel information was used to quantify rates of errors in three vessel characteristics reported in AIS data: vessel type, length, and breadth. This analysis used all verified vessels from the study area, as well as additional vessels detected in 2017 by terrestrial AIS receiver stations operated by Dalhousie University at Halifax, NS and Mulgrave, NS (Figure 2). The stations operated by Dalhousie University saved AIS data records once per minute, unlike the CCG’s terrestrial AIS receiver network that provided continuous streamed AIS data. Vessel types verified by any of the methods described in Section 2.2.2 were used in the analysis of vessel type error rates, while only the Sea-web Ships database matches were used to verify vessel dimensions. For each vessel attribute (vessel type, length and width), I determined the proportion of records from the database of vessel static information for which the AIS data included a value for the attribute. For those with both AIS values and verified values, I determine the rates of agreement between the two data types. Vessel types were deemed to be in agreement if the code assigned for the verified vessel type captured the AIS vessel type (e.g. verified vessel type code 52 captured AIS vessel types 31, 32, 52 and 57; Table 1). For vessel dimensions, I rounded values to the nearest metre before assessing agreement of the dimensions. I also examined the distribution of the magnitude of the dimension discrepancies for vessels with differences between the values from the two data sources.

2.4. ANALYSIS OF VESSEL COUNTS

Vessel counts were determined within the approximated AOI (Figure 1). Unique vessels in the area each day were averaged within months (January 2017 to April 2018) for all vessels collectively, as well as for the five most common vessel types: fishing, cargo, pleasure craft, tanker and passenger (for months with validated vessel type data: January to December 2017). I chose to use average unique ships per day for each month, rather than monthly sums, to avoid bias resulting from differences in the number of days with available data across months. Monthly vessel counts from January to April 2018 (without any breakdown by vessel type) were compared to the corresponding vessel count values from the same months in 2017, to examine the effect of missing AIS data from the Ecum Secum station for those months. Unique vessels in the approximated AOI each day were also summed across all available days in 2017, for all vessels collectively, as well as for all identified vessel types (listed in Table 1).

2.5. ANALYSIS OF VESSEL SPEED

Within the approximated AOI, I analysed vessel speed, using the available months with verified vessel type information (January to December 2017), and excluding all vessel records with a reported SOG greater than 50 knots. These records were removed because speeds greater than 50 knots were assumed to be in error, as they were sporadic, sparse and recorded for vessel types for which they were deemed implausible. Monthly average speeds were determined by calculating daily average speeds for each vessel, and then averaging across all daily vessels averages (see below):

$$\overline{\sum_{day} \sum_{ves} \left(\overline{speed}_{ves}\right)} \quad (1)$$

This value was calculated for all vessels collectively, as well as for the five most common vessel types: fishing, cargo, pleasure craft, tanker and passenger. The minimum, maximum, sample size, and standard deviations for these speeds were also determined.

2.6. GENERATING RASTER DATA

Raster layers of vessel counts were generated using ArcGIS 10.2.2. All data layers used a WGS 1984 datum, and were projected to UTM Zone 20. The resolution at which the layers were generated was 500 m, resulting in 0.25 km^2 grid cells. Vessel count within each cell was calculated as the cumulative sum of counts of unique vessels in each day in that cell (see Appendix 4 for additional details). For the collective count of all vessels, daily counts were summed across 12 April 2017 to 11 April 2018. For the counts of the five most common vessel types (fishing, cargo, pleasure craft, tanker and passenger), daily counts were summed across 12 April 2017 to 31 December 2017[2]. Data were log2-transformed (doubling scale), to allow for improved visualization of the data.

3.0 RESULTS

3.1. VALIDATION SUMMARY AND ERROR RATES

3.1.1. Validation summary

From the AIS data used in the analysis of AIS data errors, 2181 unique vessels were identified. The majority (71.7%) of these were validated by being matched to records on the Sea-web Ships database, primarily based on IMO number (68.32% of all unique vessels; Table 2). Most remaining vessels were then matched on the ITU Ship Station list, based on MMSI (21.6% of all unique vessels; Table 2). Relatively few vessels types were assigned based on a Google search to find vessel images (3.4%) or deduced based on vessel name (1.2%; Table 2). For 2.0% of vessels, vessel type could not be verified and so these vessels were assigned to the vessel category "Other/Unknown."

Table 2. Resources used to validate AIS vessel static information. The resources were used sequentially, in the order listed below, to find matches for vessel records, such that records matched using the Sea-Web Ships database were not searched on the ITU Ship Station list, etc.

Match type		Validated matches	% of matches
Sea-Web Ships database	IMO number	1490	68.32
	MMSI	53	2.43
	current vessel name	8	0.37
	past vessel name	2	0.09
	call sign	11	0.50
ITU Ship Station list		471	21.60
Deduced from vessel name		27	1.24
Google search	marinetraffic.com	74	3.39
	fleetmon.com	1	0.05
Unconfirmed		44	2.02
Total		2181	100.00

[2] This range of dates was selected to only include dates for which vessel type had been validated, and to exclude dates with poor coverage of the AOI due to the absence of data from the Ecum Secum AIS receiver station (January 1 to April 11, 2017).

3.1.2. Vessel type reliability

Out of 2181 vessels, 474 had no vessel type listed on AIS (21.7%) and 107 had an AIS vessel type that differed from their verified vessel type (4.9%; Table 3). The rate at which AIS vessel type agreed with the verified vessel type (considering both missing and conflicting AIS vessel types) differed across vessel types (Table 3). Almost all verified cargo or tanker vessels were correspondingly designated on AIS (~99% of each), as were most verified passenger vessels and high-speed craft (79.1%). However, only a third of fishing vessels and pleasure craft were identified accordingly in the AIS data.

Many misclassified vessels (those that had an AIS vessel type that differed from their verified vessel type) were those classified as “other” on AIS that could be assigned to a specific type through the verification process. Vessels classified as “cargo” on AIS were the next most common type to be reassigned to a different type during the validation process (15.89% of misclassifications). For additional details on reassignment of misclassified AIS vessel types see Table A4 in Appendix 5.

Table 3. Agreement between verified vessel type designation and vessel type designation on Automatic Identification System (AIS). Vessels with missing AIS vessel type were considered to be in disagreement with the verified type, for the purposes of determining agreement percentage. AIS data were collected from the Canadian Coast Guard’s terrestrial AIS receiver network along the Eastern Shore of Nova Scotia and terrestrial AIS receiver stations operated by Dalhousie University at Halifax and Mulgrave.

Verified Vessel Type		Vessel Count			Type agreement (%)
		Same types	Different types	Missing AIS type	
30	Fishing	71	11	138	32.3
33	Dredging, research, etc.	2	25	4	6.5
34	Diving	0	1	1	0.0
35	Military	17	0	6	73.9
37	Pleasure craft	117	6	228	33.3
50	Pilot	6	0	1	85.7
51	Search & Rescue	18	0	13	58.1
52	Tug/Towing	34	10	11	61.8
53	Port tender	8	19	13	20.0
55	Law Enforcement	1	4	1	16.7
60	Passenger & high-speed craft	68	6	12	79.1
70	Cargo	862	3	8	98.7
80	Tanker	392	3	1	99.0
90	Other/Unknown	4	16	37	7.0
100	Drilling (Rigs and Ships)*	0	3	0	0.0
	Total	1600	107	474	73.4

**Note: This category does not exist in AIS data, and thus there was no way for the types to be the same.*

3.1.3. Vessel Dimensions reliability

The majority of vessels (72.1%) had non-zero length and width values reported in their AIS static information. Two vessels had values for width but not length in their AIS static information. I was able to verify dimensions for 71.4% of all vessels, resulting in 1511 vessels with both AIS dimensions and verified dimensions. Of these, 59.1% had agreement between AIS dimensions and verified dimensions for both length and width (to the nearest metre) and for 31.2% the verified values agreed with the AIS data for one dimension. Most differences in dimensions were less than one metre (66.4% of lengths, 70.4% of widths; Figure 4). For values in disagreement, AIS lengths were more often smaller than verified lengths (n = 311) than they were greater (n = 130), while the opposite was true for widths (smaller: 148; greater: 176).

3.2. VESSEL COUNTS IN AREA OF INTEREST

In the approximated area of interest, the top five vessel types, in decreasing order based on cumulative vessels days were fishing vessels, cargo vessels, pleasure craft, tankers and passenger vessels (Table 4). In terms of the number of unique vessels the same five vessel types were most dominant, though cargo vessels and pleasure craft ranked higher than fishing vessels, and passenger vessels ranked slightly above tankers (Table 4).

The prevalence of some vessel types in the approximated AOI varied across months, while the densities of other vessel types remained relatively constant (Figure 5). The average daily counts of passenger vessels, pleasure craft and fishing vessels each had notable seasonal variation and peaked, respectively, in October, August, and June. Tanker and cargo vessels traffic, however, was consistently present at low densities and did not vary by much.

The mean daily counts of vessels in the approximated AOI were lower in January through April of 2017 compared to the same months in 2018 (Figure 5). For January and April these differences were relatively minor (2017 values were 95% and 84% of the 2018 values), but the values for February and March were reduced by approximately 50% in 2017, relative to 2018. This is almost certainly due to the lack of information from the Ecum Secum receiving station from 1 January to 12 April, 2017. As such, daily vessel counts by vessel type for these months may be underestimates, particularly for February and March (Table 4, Figure 5). For exact values see Appendix 5, Table A2.

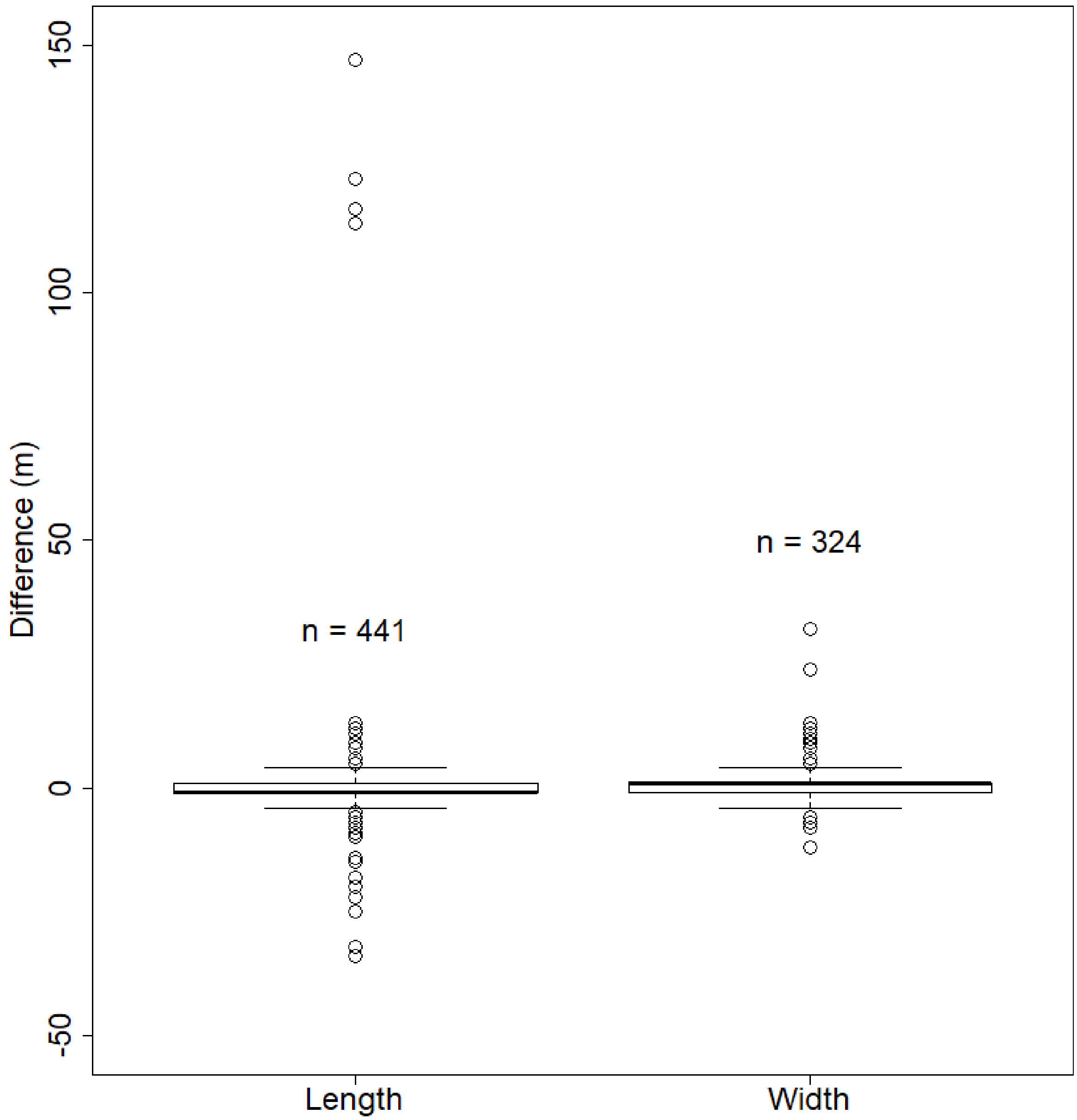

Figure 4. Deviations of vessel dimensions reported in Automatic Identification System (AIS) messages from the dimensions verified using the Sea-web online database. Only vessels with non-zero differences were included. Values were rounded to the nearest metre. Points indicate outliers beyond 1.5 times the interquartile range from the box. AIS data were collected from the Canadian Coast Guard's terrestrial AIS receiver network along the Eastern Shore of Nova Scotia and terrestrial AIS receiver stations operated by Dalhousie University at Halifax and Mulgrave.

Table 4. Vessel counts across all vessel types, for Jan-Dec, 2017, in the approximated area of interest, Eastern Shore Islands, Nova Scotia. “Unique vessels” is a count of all unique vessels identified in the area. “Cumulative Vessel-Days” is the sum of daily counts of unique vessels in the area across all available days in 2017.

Vessel Type	Unique Vessels	Cumulative Vessel-Days
Fishing	111	693
Cargo	154	536
Pleasure Craft	153	374
Tanker	37	125
Passenger	39	97
Tug & Towing	22	79
Tenders	14	57
Dredging etc.	12	56
Military	14	36
Search & Rescue	8	29
Law Enforcement	5	25
Drilling	1	2
Pilot Vessel	1	1
Other/Unknown	31	56
Total	602	2166

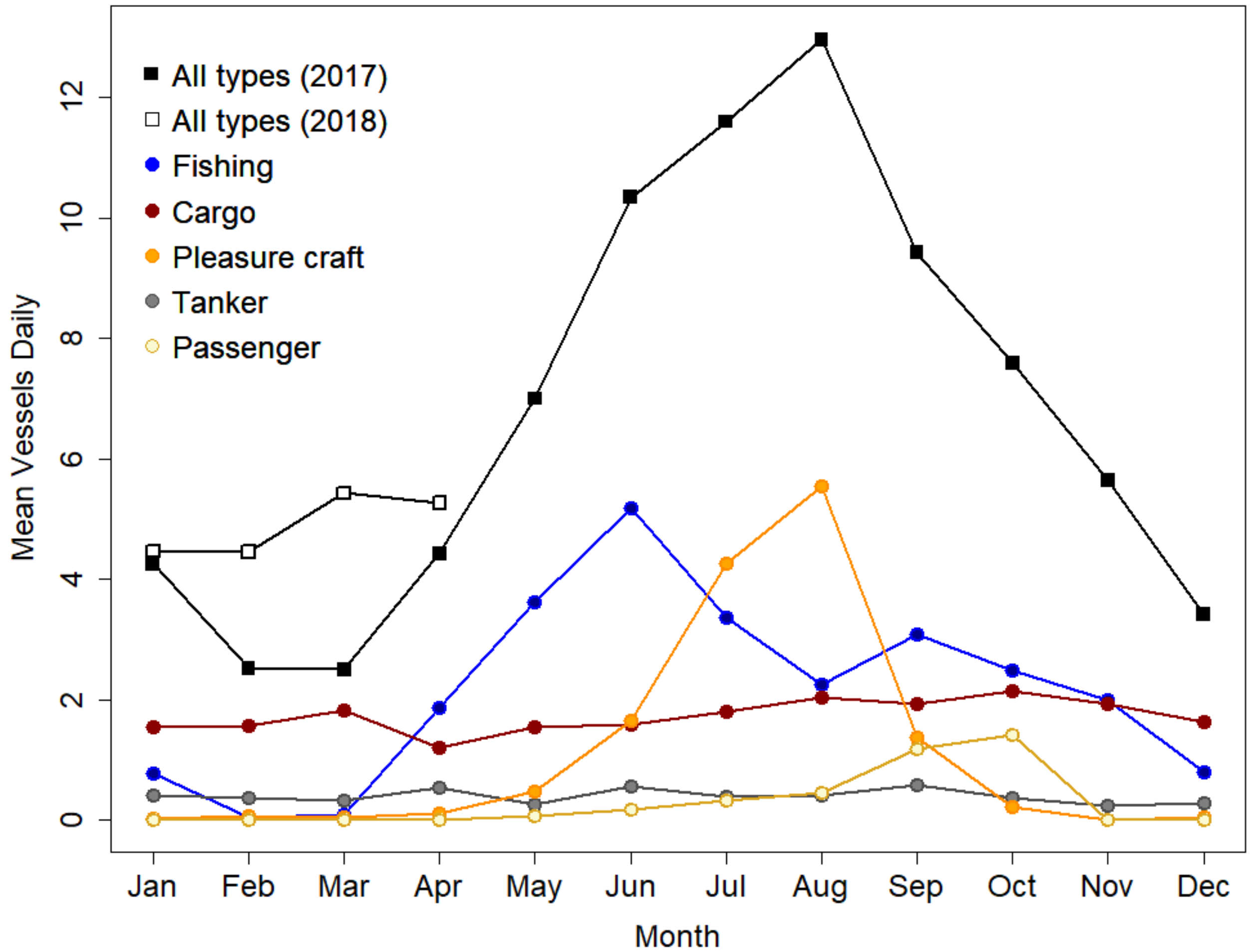

Figure 5. Monthly averages of daily vessel counts, across all vessels and for the five most prevalent vessel types, in the approximated area of interest, Eastern Shore Islands, Nova Scotia. Counts for vessel types used data from 2017 only.

3.3. NOTE ABOUT LNG AND LPG TANKERS

No LNG, LPG or combination LNP/LPG tankers were identified within the approximated AOI for the date range assessed (January through December 2017). On one occasion (14 January 2017), one LPG tanker was identified within the larger Eastern Shore area examined in this study. The closest documented approach between this vessel and the approximated AOI was 15.2 km.

3.4. VESSEL SPEED IN AREA OF INTEREST

Mean monthly vessel speeds were generally higher for cargo, tanker and passenger vessels (12 to 16 knots), and lower for fishing vessels and pleasure craft (generally ≤ 8.5 knots). The three vessel types with the largest sample sizes (fishing, cargo and tanker vessels) did not show strong seasonality in mean monthly vessel speeds (Figure 6; Table A3 in Appendix 5). Monthly means were more variable for passenger vessels and pleasure craft, but also without any seasonal trend. In the spring and summer, when slower moving vessel types (fishing vessels and pleasure craft) make up a larger proportion of all vessels (Figure 5), mean monthly vessel speed taken across all vessel types decreased (for additional details see Table A3 in Appendix 5).

The minimum and maximum daily mean speeds observed for any vessel of a given type were similar across most vessel types (Table 5). Most vessel types included a minimum daily mean speed close to the analysis cut-off of three knots, and a maximum between 21 and 24 knots. Tankers, however, had a narrower range, and passenger vessels had a higher minimum speed.

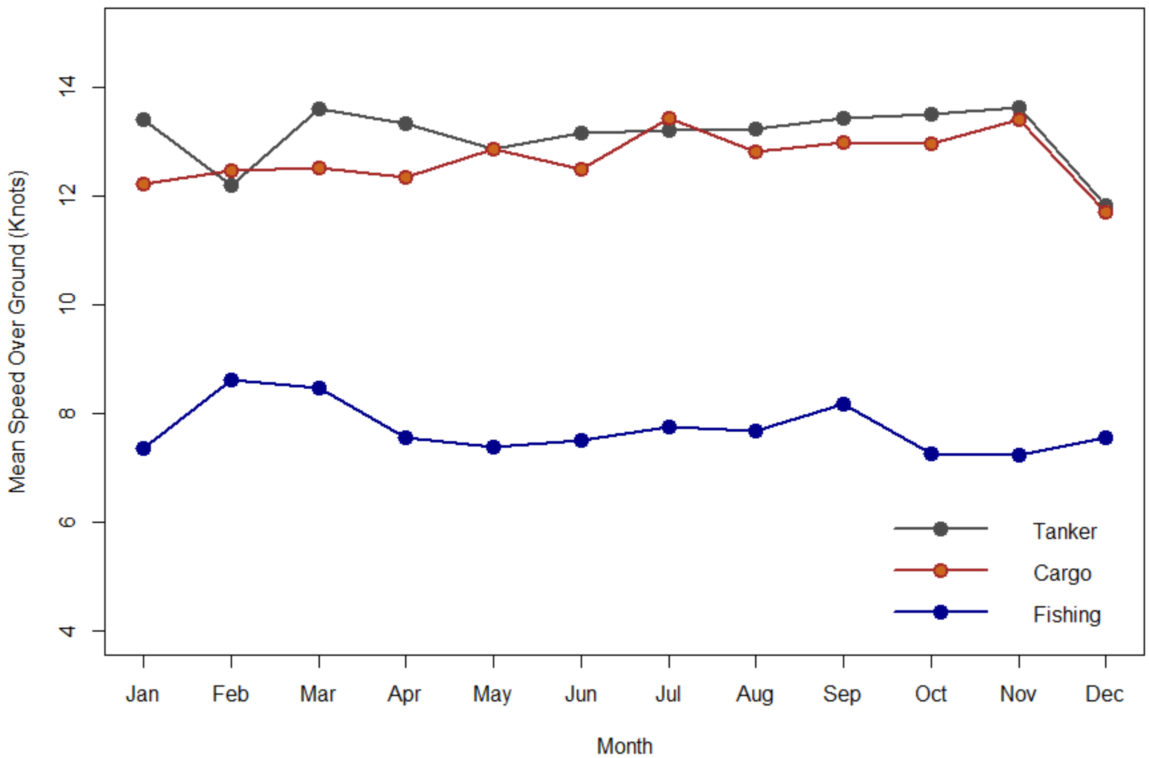

Figure 6. Monthly mean vessel speed for three common vessel types (tanker, cargo and fishing vessels) in the approximated area of interest, Eastern Shore Islands, Nova Scotia, in 2017.

Table 5. Minimum and maximum values* for an individual vessel's mean daily speed in the approximated area of interest around the Eastern Shore Islands, Nova Scotia, in 2017. The overall values ("All") encompass other vessel types in addition to the ones included in this table, e.g. tug and towing vessels.

Vessel Type	Minimum mean daily speed (knots)	Maximum mean daily speed (knots)
All	3.0	25.8
Tanker	7.8	15.5
Cargo	3.4	21.5
Passenger	4.9	23.6
Pleasure Craft	3.8	23.3
Fishing	3.2	21.8

**Note: 3 knots is the lowest possible speed, as this analysis excluded records with speeds < 3 knots. Speeds > 50 knots were also ignored.*

3.5. MAPS OF VESSEL COUNTS

Taking all vessel types together, vessel traffic was detected at some level across essentially the whole study area (Figure 7). The maps of vessel counts exclude January 1st to April 11th, 2017, because data were missing from a key AIS receiving station during this period. The area around the Port of Halifax had the highest vessel counts (Figure 7a). The vessel traffic spatial patterns for each of the five most common vessel types were distinct, in their densities across the whole study areas and in the area of interest (Figures 7 to 12). Some vessel types, namely cargo, passenger and tanker vessels, essentially remained offshore in the AOI, with cargo and passenger vessels making some trips into Sheet Harbour (Figures 8, 10 and 12). In contrast, fishing vessels and pleasure craft made more frequent use of nearshore waters among the Eastern Shore Islands (Figures 9 and 11).

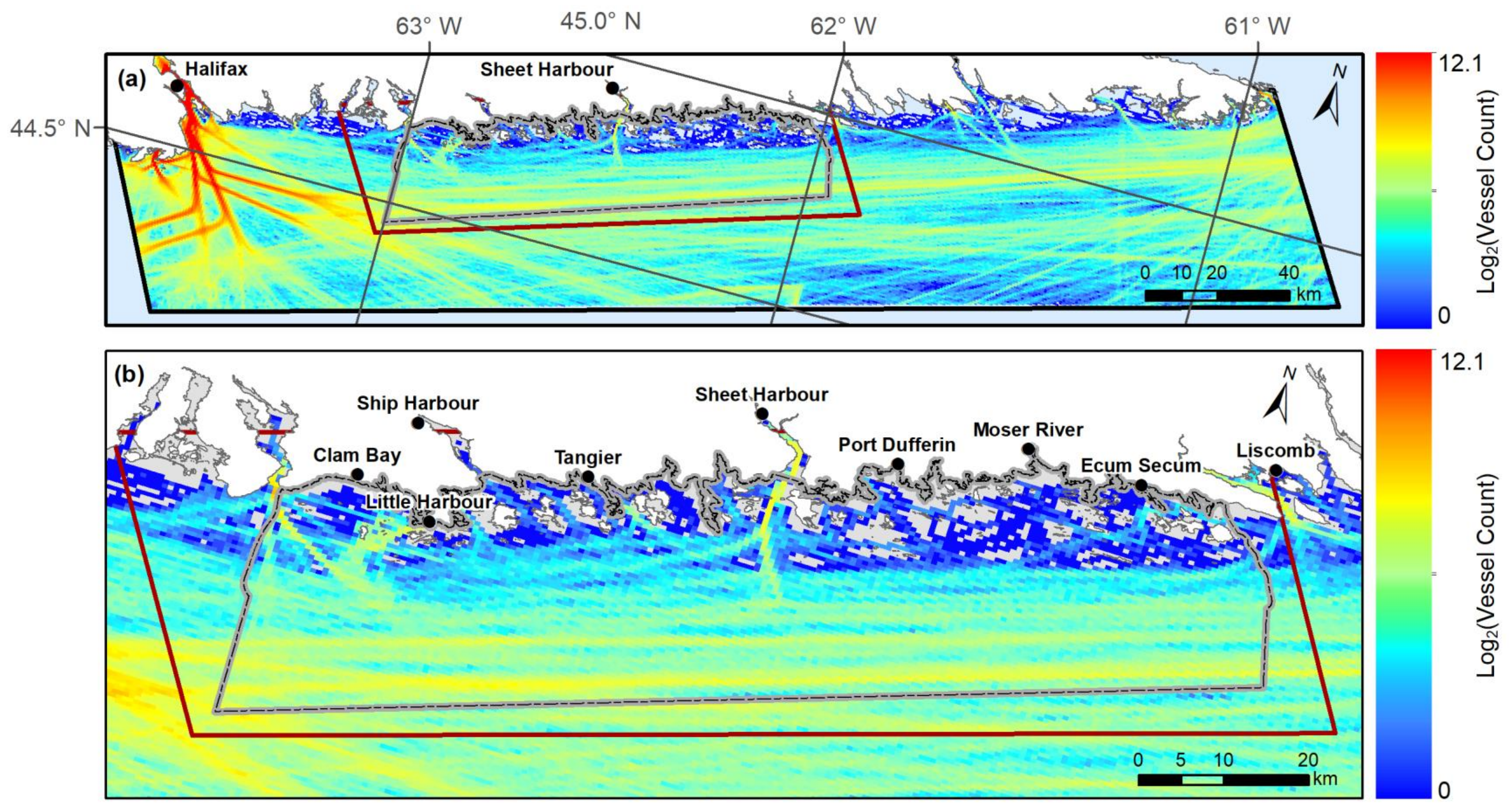

Figure 7. Cumulative vessel counts, from 12 April 2017 to 11 April 2018, for vessels of all types, as determined from Automatic Identification System (AIS) data, within the study area (black line) along the Eastern Shore of Nova Scotia (cumulative sum of counts of unique vessels in each day; see section 2.6 for additional details on how maps were generated). AIS is not mandatory for all vessels; only AIS-bearing vessels are represented here. All vessel statistics were calculated within the approximated area of interest (red line), which was used as an approximation of the Eastern Shore Islands Area of Interest (grey line with black stippling). Note that the boundaries delineate the candidate area for assessment and planning purposes, and are not indicative of the boundaries for the potential future Marine Protected Area.

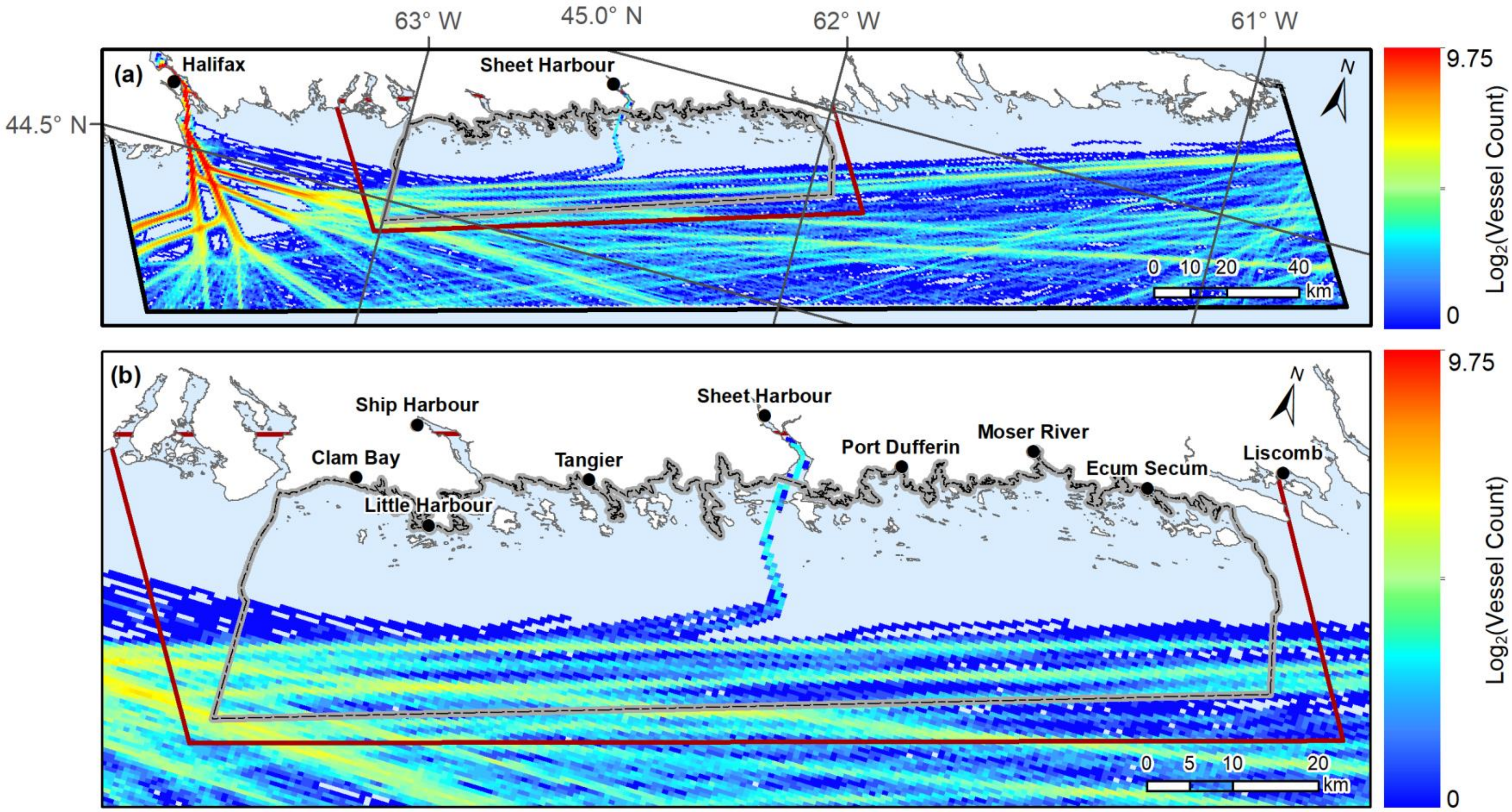

Figure 8. Cumulative cargo vessel counts, from 12 April 2017 to 31 December 2017, as determined from Automatic Identification System (AIS) data, within the study area (black line) along the Eastern Shore of Nova Scotia (cumulative sum of counts of unique vessels in each day; see section 2.6 for additional details on how maps were generated). Since AIS is mandatory for cargo vessels, essentially all cargo vessel traffic is likely represented here. All vessel statistics were calculated within the approximated area of interest (red line), which was used as an approximation of the Eastern Shore Islands Area of Interest (grey line with black stippling). Note that the boundaries delineate the candidate area for assessment and planning purposes, and are not indicative of the boundaries for the potential future Marine Protected Area.

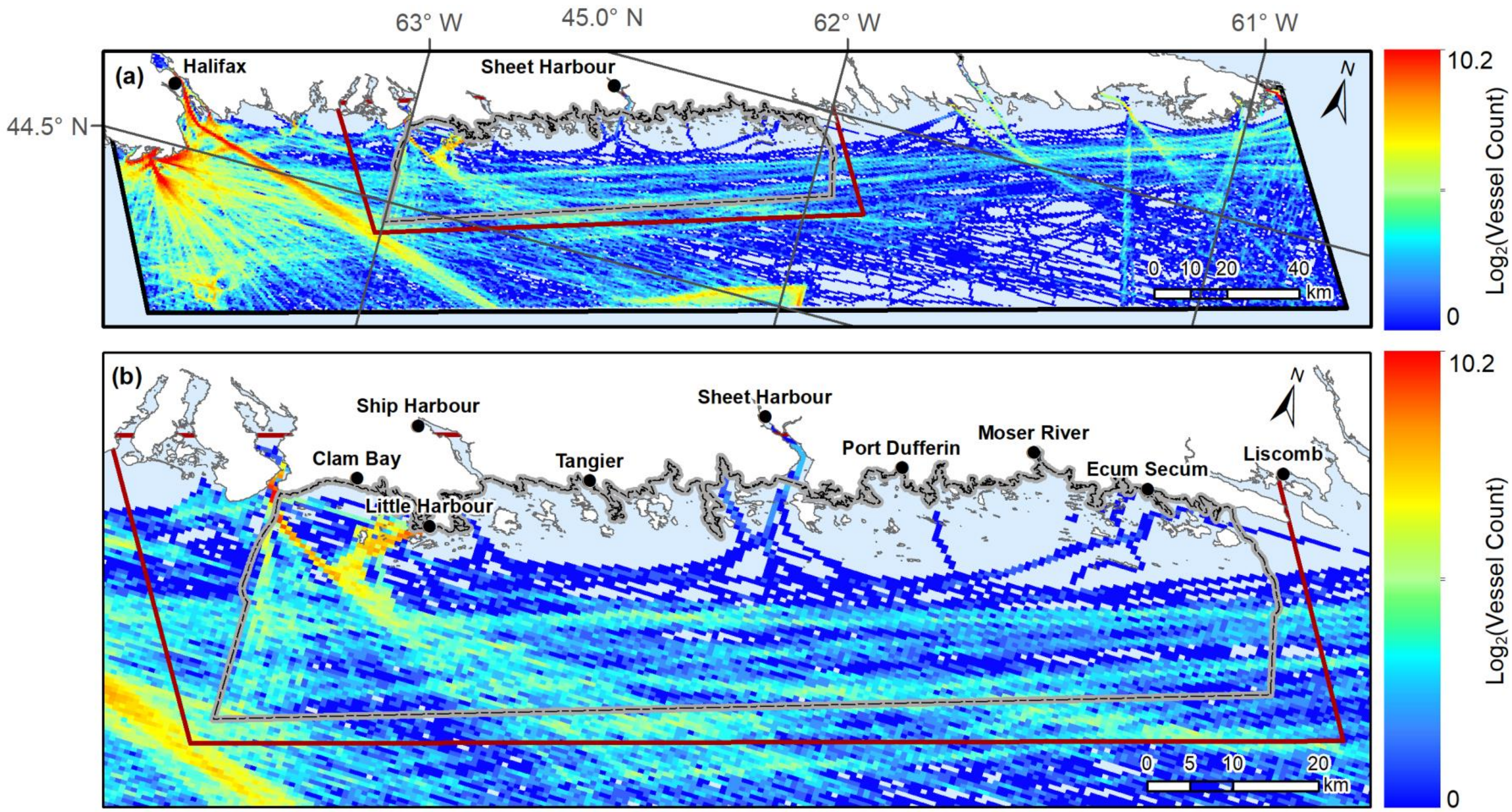

Figure 9. Cumulative fishing vessel counts, from 12 April 2017 to 31 December 2017, as determined from Automatic Identification System data (AIS), within the study area (black line) along the Eastern Shore of Nova Scotia (cumulative sum of counts of unique vessels in each day; see section 2.6 for additional details on how maps were generated). AIS is voluntary for fishing vessels; only AIS-bearing vessels are represented here. All vessel statistics were calculated within the approximated area of interest (red line), which was used as an approximation of the Eastern Shore Islands Area of Interest (grey line with black stippling). Note that the boundaries delineate the candidate area for assessment and planning purposes, and are not indicative of the boundaries for the potential future Marine Protected Area.

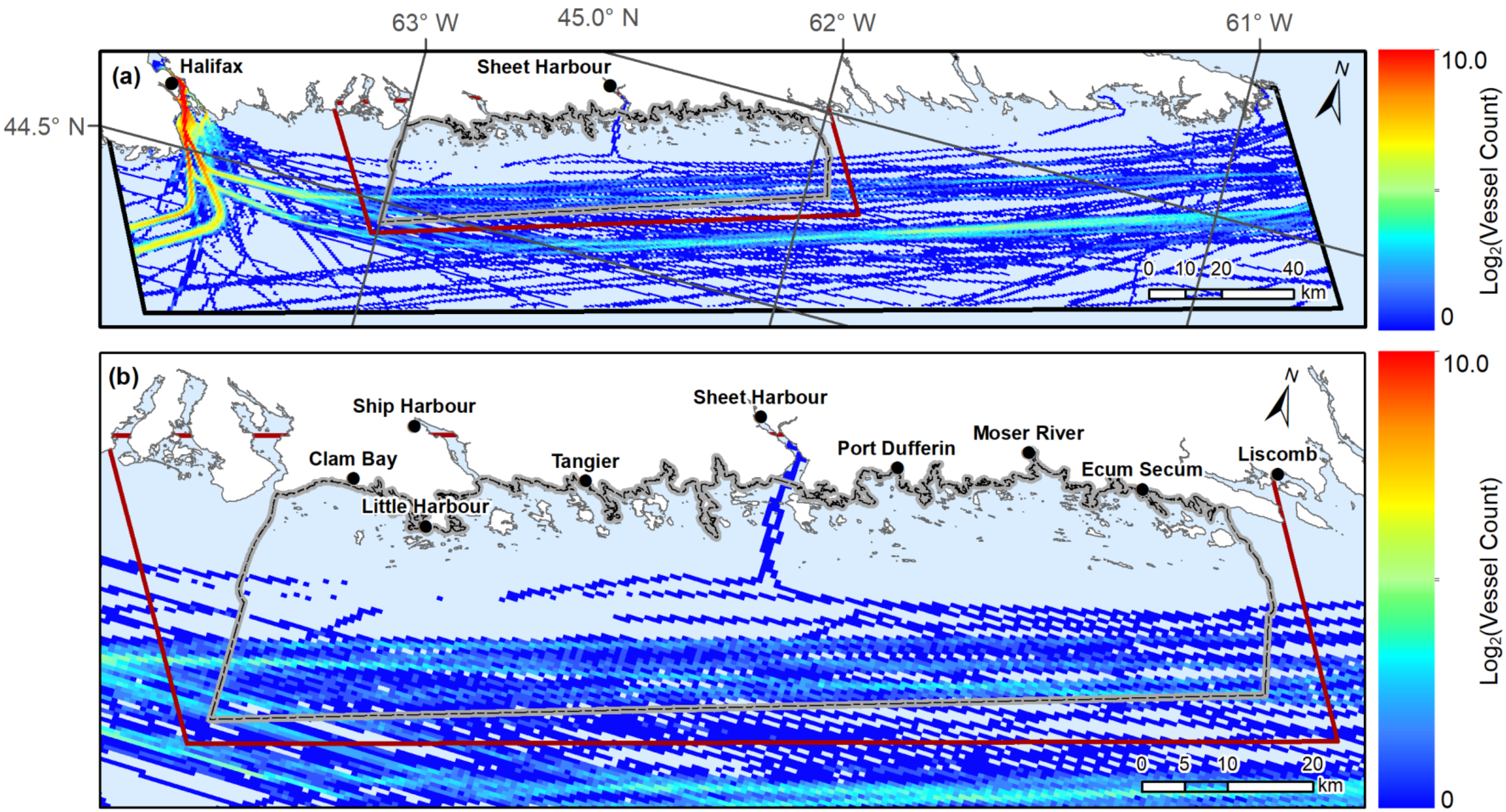

Figure 10. Cumulative passenger vessel counts, from 12 April 2017 to 31 December 2017, as determined from Automatic Identification System (AIS) data, within the study area (black line) along the Eastern Shore of Nova Scotia (cumulative sum of counts of unique vessels in each day; see section 2.6 for additional details on how maps were generated). AIS is mandatory for passenger vessels ≥ 150 GT, carrying > 12 passengers, and engaged on an international voyage; only AIS-bearing vessels are represented here. All vessel statistics were calculated within the approximated area of interest (red line), which was used as an approximation of the Eastern Shore Islands Area of Interest (grey line with black stippling). Note that the boundaries delineate the candidate area for assessment and planning purposes, and are not indicative of the boundaries for the potential future Marine Protected Area.

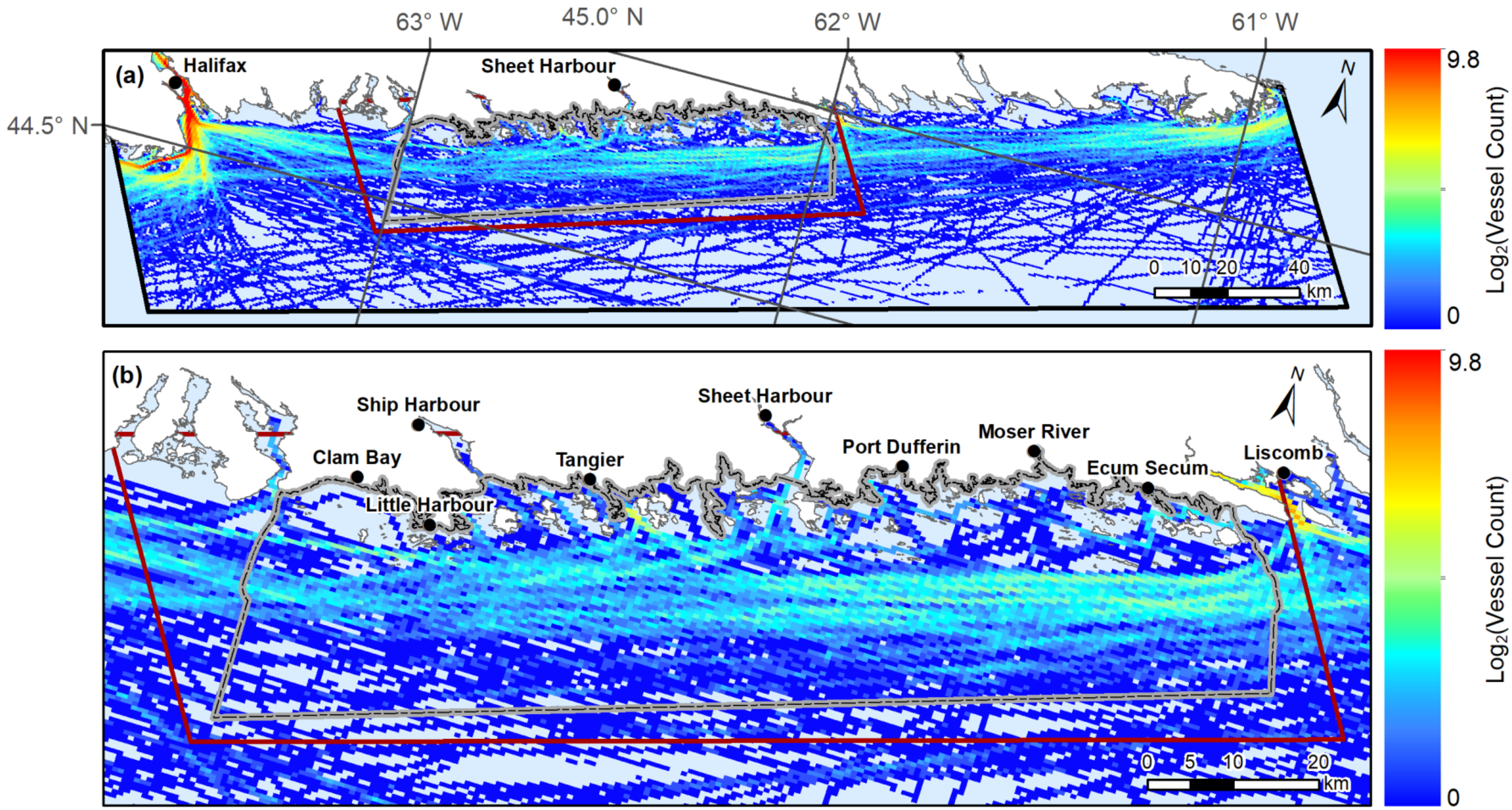

Figure 11. Cumulative pleasure craft counts, from 12 April 2017 to 31 December 2017, as determined from Automatic Identification System (AIS) data, within the study area (black line) along the Eastern Shore of Nova Scotia (cumulative sum of counts of unique vessels in each day; see section 2.6 for additional details on how maps were generated). AIS is voluntary for pleasure craft; only AIS-bearing vessels are represented here. All vessel statistics were calculated within the approximated area of interest (red line), which was used as an approximation of the Eastern Shore Islands Area of Interest (grey line with black stippling). Note that the boundaries delineate the candidate area for assessment and planning purposes, and are not indicative of the boundaries for the potential future Marine Protected Area.

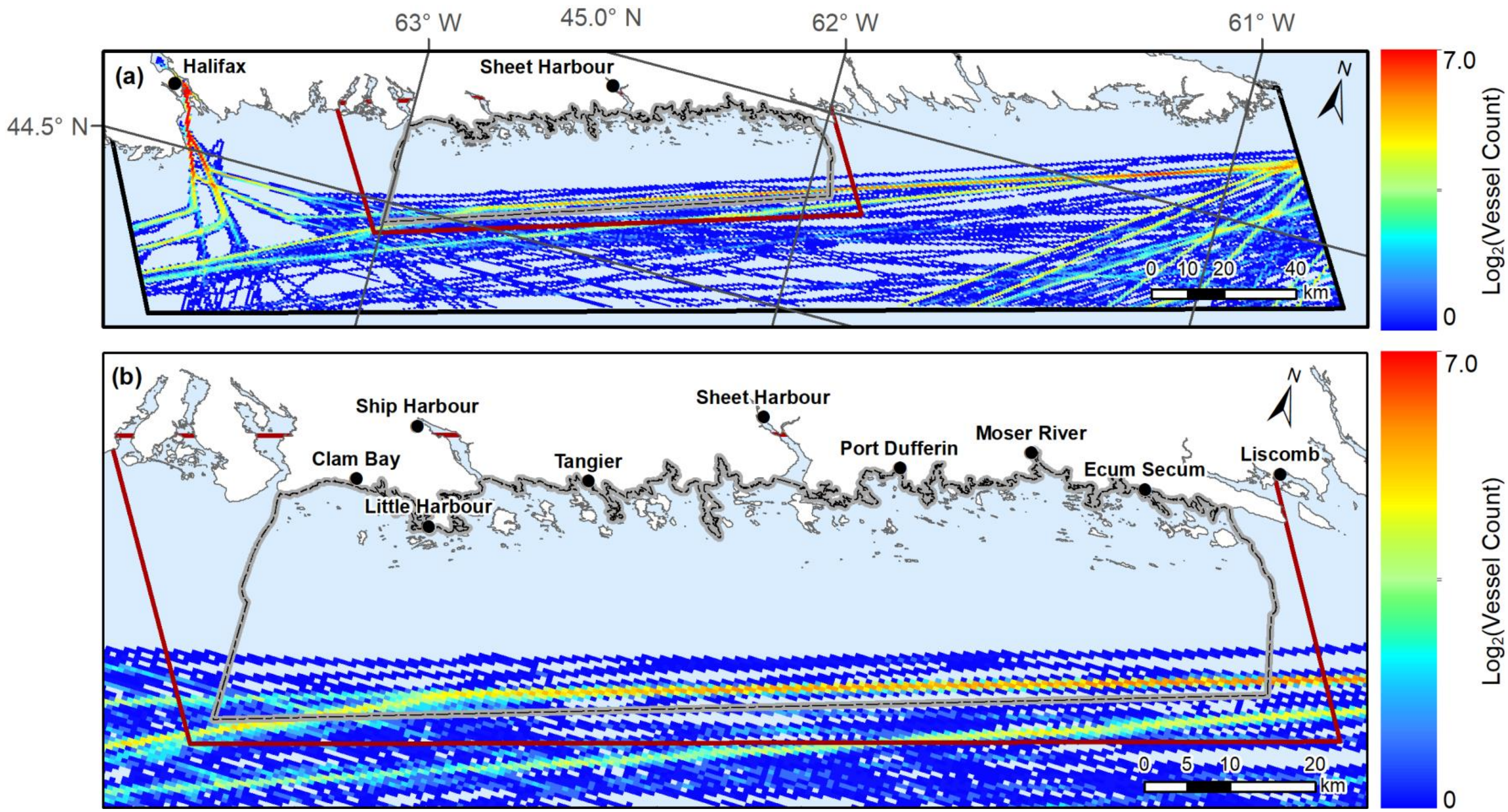

Figure 12. Cumulative tanker vessel counts, from 12 April 2017 to 31 December 2017, as determined from Automatic Identification System (AIS) data, within the study area (black line) along the Eastern Shore of Nova Scotia (cumulative sum of counts of unique vessels in each day; see section 2.6 for additional details on how maps were generated). Since AIS is mandatory for tankers, essentially all tanker vessel traffic is likely represented here. All vessel statistics were calculated within the approximated area of interest (red line), which was used as an approximation of the Eastern Shore Islands Area of Interest (grey line with black stippling). Note that the boundaries delineate the candidate area for assessment and planning purposes, and are not indicative of the boundaries for the potential future Marine Protected Area.

4.0 DISCUSSION

4.1. CHARACTERIZING VESSEL TRAFFIC ON THE EASTERN SHORE

This study was able to achieve its primary objective, by providing informative maps and summary statistics that characterize vessel traffic along the Eastern Shore of Nova Scotia at a fine-scale resolution, with an emphasis on the Eastern Shore Islands AOI. The study identified which vessel types were top users of the AOI and further examined the behaviour of these vessels in the area, including characterizing changes in vessel counts across months.

However, these results must be interpreted with the understanding that only AIS-bearing vessels were captured by the methods used. As such, the results show areas where vessels of a given type were present, but the results cannot be used to definitively exclude the possibility of vessel presence in areas where vessels were not observed. This is particularly true for vessel types where AIS in not mandatory (e.g. fishing vessels). For such vessel types, the AIS-bearing vessel traffic represents an unknown proportion of traffic of these types, which may or may not well reflect the overall use of the area by vessels of this type. For vessel types required to carry AIS, such as tankers, the results can be expected to be a good representation of vessel traffic. For such vessel types, a useful application of this information is to determine, with relatively high certainty, the presence or absence of a vessel type of concern in a given area, such as was done in this analysis to determine the absence of LNG, LPG and combination LNP/LPG tankers from the approximated AOI for the date range assessed. Even for AIS-mandatory vessels, however, station outages and other variation in coverage across space and time can affect results. For example, some passenger vessels tracks can be seen to end abruptly (Figure 10), suggesting interruptions in the reception or transmission of data. Also, AIS transmission distances are affected by the height of the receiving tower and the height of the transmitter on the vessel, because AIS transmission operates by line-of-sight. Overall, the results of this study provide an informative characterization of vessel traffic, using fine-scale and quality-controlled data, but care should be taken to ensure the results are interpreted appropriately.

4.2. AIS DATA RELIABILITY

Most vessels (78.3%) had a vessel type listed on AIS and in the majority of cases (93.7%) this listed vessel type agreed with the vessel type determined through quality control using reliable online databases (Table 3). This indicates that, overall, in the region and time period studied here, vessel types reported by AIS were relatively reliable, but that a notable proportion of vessels using AIS did not report this information. Vessels of types that are typically required to carry AIS (e.g. cargo and tanker vessels) had higher correspondence between vessel types reported in their AIS data and the vessel type verified using online databases than those for which carrying AIS is optional (e.g. fishing vessels and pleasure craft). Largely, this was due to vessel type being absent from the AIS data for the majority of validated pleasure craft and fishing vessels (65.0% and 62.7%, respectively; Table 3). This variation in AIS reliability across vessel types may be due to greater care and attention being taken when this system is a requirement, or greater knowledge of the system and vessel type coding by the personnel setting up the systems on vessels for which they are a requirement. McCauley et al. (2016) also point out that some errors in AIS data can be due to intentional misrepresentation.

Many of the deviations in vessel dimension between the AIS data and the values from the Sea-Web Ships database were negligible in size. Most differences in dimensions were less than 1 m (66.4% of lengths, 70.4% of widths) and thus can likely be attributed to rounding during the analysis. As discussed by Chang (2004), another reason for differences in dimensions may be that the values entered into the AIS could be in feet rather than metres.

4.3. APPLICATIONS OF THE VALIDATION PROTOCOL AND RELATED CONSIDERATIONS

The systematic protocol for thorough quality control of vessel identity and vessel type developed here (details in Appendix 2) was highly successful in that 98.0% of vessels were positively matched to an online resource for validation of vessel type. This validation protocol can be helpful to other studies, especially for those where a high-degree of accuracy of information on vessel type and identity is required. One example would be in the development of ocean noise models, particularly in determining vessel noise signatures by pairing recordings of vessel noise from underwater passive acoustic devices with AIS vessel locations for vessels with verified characteristics. Additionally, while the methods described here were developed using data from a terrestrial AIS receiver network they are also applicable to satellite-AIS data.

Comprehensive manual quality control can be laborious or impractical when analyses are carried out over very large areas where many ships are present, or across years if there is a high turnover in the shipping fleet, as is the case in the Bay of Fundy and on the Scotian Shelf, in Atlantic Canada (Bouwman 2016). Tools to automate the process of gathering information from online databases (such as the web scrapers described in Appendix 3) can improve efficiency. With the aid of these tools and the batch search option of the Sea-Web Ships database, finding validated matches for the majority of vessels was efficiently achieved (Table 2). However, considerable manual effort was required to match the remaining vessels (e.g. by Google searches to find images to visually validate vessel type) and to manual inspect matches that had conflicting data (see Appendix 2 for details). Further development of the methodology should reduce the manual component of this process and thus improve efficiency. However, with the currently available tools it may not be possible to eliminate this manual inspection and confirmation altogether.

To minimize the manual effort required, for projects focused on vessel types that are required to carry AIS, validation could be carried out only on data from vessels transmitting “Type A” AIS messages. These vessels are those for which AIS is mandatory, which I found required notably less manual effort to validate, because most could be easily matched to the Sea-web Ships database, using a multiple ship search based on IMO numbers. If desired, vessels transmitting “Type B” AIS messages (which would include fishing vessels and others voluntarily carrying AIS) could still be included in the analysis, with their vessel types treated as unverified or unknown.

For studies intended to characterize the general patterns of traffic of varying vessel types, without regard to movements of particular vessels, validation of a random or representative sample of vessels could be another option to balance the trade-off between data accuracy and time required for quality control. The validated sample could be used to determine rates of misclassification from one type to another, and to then adjust the observed traffic patterns according to these error rates. An important caveat is that rates of errors in vessel type can vary regionally, and may vary across time. As such, an error rate calculated for one region or study should not be assumed to be directly applicable elsewhere.

This regional variation in error rates is demonstrated by an analysis by Bouwman (2016), using satellite-AIS data from 2015 to examine reliability of AIS static data for two regions: the Scotian Shelf and the Gulf of Maine. In the Gulf of Maine, many vessels were classified in the AIS data as towing-diving-dredging, wing-in-ground[3] and future-use vessels, but were reclassified, primarily as cargo or tanker vessels, based on manual validated using online resources. In contrast, off the Scotian Shelf, these classification discrepancies were not observed. My study of the Eastern Shore Islands, which falls in the Scotian Shelf region, was consistent with this finding: misclassifications of cargo and tanker vessels were low, and no vessels were reported as wing-in-ground vessels in the AIS data. National differences in AIS guidelines may play a role in explaining these differences, particularly regarding vessels being misclassified as “wing-in-ground”. In 2012, the United States Coast Guard Navigation Center (United States Coast Guard, 2012) published a guidance document that directs AIS ship types 20-29, which are normally used to represent wing-in-ground vessels, to be used for other vessel types when operating in U.S. waters. The area of the Gulf of Maine examined in the study by Bouwman (2016) was exclusively within Canadian waters, but the proximity to U.S. waters may account for the higher prevalence of these AIS ship types.

It is also important to acknowledge that not all resources for vessel data verification are alike, due to different inclusion requirements of those resources (see Appendix 1). One key difference is in the populations of vessels contained in each. For example, the proportion of verified vessels of each vessel type that were matched on the Sea-web Ships database varied substantially, ranging from 100% of tankers to 28% of fishing vessels and 6% of sailing vessels (Appendix 1, Table A1). For comparison, the majority of fishing vessels and pleasure craft were successfully matched using the ITU Ship Station list. Without care and attention, these resource differences could produce biases in verified data. For example, if only vessels that could be verified using the Sea-web Ships database were included in an analysis, this would likely over represent the vessel

[3] “A wing-in-ground craft is defined as a vessel capable of operating completely above the surface of the water on a dynamic air cushion created by aerodynamic lift” Source: www.navcen.uscg.gov/?pageName=wig

types required to carry AIS, which are much more completely represented in the Sea-web Ships database than are vessel types such as fishing vessels and pleasure craft. For further discussion of the online data sources used in this study, see Appendix 1.

4.4. CONCLUSION

Overall, these results demonstrate the utility of AIS data for characterizing vessel traffic, using the Eastern Shore Islands AOI off Nova Scotia, as a case study. Additionally, these results highlight the importance of validating vessel information derived from AIS data, such as vessel type, to ensure accuracy of this information. The methods described and recommendations made in this study may also serve as guidelines to inform future similar work.

5.0 ACKNOWLEDGEMENTS

The author would like to acknowledge the help the following individuals in completing this report: Heather Breeze for her assistance in formatting and compiling this document, Angelia Vanderlaan for her collaboration on decoding the AIS data and her help with troubleshooting, and Diego Ibarra and Edd Clarke for their guidance on automation of the vessel verification process. The author also acknowledges the Canadian Coast Guard and Taggart Lab, Department of Oceanography, Dalhousie University for providing AIS data.

6.0 REFERENCES

Bouwman, K. 2016. An analysis of Scotia-Fundy vessel users and what this means for the North Atlantic right whale. Master of Marine Management thesis. Dalhousie University, Halifax, Nova Scotia, Canada. 68 pp.

Breeze, H. and Horsman, T., eds. 2005. The Scotian Shelf: An Atlas of Human Activities. DFO, Oceans and Coastal Management Division, Dartmouth, N.S. v + 113 p.

Chang, S. 2004. Development and analysis of AIS applications as an efficient tool for vessel traffic service. Ocean '04 - MTS/IEEE Techno-Ocean '04: Bridges across the Oceans - Conference Proceedings. 4: 2249-2253.

Harati-Mokhtari, A., Wall, A., Brooks, P. and Wang, J. 2007. Automatic identification system (AIS): data reliability and human error implications. J. Navigation 60(3): 373-389.

International Convention for the Safety of Life at Sea. 1974. Adopted 1 November 1974; Entry into force: 25 May 1980. See online: http://www.imo.org/en/About/Conventions/ListOfConventions/Pages/International-Convention-for-the-Safety-of-Life-at-Sea-(SOLAS),-1974.aspx

Koropatnick, T., Johnson, S.K., Coffen-Smout, S., Macnab, P. and Szeto, A. 2012. Development and Applications of Vessel Traffic Maps Based on Long Range Identification and Tracking (LRIT) Data in Atlantic Canada. Can. Tech. Rep. Fish. Aquat. Sci. 2966: 27 pp.

McCauley, D.J., Woods, P., Sullivan, B., Bergman, B., Jablonicky, C., Roan, A., Hirshfield, M., Boerder, K. and Worm, B. 2016. Ending hide and seek at sea. Science 351(6278): 1148–1150.

McLeod, A. 2017. AIS Whale-alert! Assessing the fleet preferences for near real-time whale conservation in the Atlantic Canada. Master of Marine Management thesis. Dalhousie University, Halifax, Nova Scotia, Canada. 98 pp.

Government of Canada. 2005. Navigation Safety Regulations. SOR/2005-134. See online: https://laws-lois.justice.gc.ca/eng/regulations/SOR-2005-134/

Shelmerdine, R.L. 2015. Teasing out the detail: How our understanding of marine AIS data can better inform industries, developments, and planning. Mar. Policy. 54: 17-25.

Simard, Y., Roy, N., Giard, S., and Yayla, M. 2014. Canadian year-round shipping traffic atlas for 2013: Volume 1, East Coast marine waters. Can. Tech. Rep. Fish. Aquat. Sci. 3091(Vol.1)E: xviii + 327 pp.

United States Coast Guard. 2012. Automatic identification system: New AIS Encoding Guidance for U.S. waters. See online: https://www.navcen.uscg.gov/pdf/AIS/AIS_Special_Notice_and_AIS_Encoding_Guide_2012.pdf

Vanderlaan, A.S.M. and Taggart C.T. 2009. Efficacy of a voluntary area to be avoided to reduce risk of lethal vessel strikes to endangered whales. Conserv. Biol. 23(6): 1467-1474.

APPENDIX 1: COMPARISON OF ONLINE RESOURCES FOR VESSEL STATIC INFORMATION VALIDATION

Please note that these comparisons reflect the state of the resources in December, 2018, and will not reflect any updates or changes to the system after this time.

MARITIME PORTAL: SEA-WEB SHIPS

Maritime Portal's Sea-web Ships database, operated by IHS Markit, is available at maritime.ihs.com. This database should contain all vessels that are registered with IMO numbers (i.e. passenger ships ≥ 100 GT, cargo ships ≥ 300 GT, and vessels that voluntarily register, such as some fishing vessels). The Sea-web Ships database allows searching for matches for many ships at once, using a list of IMO numbers. Note that while the "Upload Multiple Ships" pop-up box says that a list of MMSI numbers can also be entered, this feature is not functional (personal communication with IHS Markit, Feb 8, 2018). Instead, a web-scraper can be used to automate searching for matches for lists of vessel MMSIs, vessel names and call signs. For details see subsequent section on web-scrapers. For single ship search results the data can be exported as a text file (.txt). For multiple ship search results the data can be exported in one of several formats, including .pdf, .csv and .xlx. The desired fields to be displayed (and subsequently exported) can be specified under the "Display Fields" tab. For the purpose of this study the fields of interest that I selected were: IMO number, MMSI, Call sign, Name of Ship, Length, Breadth, Draught, Ship Type and ExName.

Pros:

- Reliably finds matches for vessels required to be registered with IMO numbers
- Provides additional vessel information (e.g. ownership, machinery, previous vessel names)
- Allows batch ship searching and data export
- Appear to be no duplicated IMO numbers or MMSIs in the database

Cons:

- Access requires a paid subscription
- Is missing many smaller vessels (e.g. pleasure craft and fishing vessels)

INTERNATIONAL TELECOMMUNICATION UNION: SHIP STATION SEARCH

The International Telecommunication Union (ITU)'s Ship Station Search is available at www.itu.int/mmsapp. In addition to vessel name, call sign, MMSI and vessel type, further information is sometimes available, such as on vessel ownership and gross tonnage. It provides information on ship stations that have been notified to the ITU Radiocommunication Bureau, which includes a broad variety of vessel types (i.e. ranging from tankers and other large commercial vessels to pleasure craft and other small vessels).

Pros:

- Finds matches for many vessels that are not registered on the Sea-web Ships database (typically pleasure craft, fishing vessels, etc.)
- Free access

Cons:

- No built-in batch ship searching and data export (but can be automated using web-scrapers)

GOOGLE SEARCH: PHOTOGRAPHIC CONFIRMATION

MMSI were searched on Google.ca to search for images of the vessel that could be used to visually verify the vessel type. The internet sites on which usable image results were viewed were marinetraffic.com and fleetmon.com.

Pros:

- Finds matches for some vessels that are not registered on the Sea-web Ships database (typically pleasure craft, fishing vessels, etc.)
- Free access

Cons:

- Images found not always adequate for confidentially assigning a vessel type
- Manual validation required, so relatively time consuming

VESSEL MATCHES ON ONLINE RESOURCES BY VESSEL TYPE

For interpreting Table A1, please note that the resources were used sequentially, in the order listed in the table (left to right), to find matches for vessel records, such that records matched using the Sea-Web Ships database were not searched on the ITU Ship Station list, etc. As such, the low percentages of certain vessel types in subsequently used resources do not necessarily indicate a lack of records on that resources of vessels of those types.

Table A1. Percentage of vessels of each vessel type matched on each online data source. AIS data were collected from the Canadian Coast Guard's terrestrial AIS receiver network along the Eastern Shore of Nova Scotia and terrestrial AIS receiver stations operated by Dalhousie University at Halifax and Mulgrave.

			Percentage of type matched on source (%)			
Vessel Type		Total Count	Sea-Web Ships	ITU Ship Station list	Deduced from vessel name	Google search
30	Fishing	220	28.18	71.4	0.0	0.5
33	Dredging, research, etc.	31	90.32	9.7	0.0	0.0
34	Diving	2	50.00	50.0	0.0	0.0
35	Military	23	30.43	4.3	60.9	4.3
36	Sailing	204	6.37	68.1	1.0	24.5
37	Pleasure craft	147	13.61	78.9	0.0	7.5
50	Pilot	7	14.29	85.7	0.0	0.0
51	Search & Rescue	31	0.00	83.9	12.9	3.2
52	Tug/Towing	55	76.36	14.5	0.0	9.1
53	Port tender	40	72.50	0.0	17.5	10.0
55	Law Enforc.	6	83.33	16.7	0.0	0.0
60	Passenger & HSC	86	89.53	9.3	0.0	1.2
70	Cargo	873	99.77	0.1	0.0	0.1
80	Tanker	396	100.00	0.0	0.0	0.0
90	Other/Unk	13	69.23	30.8	0.0	0.0
100	Drilling	3	100.00	0.0	0.0	0.0

APPENDIX 2: STEP-BY-STEP PROCEDURE FOR VERIFICATION WITH ONLINE RESOURCES[4]

MARITIME PORTAL: SEA-WEB SHIPS

Step 1: For vessels with IMO numbers provided in the AIS data, search batches of multiple ships, using IMO number

Step 2: Manually inspect and correct AIS matches where vessel identifiers (i.e. MMSI, call sign or ship name) differ between the AIS and Sea-web Ships data

- If name is essentially the same (e.g. “M/V Silver voyager” vs. “Silver voyager”) the name is considered a match, and the Sea-web Ships version of the name is used
- Likewise, for call signs with slight differences (e.g. “PCWU” vs. “P.C.W.U”, or an O instead of a 0 [zero], or “ZDNY3” vs. “ZDNY”), the call sign is considered a match, and the Sea-web Ships version of the name is used
- If a name or call sign for the AIS message was clearly invalid or generic (e.g., call sign: ABCDEF), the value is replaced with the values from the Sea-web Ships database
- If the vessel name is mismatched, examine old names in the Sea-web Ships database, and if the AIS name matches a past vessel name of the ship, it is deemed a match. Use the new name from Sea-web Ships, but retain the old name in a “previous name” column. Also, record when the ship was renamed
- If the call sign is mismatched but other identifiers indicate a match (i.e. IMO number, MMSI, and vessel name), keep call signs from Sea-web Ships in “callsign” column. Keep alternative call signs in a “othercallsign” column
- If the name matches between the AIS data and Sea-web Ships data, but the other information (namely MMSI and call sign) is missing from the Sea-web Ships data but present in the AIS data, use the call sign and MMSI from the AIS data
- If the AIS data is missing the call sign, but all other data matches, fill in the call sign from the Sea-web Ships match; and vice versa if the Sea-web Ships call sign is missing

[4] Development of this procedure was informed by the data validation process used by McLeod (2017)

McLeod, A. 2017. AIS Whale-alert! Assessing the fleet preferences for near real-time whale conservation in the Atlantic Canada. Master of Marine Management thesis. Dalhousie University, Halifax, Nova Scotia, Canada. 98 pp.

Step 3: Individually search for ships that could not be matched based on IMO numbers

- Search ships by MMSI (ignore possible matches where MMSI is the only vessel identifier or attribute that matches)
- Search ships by call sign (ignore possible matches where call sign is the only vessel identifier or attribute that matches)
    - If call sign matches a tender boat to its cruise ship, assign the Sea-web Ships vessel type as "Cruise Ship Tender Boat"
- Search ships by vessel name, including former names (ignore possible matches where name is the only vessel identifier or attribute that matches)

Step 4: Repeat Step 2.

Step 5: Manually inspect and correct AIS matches where IMO numbers do not match between the AIS and Sea-web Ships data[5]

- If the ships are deemed to be a match based on the other identifiers in the AIS data (i.e. name, MMSI, and call sign), add the IMO number from the Sea-web Ships database

Step 6: Inspect and address MMSI disparity between AIS data and Sea-web Ships data[6]

- Search on Sea-web Ships database for recent (since the beginning of the study period) changes in flag or ownership
- Retain both MMSIs along with the date of flag and/or ownership change (if known/applicable)
- If MMSIs are missing for Sea-web Ships database matches, use MMSI from the AIS data.

INTERNATIONAL TELECOMMUNICATION UNION: SHIP STATION SEARCH

Step 1: Individually search by MMSI for ship matches

- Collect data if it seems to be a possible true match based on name, call sign etc.

---

[5] In my experience, most often, the mismatching IMO value was invalid (i.e., "0" or not seven digits). Out of 47 mismatches that I inspected, only three were seemingly valid seven-digit IMO numbers that did not match.

[6] In my experience, most MMSI discrepancies corresponded with a recent change in flag country (52 out of 60) or ownership (4 out of the 8 which had not changed flag countries).

Step 2: Manually inspect and correct AIS matches where call sign or ship name differ between the AIS and ITU data

- Take applicable steps as in Step 2 of Sea-web Ships protocol

- If vessel name is mismatched, but it is deemed a match, consider the name from the ITU database to be correct, but retain the AIS name in an "other name" column

APPENDIX 3: AUTOMATION USING WEB CRAWLERS AND SCRAPERS

IMACRO FOR SEA-WEB SHIPS DATABASE

I used iMacros to semi-automate my search for vessel matches using MMSI, vessel name and call sign on the Sea-web Ships database. iMacros work by recording an action carried out online and replaying that action. Recording the action generates a script that can be manually edited to modify the action (e.g. to supply a list of MMSIs to iteratively search). iMacro extensions can be downloaded for Mozilla Firefox, Google Chrome, or Internet Explorer. There are free and paid versions available. The features available in the free version vary between browsers. I used the free version of the Internet Explorer extension because it allowed data input, and I required this feature in order to input the list of vessel identifiers to search.

Modification of the iMacro was required for each vessel identifier type (MMSI, vessel name and call sign). To use the iMacros I wrote, the user must first manually sign-in to the Sea-web Ships database to enable access, and set the desired fields to be exported under the “Display Fields” tab. The iMacro can then be used (with modification for each vessel identifier type) to automatically search for vessel identifiers of the corresponding type and download the data for the matches.

The iMacros required the addition of a “timeout_step”, when the speed of the internet connection was slow. Without this time step, when the internet connection was slow, the iMacro would attempt to download the data before the page had loaded and thus fail to successfully retrieve the data. As a result, some visual monitoring by the user was required during processing to ensure proper functioning. Further development of the iMacros could likely rework them to remove the need for visual monitoring. However, developing a script using Python may prove to be a more efficient solution.

PYTHON SCRIPT FOR ITU SHIP STATION LIST

I wrote a Python script to automate searching for vessel matches using MMSI on the ITU Ship Station list. The script uses Selenium Webdriver and a list of MMSIs to search. It generates an intermediate list of “magic numbers” that allows the script to find the webpage for each vessel match. It returns a .csv file containing MMSI, call sign, vessel name, general classification, and primary individual classification. The script can be downloaded at: https://github.com/cmkonrad/Automate-Online-Ship-Search

FURTHER CONSIDERATIONS

Python and iMacros are not the only ways to create web scrapers and crawlers. For example, the software program R could also be used to write functions for these purposes. Another consideration when using web scrapers and crawlers is that they are dependent on the structure of the website, so any updates or changes to the website may prevent proper functioning of the scripts, such that they need to be updated also. The scripts for this project were written in March 2018, based on the structure of the websites at that time.

APPENDIX 4: METHODOLOGICAL DETAILS ON GENERATING RASTER DATA

For analyses performed using ArcGIS 10.2.2, file geodatabases were used, because the monthly AIS were files too large (> 2GB) to work with as shapefiles. Figures A1 and A2 outline the models used in ArcGIS 10.2.2. to generate raster layers of monthly vessel counts. These monthly layers were summed using the “Raster Calculator” tool to generate the raster data displayed in Figures 7 through 12. The ArcGIS “Advanced” licence and the “Spatial Analysist” extension are required to run this process.

The model outlined in Figure A1 creates a grid of polygons, at a specified resolution (in this case 500 m), within a defined polygon (indicated by blue oval at start of model diagram), within which ships will be counted.

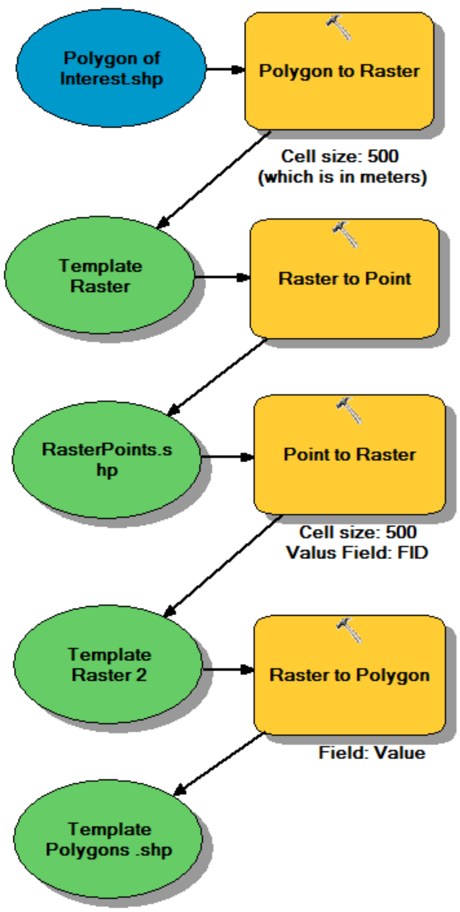

Figure A1. “CreateTemplateRaster” model built in ArcMap 10.2.2.

The model outlined in Figure A2 uses two inputs: vessel points (locations based in AIS data, with attribute values for MMSI and date) and a grid of polygons (generated by the model described in Figure A1). The model sums the daily counts of unique vessels to generate a raster layer of cumulative vessel counts. This model must be run stepwise in the “Edit…” window of the model builder, seemingly due to a software bug related to the “Add Join” tool, which prevents the model from running all the way through if used externally as a tool. An alternative to the “Add Join” tool that doesn’t cause this issue is the “Join Field” tool, however, the “Join Field” tool takes significantly longer to run. The model uses the “Near” tool rather than the “Spatial Join” tool, because the “Near” tool deals better with points that fall on the boundaries of polygons.

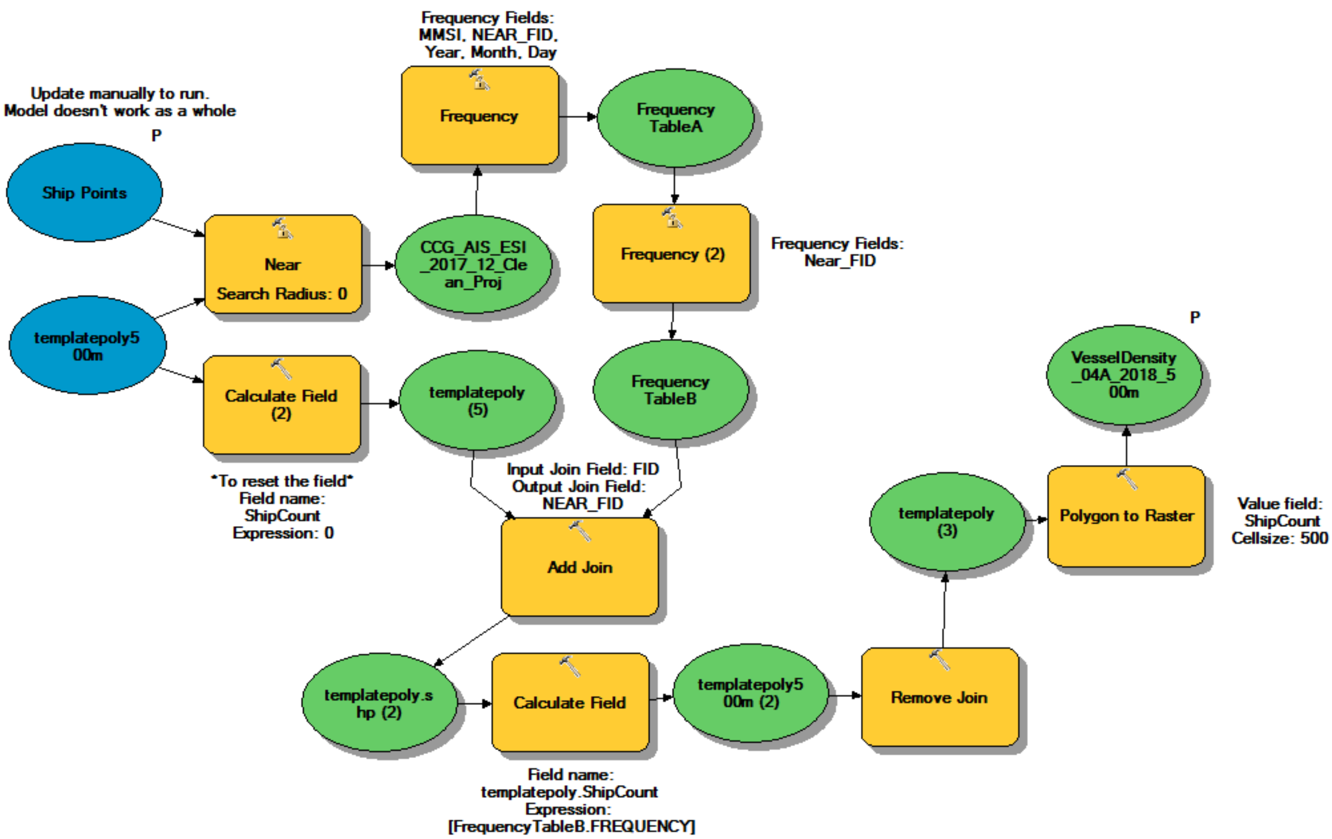

Figure A2. “MakeCountRaster” model built in ArcGIS 10.2.2.

APPENDIX 5: SUPPLEMENTARY RESULTS

Table A2. Monthly averages of daily vessel counts, across all vessels and for the top five most prevalent vessel types, in the approximated area of interest, Eastern Shore Islands, Nova Scotia.

Year	Month	All Vessels	Tanker	Cargo	Passenger	Pleasure Craft	Fishing
2017	Jan	4.26	0.41	1.56	0	0.04	0.78
	Feb	2.52	0.36	1.56	0	0.08	0.04
	Mar	2.50	0.33	1.83	0	0.04	0.08
	Apr	4.43	0.54	1.21	0	0.11	1.86
	May	7.00	0.26	1.55	0.06	0.48	3.61
	Jun	10.34	0.55	1.59	0.17	1.66	5.17
	Jul	11.60	0.40	1.80	0.33	4.27	3.37
	Aug	12.96	0.42	2.04	0.46	5.54	2.25
	Sep	9.42	0.58	1.92	1.19	1.38	3.08
	Oct	7.59	0.37	2.15	1.41	0.22	2.48
	Nov	5.65	0.24	1.94	0	0	2.00
	Dec	3.42	0.29	1.63	0	0.04	0.79
2018*	Jan	4.47					
	Feb	4.46					
	Mar	5.43					
	Apr	5.27					

**Note: Break down by vessel type only available for 2017, because vessel type data for 2018 were not validated due to constraints on time and funding for this project.*

Table A3. Monthly mean vessel speed in the approximated area of interest, Eastern Shore Islands, Nova Scotia, in 2017. Standard deviation (sd) are given, and sample sizes (cumulative unique vessels across days) are listed in parentheses. Months in which no vessels of a given type were identified are left blank.

	Mean speed ± sd (knots)					
Month	All	Tanker	Cargo	Passenger	Pleasure Craft	Fishing
Jan	10.3 ± 3.7	13.4 ± 1.4	12.2 ± 3.5		6.0	7.4 ± 0.8
	(114)	(11)	(42)		(1)	(21)
Feb	11.9 ± 3.4	12.2 ± 2.5	12.5 ± 3.5		6.7 ± 0.7	8.6
	(63)	(9)	(39)		(2)	(1)
Mar	12.3 ± 2.7	13.6 ± 1.1	12.5 ± 2.8		5.5	8.5 ± 1.1
	(60)	(8)	(44)		(1)	(2)
Apr	9.8 ± 3.0	13.3 ± 1.7	12.3 ± 2.1		5.5 ± 1.2	7.5 ± 1.0
	(124)	(15)	(34)		(3)	(52)
May	9.2 ± 3.4	12.8 ± 1.9	12.9 ± 2.7	13.4 ± 2.5	8.5 ± 3.1	7.4 ± 1.5
	(217)	(8)	(48)	(2)	(15)	(112)
Jun	8.7 ± 3.2	13.2 ± 1.4	12.5 ± 2.5	16.2 ± 2.0	6.9 ± 2.8	7.5 ± 1.4
	(300)	(16)	(46)	(5)	(48)	(150)
Jul	8.7 ± 3.4	13.2 ± 1.1	13.4 ± 3.1	11.9 ± 4.9	6.4 ± 1.4	7.7 ± 1.2
	(348)	(12)	(54)	(10)	(128)	(101)
Aug	8.6 ± 3.6	13.2 ± 1.1	12.8 ± 2.7	12.9 ± 5.9	6.9 ± 2.5	7.7 ± 0.8
	(310)	(10)	(49)	(11)	(132)	(54)
Sep	10.4 ± 3.8	13.4 ± 0.9	13.0 ± 2.9	15.3 ± 2.8	7.9 ± 3.0	8.2 ± 2.3
	(245)	(15)	(50)	(31)	(36)	(80)
Oct	11.3 ± 4.1	13.5 ± 0.8	12.9 ± 2.6	15.9 ± 3.3	13.8 ± 4.9	7.3 ± 1.2
	(205)	(10)	(58)	(38)	(6)	(67)
Nov	10.3 ± 4.0	13.6 ± 1.3	13.4 ± 3.7			7.2 ± 0.9
	(96)	(4)	(33)			(34)
Dec	10.0 ± 3.6	11.8 ± 2.6	11.7 ± 3.9		6.8	7.5 ± 1.2
	(82)	(7)	(39)		(1)	(19)

Table A4. Assignment of vessels along the Eastern Shore of Nova Scotia, in 2017, that had vessel types supplied in the Automatic Identification System (AIS) data that did not match those determined by the verification process. AIS data were collected from the Canadian Coast Guard's terrestrial AIS receiver network and terrestrial AIS receiver stations operated by Dalhousie University at Halifax and Mulgrave.

	Verified Ship Type											
AIS Type	30	33	34	37	52	53	55	60	70	80	90	100
6	0	0	0	0	0	0	0	1	0	0	0	0
9	1	0	0	1	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	1	0	1
20	0	0	0	0	0	1	0	1	0	0	0	0
30	0	1	0	0	0	0	0	0	0	0	1	0
33	0	0	0	0	0	0	0	0	0	0	2	0
35	0	1	0	0	2	0	0	0	0	0	0	0
37	3	0	0	0	1	0	0	1	0	0	4	0
39	0	0	0	1	2	0	0	0	0	0	0	0
50	1	1	0	0	0	0	0	0	0	0	2	0
51	0	5	0	0	0	0	3	0	0	0	1	0
52	1	1	0	0	0	5	0	0	0	0	0	0
53	0	0	0	0	0	0	0	0	0	0	2	0
60	1	1	0	1	0	0	0	0	0	0	2	0
70	1	3	0	1	0	7	0	3	0	0	2	0
80	0	0	0	0	1	0	0	0	1	0	0	0
90	3	12	1	2	4	6	1	0	2	2	0	2