UNIVERSITY OF CALIFORNIA
Santa Barbara

Assessing ship movements using volunteered geographic information

A Thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Arts

in

Ecology, Evolution, and Marine Biology

by

Shaun Walbridge

Committee in Charge:

Professor Steven D. Gaines, Chair

Research Professor Robert R. Warner

Assistant Professor Krzysztof W. Janowicz

March 2013

The Thesis of
Shaun Walbridge is approved:

Research Professor Robert R. Warner

Assistant Professor Krzysztof W. Janowicz

Professor Steven D. Gaines, Committee Chairperson

December 2012

Assessing ship movements using volunteered geographic information

by

Shaun Walbridge

Acknowledgements

No work is the act of one alone, and this thesis is no exception. This masters project would not have been possible without the inspiration and guidance of many. I owe the foundational ideas of the project to Ben Halpen. Without his advice and encouragement, I may have never worked on interesting marine problems, or even thought twice about shipping. Steven Gaines and the entire Gaines lab have been tremendously supportive even when it required rose-colored glasses to see the connections between our work. Robert Warner was gracious enough to join my committee on short notice, and provide insightful context. Krzysztof Janowicz kindly connected this work with the geographic world. The National Center for Ecological Analysis and Synthesis (NCEAS) provided me with the intellectual and technical means to carry out this work, including generous use of its computing resources. NCEAS staff Mark Schildhauer and Jim Regetz helped me see new ways of marrying technology to science. Feedback from NCEAS postdoctoral researchers, including Jennifer Balch, Jai Rangnathan, Carrie Kappel, and Darren Hardy have left this work richer. NCEAS staff member Ben Best has been a wonderful resource for understanding the ecological context of shipping. Megan McKenna and Steve Katz were generous enough to share with me an application area, and taught me how to intepret ship observations. The manuscript is improved thanks to the thoughtful and careful feedback of Oliver Soong and Jamie Walbridge. I am indebted to Dawn Wright, for seeing me through the fire, and supporting me through the transition between academia and industry. Thanks to my family and friends, for being there, even when they didn't understand what I was working on, or why anyone would care. To Charlie, who has been a great morale booster on many late nights. And lastly, thanks to my fiancée, Jessica Phillips, who has helped me in too many ways to list.

> *If you want to build a ship, don't drum up the men to gather wood, divide the work and give orders. Instead, teach them to yearn for the vast and endless sea.*
>
> —Antoine de Saint-Exupéry

Abstract

Assessing ship movements using volunteered geographic information

Shaun Walbridge

Shipping, the ocean transportation of people and goods, moves most world trade, and understanding its effects is required to assess human use of the oceans. This work examines the shipping trade by combining global observations of ship location with vessel identification records, and interpreting the results in an ecological context. By incorporating quality checking methods with volunteered geographic information, I provide a spatially resolved high resolution dataset which links individual vessels with their movement patterns and attributes. This contributes knowledge on the state and distribution of shipping, and identifies areas where mitigation of impacts are achievable.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The ocean is integral to supporting life, by producing much of the planet's biomass, acting as a critical step in the hydrologic cycle, and by regulating global climate. Oceans also support human needs by providing nutrition, resources, and an efficient medium for transportation. However, a confluence of changes is compromising the ocean's ability to support both marine life and meet human needs, with a recent study identifying climate change, fishing, and shipping as the greatest anthropogenic threats to marine health (Halpern *et al.*, 2008). Shipping, the ocean transportation of people and goods, moves \$1.8 trillion dollars of trade annually, constitutes a \$300 billion dollar maritime industry, and plays a crucial role in the global transportation network, where it conveys 90 % of world goods (Organisation for Economic Co-operation and Development, 2010; Rodrigue *et al.*, 2009). Preservation of the ocean's benefits calls for holistic management (Lubchenco and Petes, 2010), which in turn requires detailed quantitative information. Current efforts have emphasized broad abiotic factors such as climate, and biotic factors such as fishing, but little work has examined the role of shipping, despite its importance to both the economy and the environment. This manuscript seeks to bridge this gap, by expanding our knowledge of shipping to address issues of marine conservation, marine spatial planning, transportation geography, and navigation.

Shipping produces many ecological impacts, but the effects largely remain out of sight, and little research looks beyond local scales. Ecological effects which have been examined include groundings, which cause direct effects like oil spills and habitat destruction, such as the $3 billion Exxon-Valdez spill[1]. Shipping contributes to climate change, where greenhouse gas emissions of the industry accounts for five percent of total man-made emissions (Eyring *et al.*, 2009), and its pollution causes 60,000 premature deaths per year (Corbett *et al.*, 2007). Ship strikes also cause injury and death to marine mammals, and have driven whale population declines, for example the North Atlantic right whale (*Eubalaena glacialis*), where human activity accounts for half of all fatalities, primarily caused by ship strikes and gear entanglement (Moore *et al.*, 2007). Globally, shipping has increased three-fold in the last 50 years, coinciding with increases in average vessel speed (Vanderlaan *et al.*, 2009).

[1] The $5 billion liability for the spill was so great that a now infamous financial instrument was manufactured to absorb it: the credit default swap.

This combination has led led to large increases in shipping-induced marine mammal fatalities, despite efforts to track ship-whale interactions (Jensen *et al.*, 2004) and improvements to shipping lanes (Lagueux *et al.*, 2011; McKenna *et al.*, 2012).

Ship ballast water, used to control vessel stability, also transports invasive species long distances, leading to widespread economic and biological losses (Ruiz *et al.*, 2000; Rodrigue *et al.*, 2009). Transported species include plants, animals, bacteria, and human pathogens, resulting in major changes to many nearshore systems, particularly coastal lagoons and inlets (Leppäkoski *et al.*, 2002). Sound, which is efficiently transmitted by water, is used by many-marine species who have adapted sensitive hearing to forage and communicate. Engine noise from ships has led to extensive noise pollution, causing behavioral changes in a variety of marine species, and the loud sounds of naval activities can cause mortalities (Hatch and Fristrup, 2009). With an improved understanding of shipping, further ecological and human impacts might also be identified.

Little research has evaluated shipping as a global, interconnected system, where vessels move between a complex network of ports spanning the globe. The first work to examine the global shipping trade in an ecological context was Halpern *et al.* (2008), a synthesis effort to collect information across habitat types and impacts. The work evaluates many human uses of the ocean, and gathers together global data to examine the cumulative ocean impact of humans, ranging from changes in sea surface temperature due to radiative forcing, to land-based nutrient runoff. The data used in this analysis to represent shipping, however, has limitations: It ignores vessel type, and contains information on only 12% of the fleet. The ships it does include are also a spatially- and statistically-biased sample of the population (Wang *et al.*, 2007), and the data vary between vessels, leaving many areas with known use missing from the model. Finally, the paper does not tackle modeling issues, but instead provides patterns without the context necessary to address deeper questions.

Another paper examining global shipping, Kaluza *et al.* (2010) instead focuses on the networks contained within shipping, providing an abstract network model of ship movement. Licensing data from Lloyd's Register Fairplay, the authors aggregate port of call records, which are sequential lists of vessel location. They then extrapolate port-to-port links into routes, by combining land-based barriers with great circle distance calculations to select the path of a ship between two ports. This novel approach provides a measure of connectivity between port locations, and improves on the gravity model widely used to predict connectivity. However, their work provides insufficient detail to extract true ship movement paths, a necessary step for effective spatial management. Instead, geographically referenced facts about movement are necessary. The data used also requires expensive licensing, making any extension of the work both financially prohibitive and limiting its accessibility. Despite largely ignoring geography, Kaluza *et al.* (2010) is currently the best examination of the global shipping trade.

As we try to understand the health of the ocean (Halpern *et al.*, 2012), we require quantitative data on the marine environment and its human use, both of which pose acute challenges to ocean scientists, due to our limited means to capture information in both the open ocean and at depth (Wright and Goodchild, 1997). These data are necessary to inform decision-making, and are also important to minimizing the regulatory burden on the shipping industry. By managing shipping as a complex system, efficiency can be tied to improved environmental conditions. This work contributes foundational knowledge on the state and distribution of shipping, which can help understand this key user of the ocean, and identify key areas to examine for mitigating ecological impact.

This is the first work to combine extensive global ship observations (Freeman, 2010; Tetreault, 2002) with vessel identification records to both address questions on ecological effects of shipping, and to expand our understanding of the global shiping trade. Most previous works focused on regional problems, and the limited global work available used poor data (Corbett *et al.*, 2007; Halpern *et al.*, 2008). Incorporating geographic quality checking methods (Goodchild and Li, 2012) with volunteered geographic information (VGI, Goodchild, 2007), I provide a spatially resolved high resolution dataset which links individual ships with their movement patterns and vessel attributes. This dataset, when aggregated into multiple representations, allows us to answer open questions on topics including noise pollution, ship strikes, and invasive species introduction.

Chapter 2

Methods

2.1 Volunteered Vessel Information

Historically, ship data has been collected both by governments for internal use, and by private corporations for commercial use. As elsewhere in the production of geographic facts, a shift is underway which moves emphasis away from top-down collection, and toward volunteered information (Goodchild, 2007; Elwood *et al.*, 2011). This shift has the potential to change our understanding of the ocean, by providing inexpensive and ubiquitous data.

Ship captains have long taken climate observations alongside known locations (Brohan *et al.*, 2009). Building on this history, the Voluntary Observing Ship program (VOS, Fletcher, 2010) collects a dataset spanning over 20 years and covering 10–20% of commercial traffic within each year. As the intention of VOS is to collect open-ocean climate data, many participating ships remain anonymous, making reconstruction of ship movements difficult. The observations are contained within the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, Woodruff *et al.*, 2010), and though the data are both spatially- and statistically- biased (Wang *et al.*, 2007), VOS data serves as a useful training dataset on ship movements in the open ocean. Here, I use VOS records from 2003–2011, supplemented with opportunistic observations collected by the vessel tracking site SailWX (Mueller, 2011). Many vessels report both attributes and locations at regular intervals, filling a gap in open ocean observations.

The Automatic Identification System (AIS, IALA AISM; Tetreault, 2002) was developed to improve maritime safety by providing mariners local situational awareness (Figure 2.1). By locating the ships via global positioning satellite (GPS), and broadcasting vessel location and attributes (Table B.1) via VHF transceiver, mariners gain real-time vessel traffic, invaluable to prevent ship collisions, groundings, and during rescue operations (International Telecommunications Union, 2010). The International Maritime Organization (IMO), a branch of the United Nations which enforces rules and standards, mandates that all ships $\geq$150 gross tonnage, and ships bearing passengers, must carry AIS units (International Maritime Organization, 1980). This has led to approximately 200,000 ships being outfitted with AIS equipment, including all licensed tankers, cargo ships, and passenger vessels. AIS uses simple to operate VHF radios, broadcasting to about 40 kilometers (km) between ships. Since the system's implementation, land-based users, including ports, maritime professionals, and amateurs, have set up VHF antennae, providing low-cost local ship traffic at a range of up to 100km. Numerous sites, such as MarineTraffic (Univeristy of the Aegean, 2010) and SailWX, provide real-time feeds by aggregating the records from both land-based antennae and satellites, then sharing them using both web maps and Google Earth. These sites augment their networks by providing incentives to users willing to set up new AIS stations, improving coverage over time.

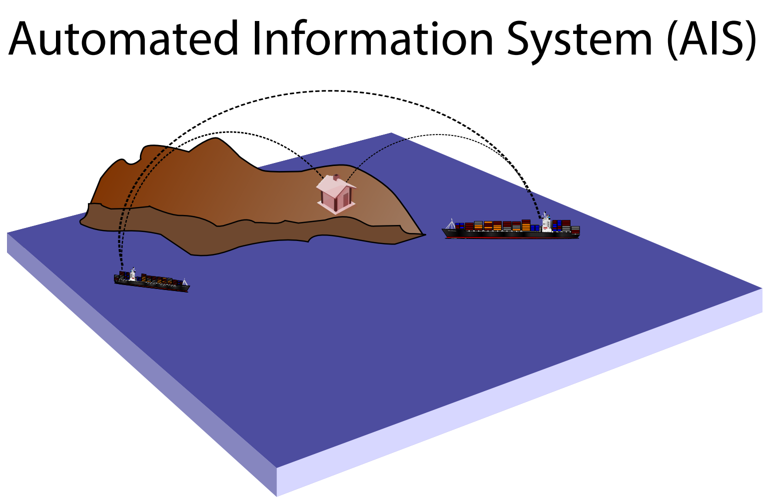

Figure 2.1: Automated Information System diagram. Ships communicate to one another via VHF broadcast. The same signals can be captured and stored by land- and satellite- based observers.

2.2 Data Collection

For this study, fifteen months of AIS observations (November 2010–December 2011) were collected, aggregating records from three online AIS providers: FleetMon, VesselTracker, and MarineTraffic. All three share Keyhole Markup Language (KML, Open Geospatial Consortium Inc.) files, intended for use within Google Earth. Examining data availability showed these providers differed in coverage area (Figure D.2). At ten-minute intervals over the study period, I automated downloading these KML files of real-time ship traffic, parsed the files to extract each observation within the KML document, normalized differences between the AIS providers, and finally inserted the results into a spatial database, (PostGIS, Ramsey *et al.*), an extension that provides support for OGC simple features (Open Geospatial Consortium Inc., 2010) on top of the PostgreSQL (PostgreSQL Global Development Group, 2012) object-relational database engine.

Over the study period, this resulted in 2.37 billion observations. By comparison, earlier work by Halpern *et al.* (2008) relied on 2.58 million observations, and that of Kaluza *et al.* (2010) relied on 490,517 journeys. This many-fold increase requires new methods for analysis, but rewards us with a rich view of shipping. AIS observations include both ship location and time, and

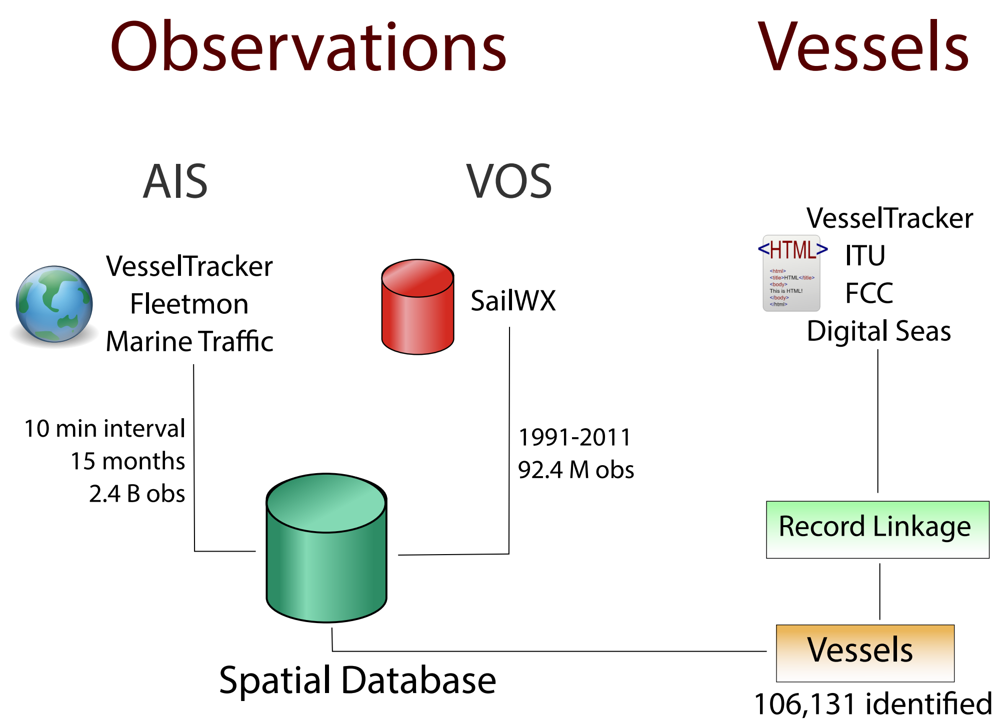

Figure 2.2: Data fusion workflow. Observations from two sources, AIS and VOS, are fused with vessel records to produce tracks of validated vessels.

frequently include additional attributes useful for addressing ecological questions (Table B.1). The VOS/SailWX dataset, consisting of 92.4 million records covering February 1991–September 2011, was provided in a MySQL database, which was converted and imported into PostgreSQL. These two sources of observations were then combined with vessel records to provide the validated vessel records used in the analysis (Figure 2.2). To validate the collected ship observations, vessel attribute records and additional ancillary data were identified, as described in subsequent sections.

2.2.1 Vessel Attribute Data

Tabular lists of individual ship attributes were collected from both authoritative and crowd-sourced media (Table 2.1).

1. Within the United States, the Federal Communications Commission (FCC) regulates use of the radio spectrum. In order to manage radio licenses, the FCC developed the Consolidated Database System (CDBS), which includes information on all vessels with US-registered radios.
2. The International Telecommunications Union (ITU) developed the Maritime mobile Access and Retrieval System (MARS) database to support the maritime community with up-to-date vessel information. Due to their role as a regulating body, participant states are required to provide information at regular intervals. This database is particularly useful as it includes details not available through other public means, such as the vessel owner, and vessel passenger capacity.
3. DigitalSeas includes volunteered vessel attribute information, collected primarily through corrected AIS observations, but has since gone off-line.
4. VesselTracker includes ship tracking, reporting and vessel records, alongside real-time AIS position data. Both their vessel data, and their AIS

observations were recorded in this study. The dataset shows particular strength for vessels located within European waters.

2.2.2 Land-sea Mask

A high resolution land-sea mask, derived from the Shuttle Radar Topography Mission (SRTM) Water Body Data (SWBD, Slater *et al.*, 2006), classifies the world into either land or sea at a three arcsecond resolution (~90m) for much of the world (56° S to 60° N). As a by-product of the SRTM digital terrain project (Rabus *et al.*, 2003), it has the advantage that it was collected over a short period, increasing self-consistency.

For the areas beyond that covered by SRTM, the Global Self-consistent, Hierarchical, High resolution Shoreline Database (GSHHS, Wessel and Smith, 1996) was used, an amalgamation of publicly available shoreline data. GSHHS is lower resolution than SWBD, but the vast majority of ship observations lie within the SRTM study area. The transition between these two layers was manually corrected to make a single, consistent, high-resolution land-sea mask at a three arcsecond (~90m) resolution.

2.2.3 Supplemental Data

Port databases were collected, containing coordinates and berth details for ports globally from two sources, the National Geospatial-Intelligence Agency's World Port Index (National Geospatial-Intelligence Agency, 2012), and the environmental impact of ports database developed in Halpern *et al.* (2008). Approximately 5,000 ports were identified from these two sources. Qualitative ship movements was also collected, based on historical charts such as a Cold War era CIA chart (Figure D.5). The original ship model produced in a previous modeling effort (Halpern *et al.*, 2008) was also helpful for comparison.

2.3 Validation

The observations are laden with caveats: because of limitations in the AIS protocol design, there is no direct way of validating a datum (Raymond and Schwehr), and the AIS radio broadcasts can be corrupted during transmission. As a result, many terrestrial locations have observations, including a particularly thick band centered around the prime meridian (Figure D.1). These records are more likely due to corruption of the longitude coordinates than a reverence for Greenwich. In addition to transmission errors are operator errors in any attribute set by the mariner, which includes all attributes other than the location and time provided by the GPS unit. These attributes are frequently either input incorrectly, or not kept up to date for attributes which change over time, such as the destination field. Observations in this dataset are suspect, and here the data are treated as guilty until proven innocent.

While there are inherent errors with the observations, the volume of data and the compiled ancillary datasets still allow for validation. This validation improves accuracy and minimizes the observations necessary to construct a model representation. Two useful approaches to addressing the validation of large geographic datasets are geographic data mining (Miller and Han, 2009) and geographic quality checking Goodchild and Li (2012). Here, I borrow the framework described in the latter work, and explore three avenues of quality assurance: Crowd-sourcing; social; and geographic approaches.

2.3.1 Quality Assurance

Crowd-sourcing While Goodchild and Li (2012) found that crowd-sourcing was generally ineffective for geographic facts, it can function when the domain is limited and the pool of expertise is vast. Crowd-sourcing becomes useful with AIS data when multiple receivers collect the same observation. Cross-referencing these sources provides confirmation, and can be used to measure transmission error.

Social Many mariners provide information to online services, and attribute quality from these sources is high. Vessel operators have direct knowledge of a ships' vitals, similar to a neighborhood resident who understands of local geography. The vessel operators then communicate to shipping aggregation sites, who organize a broad collection of vessel data, relying on trusted users to vet updates. These socially filtered records are then used as sources of vessel attribute data (Section 2.2.1).

Geographic For the error-prone vessel observations, geographic validation is key. Individual observations are point locations with time, and I rely on a handful of tests to provide validation. Each spatio-temporal observation $\langle x, t, z \rangle$ provides point location x, time t and multiple vessel attributes z. By cross-referencing the attributes, the joint probability of each attribute can be computed, inferring likelihood on geometry and time from the known traits of a particular vessel.

I impose basic validation on the location by referencing other geographic facts, as is used in an essential geographic information system function, the overlay operator. By checking the observations against a composite land-sea mask (Section 2.2.2), I estimate when the provided location is likely. This can be challenging, as many ships travel both by inland waterways such as rivers and canals, so these rules must be careful to define what is traversable, but most vessel classes are restricted to major water bodies and a handful of large canals. Shallow water imposes an additional constraint, depending on ship draft. This can be used to show, for example, the distance oil tankers keep from land masses.

For most vessel transportation classes, ships move between ports. In these classes, ships exhibiting patterns inconsistent with port-to-port movement are suspect. However, there are other fixed locations besides ports which require inclusion, such as ballast water exchange points, like one located 100 nautical miles offshore of California (Figure D.4), and canals, which enable movement between otherwise disconnected locations. Generally, port location is useful: because AIS receiving towers tend to be near ports, and thus the data captures the origin and destination pairs of most vessel journeys. Finally, I require observations have coordinates bounded by the coordinate space, and exclude any beyond the earth's surface.

2.3.2 Ship Types

An open problem in the maritime community is how to designate ship types. Ships can be classified on a variety of dimensions, including: Engine type, hull material, vessel function, and size. Ontologies have been developed to address the problem, but remain incomplete (Vries *et al.*, 2009). Here, I focus on use, and rely on the primary activity in which the vessel is engaged. Starting with the classes provided by our AIS and VOS sources, I collapsed them into nine major functional groups: authority, cargo, fishing, high-speed, passenger, pleasure, support, tanker, and "other" for vessels which don't map into any of our primary classes. Because the full attributes of each vessel are retained, and frequently includes multiple type labels, it is possible to break this down further for future analyses, but these broad classes served well for classifying distinct movement patterns. A full list of vessel classes is provided in Appendix B.3.

2.4 Record Linkage

Before producing derived representations of ship movement, I first must link observations to specific vessels. Here I rely on four vessel record datasets (Table 2.1), two provided by authoritative government bodies, and two from crowd-sourced websites. Each source contains inconsistent information, and before reconciling ships to observations, I cross-linked vessels into a single set of known vessels. Describing matching entities across datasets was initially

Table 2.1: Ship data sources.

Source	Code	Records	Linked	Attributes
Digital Seas	DS	212166	68002	name, IMO, MMSI, callsign, type, width, length
FCC[1] ULS[2]	FCC	319964	24531	name, MMSI, callsign, class, gross tonnage, length
ITU[3] MARS[4]	ITU	372183	75928	name, IMO, MMSI, callsign, class, owner, gross tonnage
VesselTracker	VT	126534	83372	name, IMO, MMSI, callsign, class, length

1. Federal Communications Commission; 2. Universal Licensing System; 3. International Telecommunication Union; 4. Maritime mobile Access and Retrieval System

developed with medical records, and more recently has advanced as the record linkage field in computer science (Christen, 2012).

By evaluating all possible pairwise combinations between source records, I map vessel records between sources. Using the methodology of record linkage, a set of rules maps records between the six possible source pairs. Each pair was evaluated for common, consistent attributes, and compared against these columns. The software package used, (FRIL, Jurczyk *et al.*, 2008), provides an Expectation Maximization algorithm to iteratively optimize the column weighting, but due to computational limits, this proved ineffective. Instead, samples were examined, and matching criteria were set by tuning both the weightings and acceptance levels to fit a training set of valid linkages (Table B.2).

For most attributes, the equal fields metric which requires both values to be the same, or the Jaro-Winkler distance metric were used. The Jaro-Winkler metric has useful properties for this data: Effective in both numeric and textual comparisons, this approach is good at mapping between noisy sources, and a study of string comparison metrics found it to be both efficient and effective, with a high match rate on diverse data (Cohen *et al.*, 2003). Details on the Jaro-Winker distance are described in Appendix A.1.

Once each pairwise combination between sources was completed using FRIL, a second rule-based method was developed to capture missed valid pairings. If vessels had equality in any two attributes of the set {MMSI, callsign, name}, if callsign and vessel length matched, or if the IMO number provided was a valid seven digit number, then the pair was linked. The rule-based method successfully found many additional valid linkages.

2.4.1 Linkage Validation

Detecting linking errors is problematic, as vessel records are correlated – IMO number, callsign and name may differ by only one character for two ships in the same fleet. Linkage validation serves to retain these distinct vessels, while simultaneously eliminating differences due to entry error. I developed a validation score to remove over-aggressive links, which falsely linked two different vessels. This provides a second quality check which can be used to threshold the results. I sampled invalid joins produced in the record linkage process, and found the specific traits that were common across the sample:

1. >6 linked records
2. >1 radio callsign
3. >1 MMSI
4. clear name mismatches
5. vessel assigned to multiple incompatible classes

Using these traits, I assigned validation scores and probable vessel class. The Jaro-Winkler metric was again employed to compare attribute matches for ship name and radio callsign, with $1 - d_w$ being added to the validation score. For attributes that had a single value, the validation score was increased by one, otherwise vessels which had more than two MMSIs or five linked records had one point removed from their score. Finally, if a ship was identified as being in multiple incompatible vessel classes, one point was subtracted for each incompatible class. Final scores for all vessels are shown in Figure 2.3.

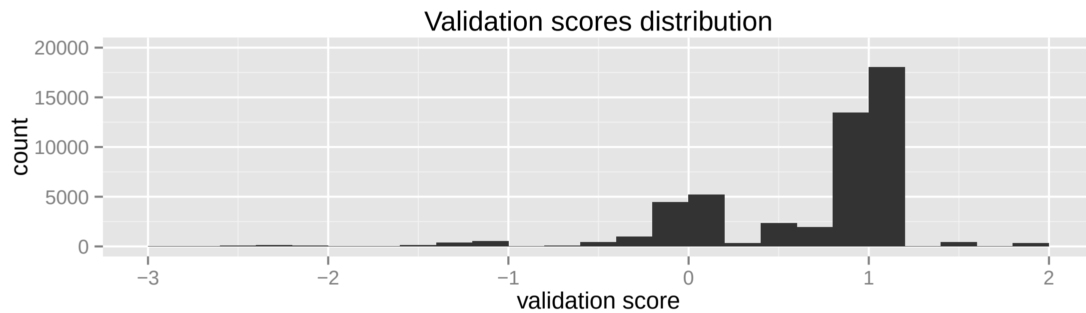

Figure 2.3: Linkage validation scores, bin width=0.3

Only vessels with validation scores above zero were used in later steps. Vessels are also assigned to a single vessel class by the validation. While two vessel record sources are authoritative, the best available vessel record data remains commercial. A 1% sample of the records was compared to those provided in Equasis (Equasis, 2012), which includes validated records from the commercial fleet. After cross-linking, the validation scores showed good correspondence with Equasis vessels.

2.4.2 Ship Identifiers

Identifiers to designate unique vessels also poses a problem, as there is no standardized and universal identifier. By incorporating information from many different identifiers and focusing on those less fluid, such as IMO number and radio callsign, increases the odds of valid matches. However, additional attributes such as the Maritime Mobile Service Identity (MMSI), and vessel name remain useful, particularly because they are broadcast alongside each AIS record. Vessel operators may not maintain this information, making it important to cross-validate the attributes to improve match rates.

2.5 Data Representations

Ship observations are contributed by vessels on both a voluntarily and mandated basis, and using geographic quality checking methods produces a validated set of observations. While individual observations have a simple representation (a location, time, and attributes), effective use requires multiple representations (Goodchild, 1992) and the spatio-temporal modeling techniques of time geography (Miller and Bridwell, 2008). Because there is no single, optimal representation of data, I produce a set of representations which, when matched to particular uses, can provide insight. Here, I maintain both discrete object and continuous fields representations (Figure 2.4), to address questions about the ecological effects of shipping. While the point representation alone is too simple for understanding shipping's effects, incorporating too much complexity in the representation risks making computation infeasible (De Smith *et al.*, 2007).

One representation developed is speed density (Section 2.5.3), useful for understanding whale-vessel interactions. Combining vessel type, vessel draft, and vessel length with a derived speed estimate, provides a way to spatially model vessel risk in whale-vessel encounters. Similar combinations of vessel attributes and derived representations are useful for other ecological questions.

2.5.1 Points

AIS records are sampled at 600 second intervals (10 minutes). For most vessel classes, this sample rate retains the movement patterns, and is a similar rate to that used in studies using aggregation to remove noise. For example, Vries *et al.* (2009) uses piecewise aggregate approximation at a 300 second interval to "strike a good balance between capturing the general movement and ignoring noise". AIS observations are recorded via GPS, and for validated observations, positional accuracy is high. Because vessels are much larger than GPS position errors, uncertainty in where the GPS unit is located on the vessel is the primary source of positional error for most vessel classes.

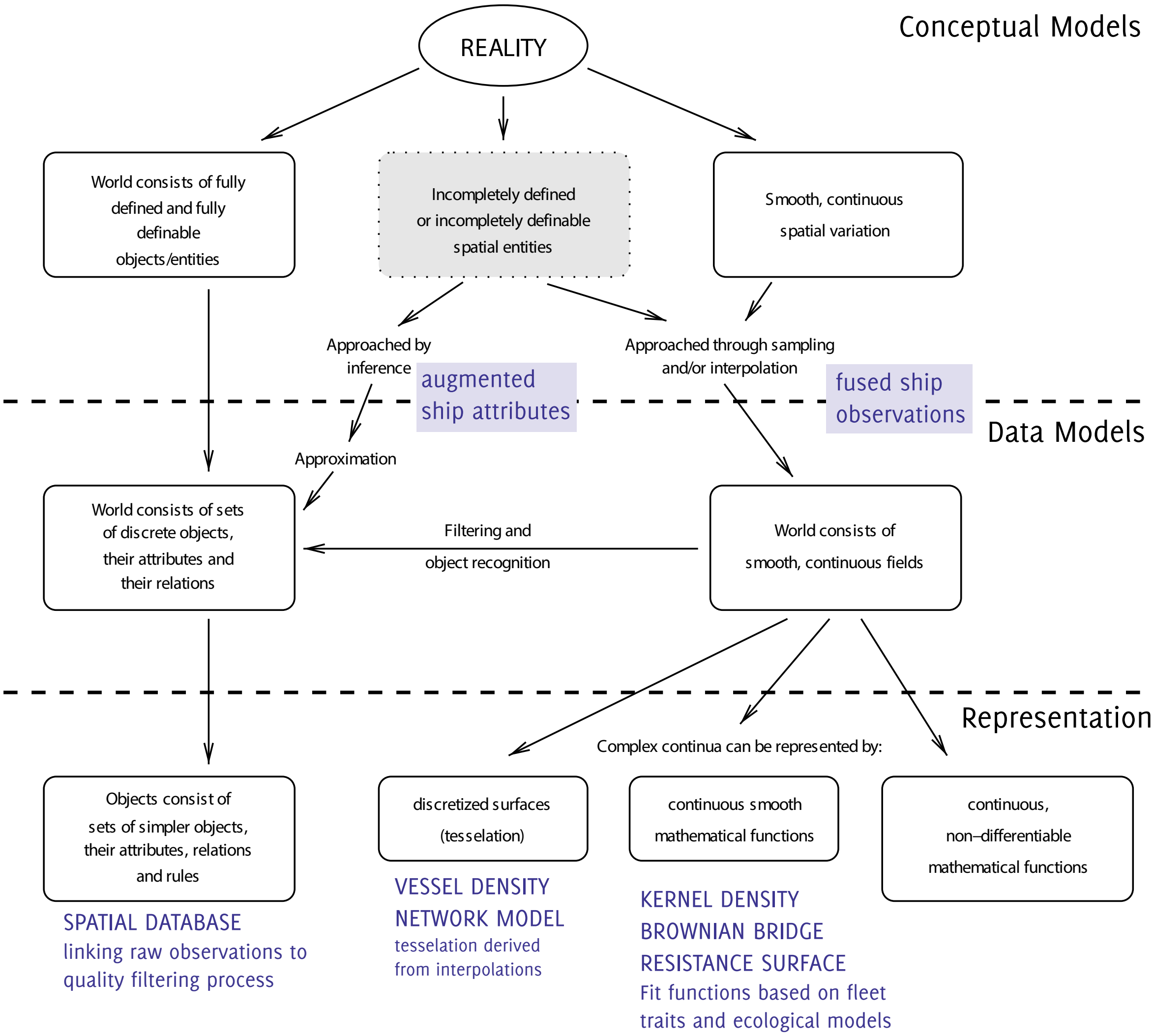

Figure 2.4: Spatial abstractions, this paper's representations in blue. Adopted from Bivand (2011), based on original by Burrough *et al.* (1996).

2.5.2 Tracks

Tracks interpolate discretely sampled time points $\langle x, t \rangle$ into a continuous phenomenon. The uncertainty of the interpolation can be described by a time

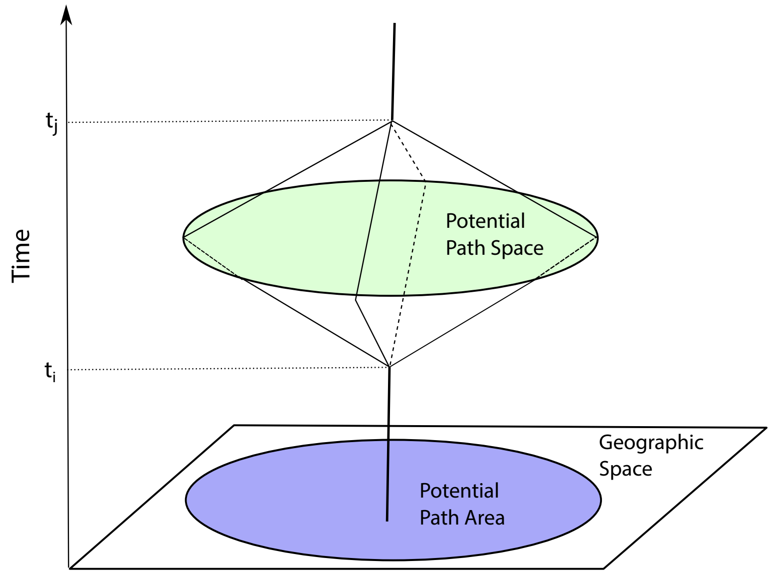

Figure 2.5: A simple space–time prism; based on Wu and Miller (2002).

prism (Figure 2.5), which bounds the area based on the maximum velocity of the object at motion. Here, I constrain the track segments to have movement speed above 0.5 Knots (Kt)/hr, below which vessel control is difficult, and speed below 40 Kt/hr, which derived empirically from our ship speed distributions (Figure 3.4) and vessel engine characteristics. I further require that to be considered a track, the above criteria must be met for at least one segment of that vessel's transit. The ships are then constrained to move along the shortest distance on a geoid, or great circles, using GeographicLib (Karney, 2012). This information is reused in computing the speed along each segment. Individual vessel tracks are stored as a continuous line segment, with vertices along the line reflecting known observations. One limitation of this representation is segments which pass through landmasses can exist.

2.5.3 Field-based Models

Initially, kernel density estimation (KDE) was used to produce field based estimates. However, based on the data volume, KDE proved more useful to provide models directly from discretized tracks, which are used in the two following representations.

Ship density by type The primary output of this work is a field-based density model of ship movement. This view is useful in a wide variety of contexts, from exercises in marine spatial planning to detecting conflict zones between resource users, and the simple density estimation in Halpern *et al.* (2008) remains a highly requested product. By retaining separate vessel classes, specific ecological concerns can be addressed in a way not possible when weighting all vessels as equivalent. Hull type, cargo, engine, length, draft, speed, and movement patterns all vary based on vessel type.

Each vessel track was converted to raster on a 90 arcsecond grid (~5.5km at the equator) and an equal area grid in the Hobo Dyer projection (Figure 3.2), as detailed in Appendix A.3. The projected version assures that the density function is computed on grid cells representing the same area for each cell, unlike the geographic grid where area varies by latitude. A vessel is counted only once for each cell the vessel passes through, as the focus here is on overall movement patterns, de-emphasizing vessels with limited movement. A single raster is produced for each validated vessel contained within either the AIS or VOS dataset. Next, the rasterized vessel tracks were combined using map algebra to produce density maps, separately for AIS and VOS data, and for each of our vessel classes. This leaves us with nine vessel classes and two vessel density estimates per class, or a total of eighteen vessel density estimates.

To merge the two vessel density estimates (VOS and AIS), which characterize different parts of ship movement patterns, a weighting scheme is used. For each cell, the output density, s, is calculated as the standardized equal-weighted addition between the two inputs:

$$s = \frac{R_{AIS}}{max(R_{AIS})} + \frac{R_{VOS}}{max(R_{VOS})} \tag{2.1}$$

Speed density estimations of ships by ship type Ship speed plays a critical role in determining the survivability of collisions for many marine species (Vanderlaan *et al.*, 2009). Speed also directly relates to the emissions profile of a vessel, and it has been shown that speed reduction alone can reduce 50-80% of vessel greenhouse gas emissions (Lack *et al.*, 2011). Conversely, decreased speeds require more vessels to ship the same volume of goods or passengers, however container shipping companies are mitigating this problem by moving to significantly larger capacity vessels (Notteboom, 2004).

Average speed per cell was calculated as sum speed over all observations $R_{AIS} \cup R_{VOS} = n$, and dividing it by the total number of observations, but only for those locations where a sufficient density s is present:

$$\bar{s} = \begin{cases} \frac{\sum_{i=0}^{n} s}{x} & s \geq 10 \\ 0 & s < 10 \end{cases} \tag{2.2}$$

This paper contributes rich data predominantly within 100km of shores, where most human and biological users of the ocean persist, and building our knowledge of these areas is particularly valuable. Ship traffic is most dense, regulated, and complex within the exclusive economic zones of nations, and it is useful property that this data are primarily within these areas.

Chapter 3

Results

3.1 Vessel Results

For vessel classes with mandatory AIS, the majority of vessels are identified, and comparing against commercial estimates of fleet size, this data contains of 75.6% for cargo vessels, 69.4% for tankers, and 62.9% for passenger ships. Details for all classes are provided in Table 3.1. In vessel classes without AIS requirements, the data are a small sample of the vessel population. For example, authority vessels here are a small subset of search and rescue vessels, and does not contain naval or police vessels. Fishing vessels are similarly limited, as only areas which mandate AIS have observations, leaving out classes such as artisanal fishing.

Table 3.1: Summary of analyzed ship data

Type	Ships (AIS)	Ships (VOS)	fleet size	coverage (%)	observations (M)
Cargo	25214	5838	33392[1]	75.6	665.45
Tanker	9758	2375	14068[1]	69.4	264.42
Passenger	4007	777	6370[1]	62.9	142.16
Support	9954	735	25234[1]	39.4	298.02
High-speed	404	81	1178	34.3	2.52
Fishing	11186	349	51200[3]	21.8	6.87
Pleasure	20727	661	800,000[2]	2.59	267.48
Other	9507	1400	–	–	4.75
Authority	656	55	–	–	1.44

1. Equasis (2012) 2. Westwood *et al.* (2001) 3. Food and Agriculture Organization (2012).

3.2 Record Linkage

Record linkage matched 30-50% of each record source pair (Table 3.2). Comparing only between the two authoritative sources, the FCC–ITU pair produces inconsistent results, pointing toward a continued need for better unified and public vessel identifiers. After pairwise linkage, matched records were further cross-linked, to account for vessels appearing in multiple sources. The number of links per ship averaged 3.5 ($\mu = 3.49, \sigma = 0.828$), although this distribution is skewed by a handful of incorrectly linked vessels.

Table 3.2: Ship record linkage results.

Source A	Source B	matched records	match rate
DigitalSeas	FCC	3481	50.35%[1]
DigitalSeas	ITU	41380	30.73%
DigitalSeas	VesselTracker	72286	53.68%
FCC	ITU	27874	50.58%[1]
FCC	VesselTracker	5282	53.23%[1]
VesselTracker	ITU	54727	43.25%

1. FCC data are US only, match rate is of US-only data from these sources.

3.3 Geographic Validation

The land-sea mask developed, as described in Section 2.2.2, was compared against each observation to detect if the point was located in a water body, which should be true of any ocean vessel. However, this validation technique was insufficient to resolve many observations, which lie at the edge of the land-sea mask when docked or when moving near shore (Figure D.3). Due to these limitations, the observations "on land" were retained. A improved approach might employ a local density estimation on the vessels, and determine a mask based on known transits. Here, on-land observations were filtered during the track generation process, which restricts vessel movements based on speed, eliminating erroneous observations.

3.4 Ship Results

3.4.1 Density

The resulting vessel density model shows strong differences in the movement patterns between vessel classes.

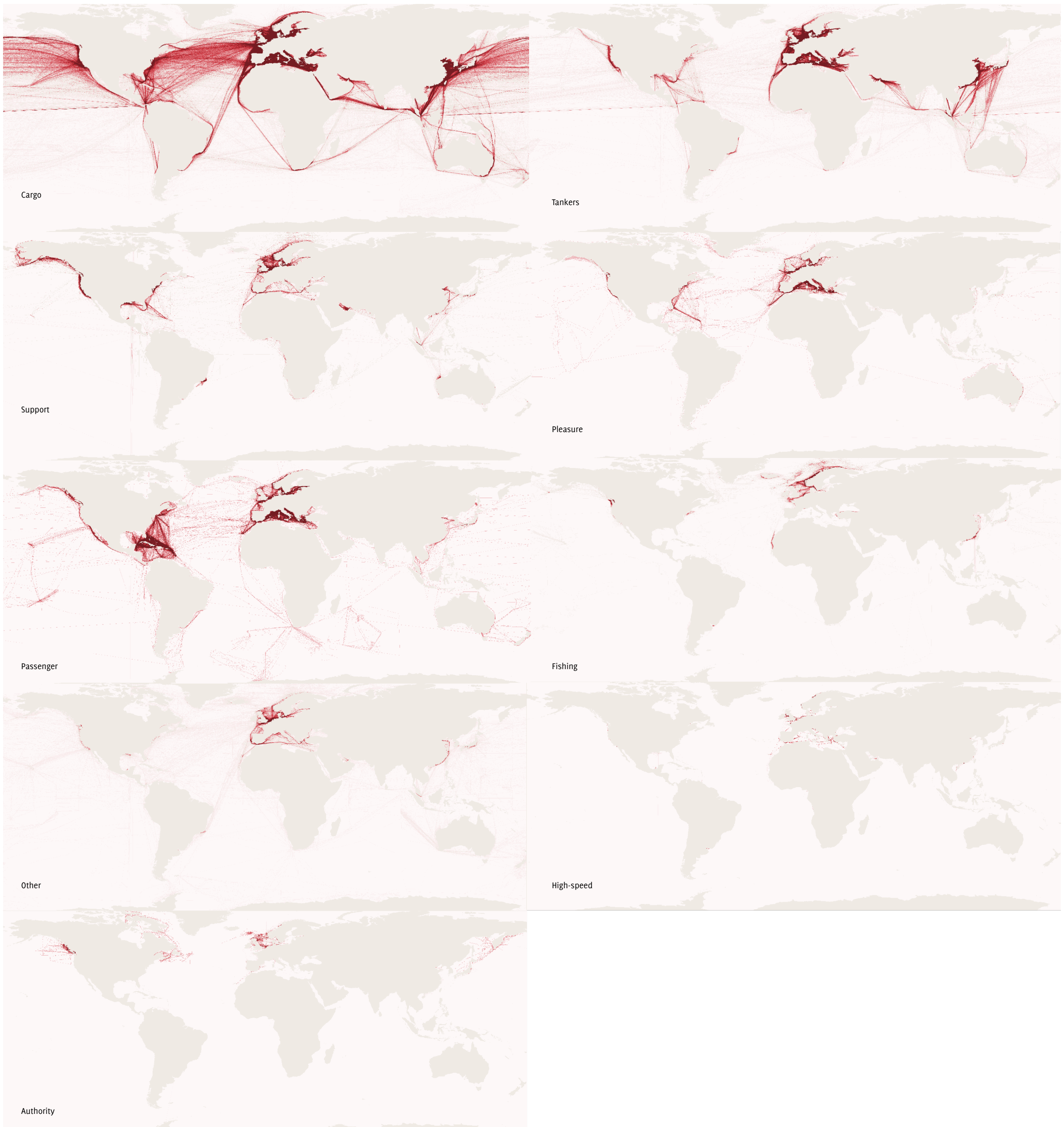

Figure 3.1: Global ship movement densities. Vessel classes rescaled to show pattern.

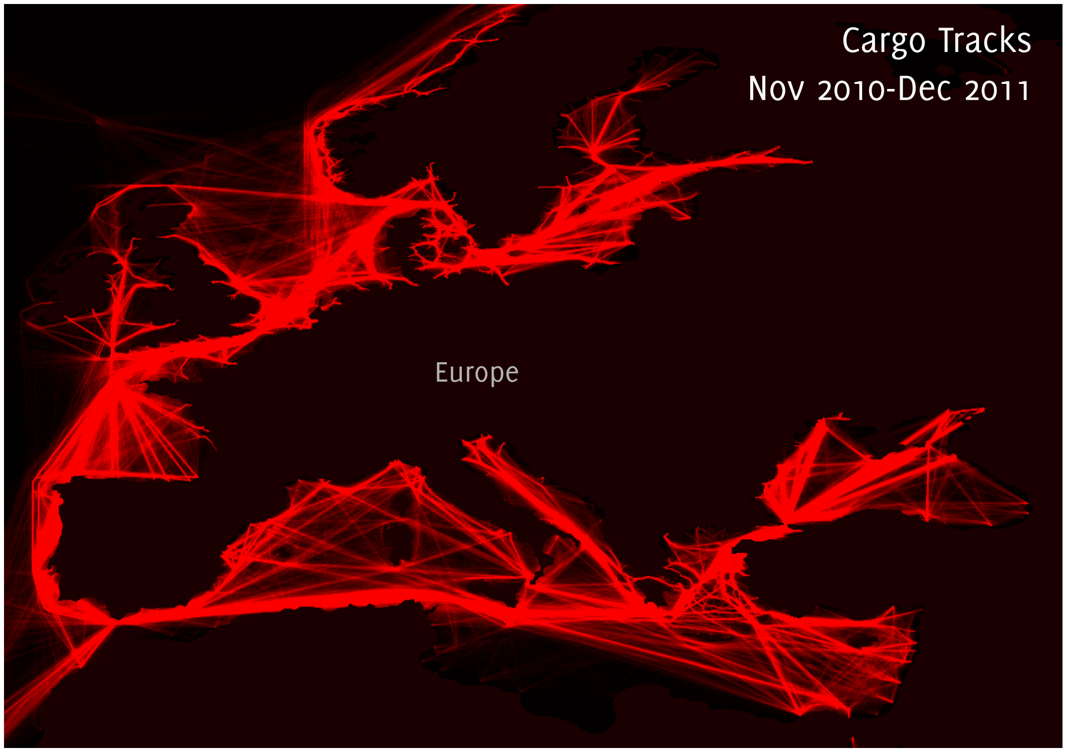

Figure 3.2: Cargo density grid, generated by combining AIS and VOS records.

3.4.2 Speed

Ship speed is an important indicator to answer a variety of questions, but is challenging to represent spatially. Ship speed is sensitive to the observation frequency at a particular location, and sensitive to accurate distance and time measurements, all three of which are needed to compute averaged velocities. The speed distribution of each vessel class was examined, through boxplots (Figure 3.3) and kernel density estimation (Figure 3.4). The density estimation shows that for many vessel classes, distinct speed patterns exist. For example,

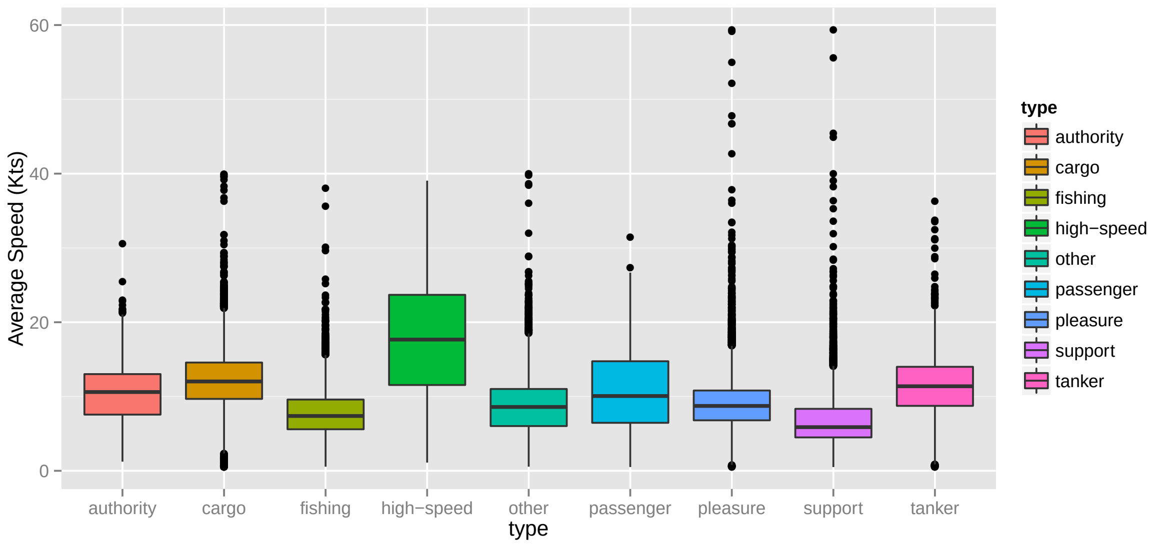

Figure 3.3: Boxplot of validated ship speeds by class.

support vessels, which include tugs and barges move much more slowly than high-speed transport vessels. Other classes are less distinct in speed signature: cargo and tanker vessels have surprisingly similar average speeds, though disaggregating the results spatially (Figure 3.5) shows they have strong regional differences masked in the overall average.

3.5 Uncertainty

AIS observations rely on terrestrial radio antennae to collect data, and their coverage of the near shore is incomplete. Many areas lack AIS observations, but are well-represented in the VOS data, showing data gaps. To account for

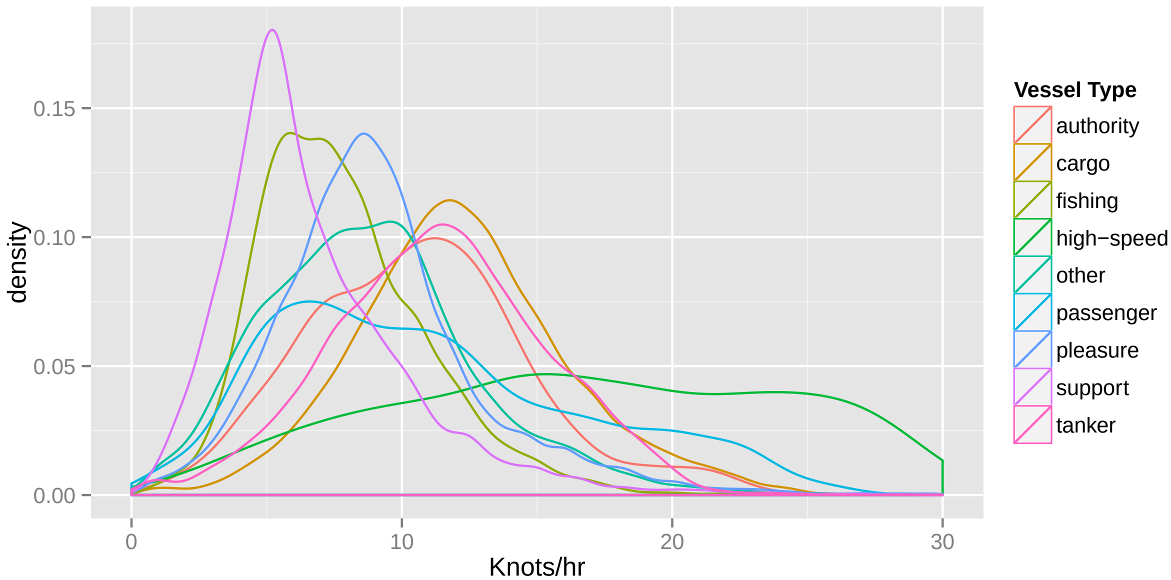

Figure 3.4: Distribution of validated ship movement speeds, computed using kernel density comparisons, as described in Bowman and Azzalini (1997).

this gap, I examined the coverage of the largest world ports, as identified in the World Port Index (National Geospatial-Intelligence Agency, 2012). Of the large ports in the index, 112 of 155 (72.3%) had 100 or more unique vessels, and of medium ports, 187 of 357 (52.4%). The identified ports show concordance with the major ports identified in Ducruet and Notteboom (2012).

Individual vessel movement is also modeled simply, and barriers to movement are not accounted for, which can be important in areas with geographic and political features which limit movement. Future work should address this and other sources of uncertainty within the results.

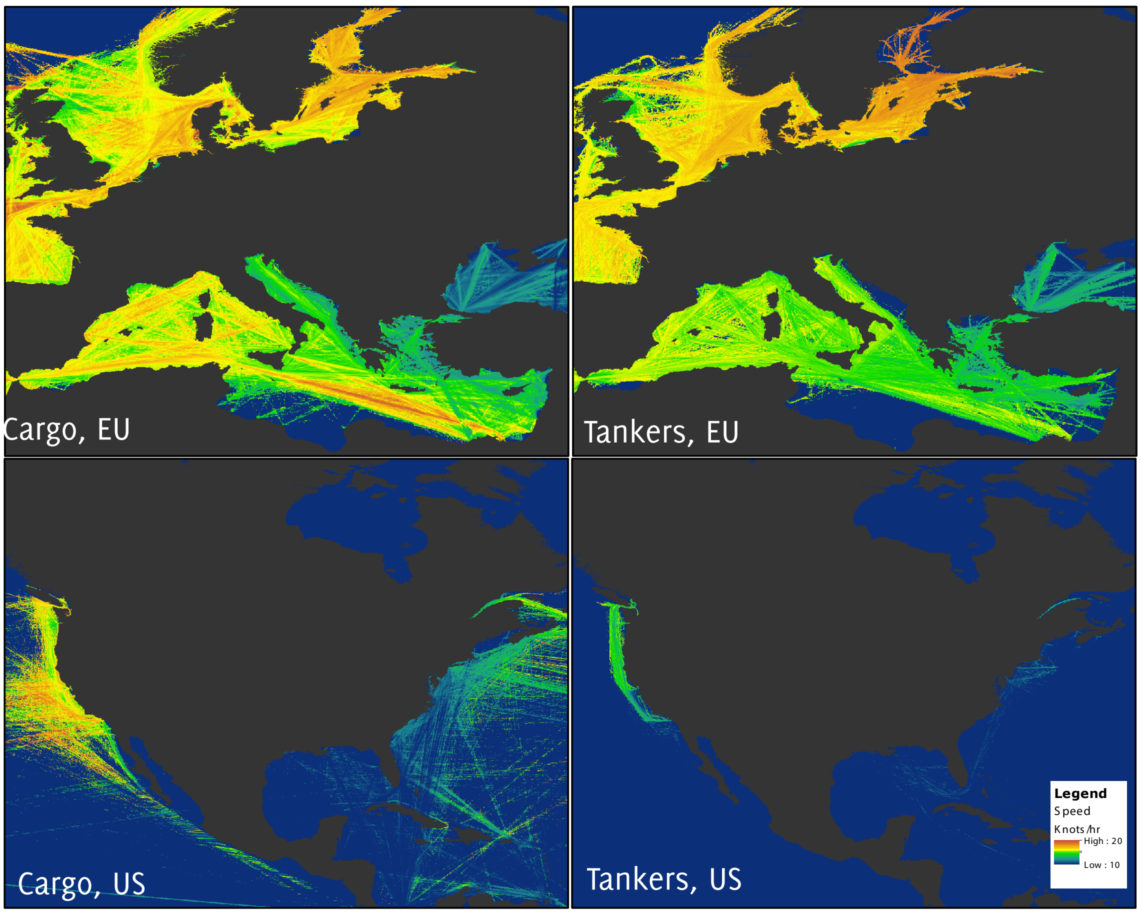

Figure 3.5: Average ship speed examples. Scale shows 10-20 Kts/hr range for all vessels. Of note is the large difference in speeds between the opposite shores of North America.

Chapter 4

Discussion

Shipping is a major user of the ocean, but little is known about its distribution and effects. Here I built the first validated and global models of ship movement. Shipping companies acknowledge the importance of managing the ocean holistically, but lack the scientific knowledge and tools to do so effectively. By incorporating ecological information alongside logistical efficiencies, system robustness can be improved.

4.1 Protected Areas and Spatial Planning

One approach for incorporating ecological information in ocean decision making is with marine spatial planning (MSP), a promising avenue for bringing stakeholders together (Merrifield *et al.*, 2012). This approach requires extensive data, particularly on the three major areas of use: fisheries management, transportation, and energy management (Lubchenco and Petes, 2010). Transportation in the ocean is the least studied of the three, and here I have shown how volunteered geographic information methods, along with volunteered observations, can provide us a way to approach the data-poor problem of shipping.

Marine protected areas (MPAs) have been shown to be effective (Halpern and Warner, 2002), however multi-dimensional ocean use can also take advantage of dynamic MPAs, to mitigate against infrequent but high-risk events (Boersma and Parrish, 1999). Dynamic MPAs may also be introduced to manage seasonal events, such as species migrations, in a way that shipping lanes and general speed reductions can not accommodate. Dynamic MPAs will likely rely on providing users, such as ship operators, with real-time information about the state of the environment. Incorporating environmental data directly within AIS systems (Schwehr *et al.*, 2011) provides a useful way to integrate across these domains.

4.2 Ecological Effects

The ecological effects of shipping have been touched on in this manuscript, an example of ecology coming to terms with the spatial context at the unit of analysis, allowing model scope to extended across disciplines (Tilman and Kareiva, 1997). A few illustrations of how this work can be applied to ecological questions follow.

Ballast Ballast water is used to prevent vessel capsizing by weighting the ship within the water, vessels can transport upwards of 20 million gallons of water for ballast. By transporting large parcels of water between otherwise disconnected areas of the ocean, ballast water introduces ecological coupling and can transport invasive species (Ruiz *et al.*, 2000; Keller *et al.*, 2011). Ballast water exchange also can transport viruses and bacteria, which have human health consequences beyond ecosystem function. By providing global, per-vessel movement, ballast exchange can be examined in greater detail, as can vessel use of ballast exchange points.

Ship Strikes Ship strikes, or species to vessel collisions, also cause injury and death to marine mammals, driving whale population declines, as is the case with the North Atlantic right whale (Moore *et al.*, 2007). Globally, shipping increased three-fold in the last 50 years, coinciding with increases in average ship speed (Vanderlaan *et al.*, 2009). This combination has led led to large increases in shipping-induced marine mammal fatalities. The IMO has helped implement improvements to shipping lanes (Lagueux *et al.*, 2011; McKenna *et al.*, 2012), but the model produced here can help us in locating key whale-vessel areas in those locations not as intensively monitored as the US coastal waters. Dynamic MPAs may also serve a role to mitigate this impact.

Noise Pollution Sound, which is efficiently transmitted through water, is used by many species who have adapted sensitive hearing to forage and communicate. Thus, engine noise from ships has led to extensive sound pollution, causing behavioral changes in a variety of marine species, and mortality when exposed to loud sounds from naval activities (Hatch and Fristrup, 2009). NOAA recently created a working group to evaluate the impact of noise pollution on marine life, and the model created here improves our understanding of the distribution of shipping, which can be incorporated into future models of ship noise to improve ocean management.

4.3 Regulation

The primary extractive use of the ocean is fishing, where our understanding has been advanced by intensive investments to study its effects. Many fisheries are now managed systems (Worm *et al.*, 2009), as unmanaged fish stocks have repeatedly collapsed (Costello *et al.*, 2012). Similar management and regulation is noticeably absent from the shipping industry.

Shipping is regulated by the International Maritime Organization (IMO), which enforces industry rules and standards. However, the transnational nature of the industry has led to limited enforcement. As is common in consensus-based international bodies (Cogan, 2009), the IMO can be slow to respond to issues. For example, while the IMO acknowledges that a 20% reduction in greenhouse gasses can be accomplished this decade without additional costs to operators (International Maritime Organization, 2009), it has not implemented new emissions rules. In light of the regulatory environment, alternatives, such as tying ship insurance rates to ecologically important areas of the ocean, may be preferable.

4.4 Transportation Networks

One approach to managing a complex system like shipping is to study its network representation. Transportation networks with geographically-fixed edges and nodes, such as road and rail, form the basis of transportation geography (Rodrigue *et al.*, 2009). Unlike land-based networks, shipping exists in between the extremes of a fixed network and two dimensional Brownian bridges, as vessels must transport both goods and passengers to specific locations (primarily ports) but are free in choosing movement between destinations, except in near shore areas regulated by shipping lanes. Air transportation shares network similarities, but strong regulation has led to predetermined flight paths, with deviations generally limited to emergencies or extreme weather.

Because of inherent risks in air transportation, detailed, well-vetted information on flight paths is provided by government agencies (Guimera *et al.*, 2005), simplifying modeling. Ship movements have no equivalent top-down data collection effort, and while organizations such as the US Coast Guard have made efforts, the system remains rudimentary. Most information is provided via private contracts, through organisations such as Lloyd's of London, who have been recording information on shipping since 1774 (Lloyd's Register-Fairplay Ltd., 2010). As a result, limited public data are available on the shipping trade, despite being identified by the Federal Geographic Data Committee (Federal Geographic Data Committee, 2008) as a key transportation component to the national spatial data infrastructure (Currier *et al.*, 2012).

Transportation networks play an important economic role (Canning and Fay, 1993). Because this role, information on shipping is valuable, leading to a cottage industry of information sales and limiting its public availability. A future extension of this work might incorporate network theory, which has been applied to port-to-port networks (Kaluza *et al.*, 2010; Ducruet and Notteboom, 2012), but has not yet been applied to spatially explicit vessel movement.

Chapter 5

Conclusion

As noted by Goodchild, “changing technology and economics are moving map production from a system of unified central production to a local patchwork, and the old radial system of dissemination is being replaced with a complex network”(Goodchild, 1999). By using the quality assurance methods of record linkage and geographic validation, provides an approach to filter unreliable inputs of shipping data, and take advantage of the newly formed patchwork to broaden our understanding of ocean transportation. The near future will involve global, real-time, high resolution ship data (Jones, 2012; Carson-Jackson, 2012), but we need methods which accommodate data curation and integration. By using quality assurange methods, I have shown that multiple data dimensions can be incorporated, and heterogenous errors can be rectified.

Recent calls for increased marine spatial planning at both the national and international level should be met with increased production of fundamental datasets required for effective planning. This work can help advancement of both marine spatial planning and ecosystem-based management, and to help organizations like the IMO provide more effective regulation of shipping, perhaps using insurance-based incentives to reflect environmental costs. The movement toward ubiquitous and real-time data provides an opportunity to greatly improve our management of ocean resources.

This work is a step toward understanging the global effects of shipping. True cost-path movements, which account for vessel preference and barriers, will give us a way of understanding the relative value of different areas within the ocean to the shipping industry. An abstract network model, which incorporates the detailed movement model developed, would allow us to interact with this complex data in a much simpler way, and potentially lead to breakthroughs in marine spatial planning at regional scales. More immediately, these results can feed into understanding the anthropogenic sources of sound in the ocean, and improve models of ship strikes, by providing detailed, holistic speed models for most of the oceans.

Appendices

Appendix A

Method Details

A.1 Record Linkage

The Jaro-Winkler formula is defined in two steps. First, The Jaro distance, d_j, is defined as:

$$d_j = \begin{cases} 0 & \text{if } m = 0 \\ \frac{1}{3}\left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m}\right) & \text{otherwise} \end{cases} \tag{A.1}$$

Where:

m = number of matching patterns

t = number of transposed characters

|s1| = length of first string

|s2| = length of second string

Two characters are considered matching when they are no further apart than:

$$\left\lfloor \frac{\max(|s_1|, |s_2|)}{2} \right\rfloor - 1 \tag{A.2}$$

The second component, added by Winkler, preferentially weights strings which match from the beginning, set by the prefix length l. Thus, the Jaro-Winkler distance is defined:

$$d_w = d_j + (\ell p(1 - d_j)) \tag{A.3}$$

Where p is a constant scaling factor to adjust for the strength of common prefixes. In its usage here, $p = 0.1$ and $l = 4$.

A.2 Observation Filtering

As in many large datasets, the distribution of observations per ship follows an approximate power-law distribution. Using the raw number of observations received in our 15 month window provides a simple filter. The peak in the kernel density estimation (Figure A.1) is seen around 10^4 observations, with a clear drop-off after 10^5, which is consistent with the theoretical maximum (one observation in every sample) of about 6.8×10^5.

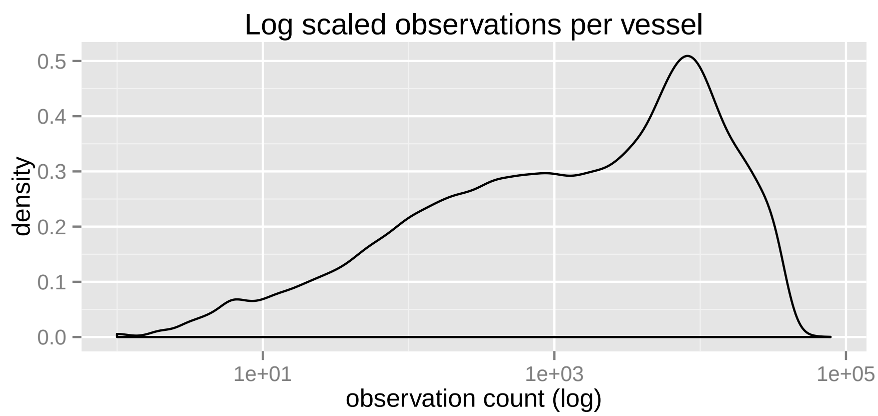

Figure A.1: Observations per vessel, kernel density estimation.

A.3 Movement Modeling

After storing the raw observations in a spatial database (PostGIS), I used parallelized code to quickly aggregate the large volume of observations. Taking advantage of the database, observations were ordered on disk based on vessel, and spatial indices were created. This allows us to quickly filter the data both based on attributes and spatial traits.

To rasterize vessel tracks, gdal_rasterize from GDAL was used to convert between vector and raster representations. Bresenham's line algorithm Bresenham (1965) determined the output cells considered used within each track. I only counted a ship moving through a single cell a single time to capture overall movement patterns, but future work should also incorporate total trips to capture movement patterns such as regular short distance trips passenger ferries engage in. When combining the data, a parallelized version of Matthew Perry's `gdal_add.py` script was used to quickly combine thousands of raster ship tracks.

Appendix B

Tables

Type	Sub-type	Vessels
cargo	cargo ship	30355
	merchant	2919
	bulk carrier	1338
	container ship	669
	general cargo	608
	vehicle carrier	56
tanker	tankship	10460
	tanker	8731
	oil tanker	1312
	liquefied gas carrier	112
	chemical carrier	39
other	merchant	17460

Type	Sub-type	Vessels
	other ship	9421
	unspecified	4580
	motor boat	4231
	inland waterways	3825
	sloop	2799
	reserved for future use	2164
	all other activities	1345
	reserved for regional use	712
support	pusher/tug	4392
	tug	7885
	towing vessel	2747
	supply vessel	1501
	service vessels	1388
	trawler	1056
	dredger	995
	vessel engaged in dredging or underwater operation	776
	pilot vessel	685
pleasure	pleasure/leisure	22013

Type	Sub-type	Vessels
	pleasure	14573
	pleasure craft	7290
	sailing vessel	5527
	yacht	8214
fishing	fishing vessel	11110
	fishing boat	8454
	fishing industry	5767
	fishing	3213
passenger	passenger ship	6478
	ferry	704
high-speed	high-speed craft	658
	high speed craft	520
authority	sar-vessel	699
	search and rescue vessel	609
	rescue vessel	102

Table B.3: Detailed ship classes, derived from observations.

Table B.1: AIS broadcast attributes. Update frequency depends on ship speed, but varies between a minimum of a record every 2 seconds for quickly moving vessels, to once per 3 minutes for moored vessels. Additional attributes are available, but infrequently used.

Attribute	Accuracy
Location (fixed from GPS signal)	$\simeq$10 meter
Timestamp (on broadcast)	$\pm$100 ms *transmission & processing*
Name	
Call Sign	
Maritime Mobile Service Identity	
Heading	
Speed	
Destination	*often incorrect*

Table B.2: Ship record linkage methods used.

Source *A*	Source *B*	acceptance level	column	distance metric	weight
DS	ITU	92	callsign	Jaro-Winkler[1]	50
			MMSI	equal	40
			name	Jaro-Winkler	40
DS	VT	85	callsign	Jaro-Winkler	60
			IMO	Jaro-Winkler	20
			name	Jaro-Winkler	20
FCC	ITU	85	callsign	Jaro-Winkler	95
			name	Jaro-Winkler	5
FCC	VT	95	callsign	Jaro-Winkler	66
			name	Jaro-Winkler	5
			MMSI	Jaro-Winkler	24
			length	equal	5
VT	ITU	80	callsign	Jaro-Winkler	20
			MMSI	Edit Distance	30
			name	Jaro-Winkler	10
			IMO	Jaro-Winkler	40
DS	FCC	95	callsign	equal	99
			name	Jaro-Winkler	1

1. Jaro-Winkler distance (Winkler, 1990): length $l = 4$ and scaling factor $p = 0.1$

Appendix C

Software

This work would not have been possible without extensive contributions from others in the form of software, both commercial and open source. An overview of the software used in the project is provided, and code produced is available at `https://github.com/scw/thesis`.

To store the raw observations, SQLite was used for rapid development, but the bulk of the observations were stored in the object-relational database PostgreSQL, where analysis was carried out using either native SQL or with the spatial extension PostGIS. The GDAL library was used extensively for transforming both raster and vector data. Record linkage was performed using FRIL, with some record linkage integrated into the database using pg_similarity, and in Python using the jellyfish module. Integration and data processing used Python, which combines a wide base of environments within a single language. R was used to perform statistical analyses and summarize results. GRASS was used to compute model outputs and spatial statistics, and maps were produced in ArcGIS and Quantum GIS.

Table C.1: Software used. A complete listing of software packages used in the analysis.

Software package	version
ArcGIS	10.1
Bash	4.2.24
FRIL	2.1.5
GDAL	1.8.0
"	1.9.1
"	1.9.2
Git	1.7.9.5
GRASS	6.4
"	7.0
pg_similarity	0.0.19
PostGIS	1.5.3
"	2.0.1
PostgreSQL	8.4.14
"	9.2.1
Python	2.6
"	2.7.2
Quantum GIS	1.8
R	2.14
"	2.15
SQLite	3.7.9

Table C.2: External libraries used. Python was used extensively throughout the project, and R was used for calculations, validation, and statistics. The following external libraries were used in each environment.

R packages	Python libraries
foreign	BeautifulSoup
geosphere	gdal
ggplot2	geographiclib
maptools	geojson
plotrix	geopy
raster	grass
RPostgreSQL	jellyfish
rgdal	numpy
rgeos	ogr
sm	psycopg2
sp	requests

Appendix D

Figures

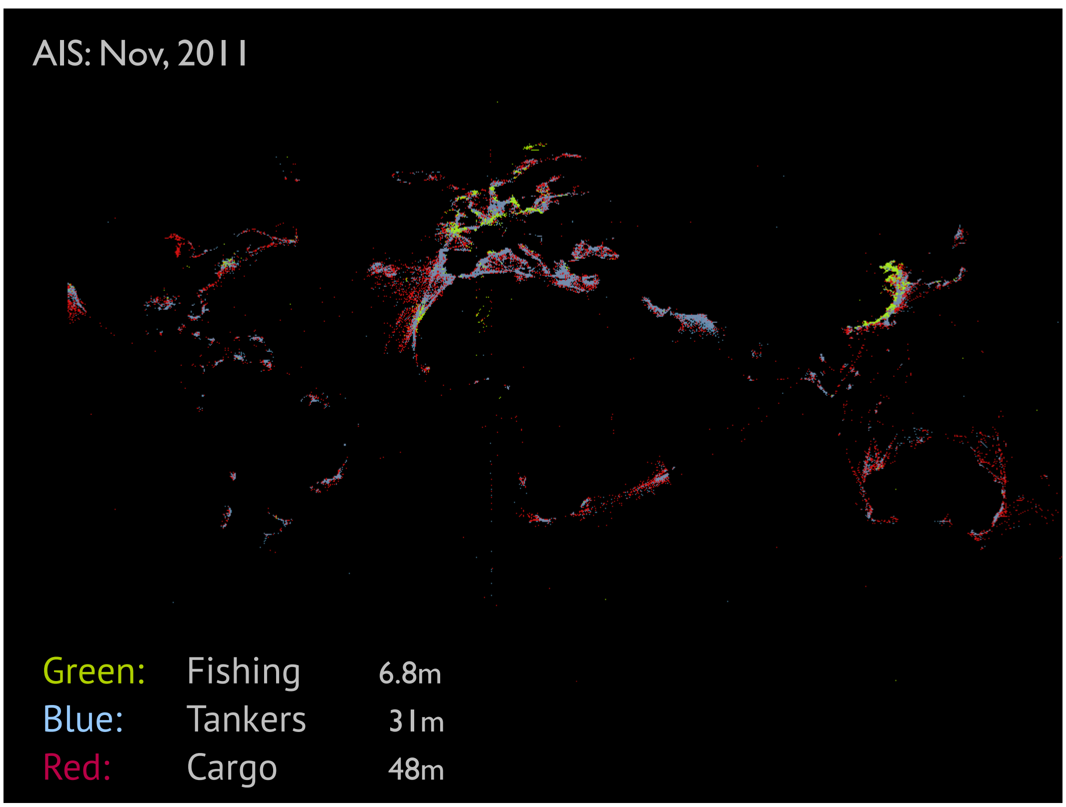

Figure D.1: Raw AIS observations, November 2011. Note the observations located in the Hoggar Mountains in Algeria.

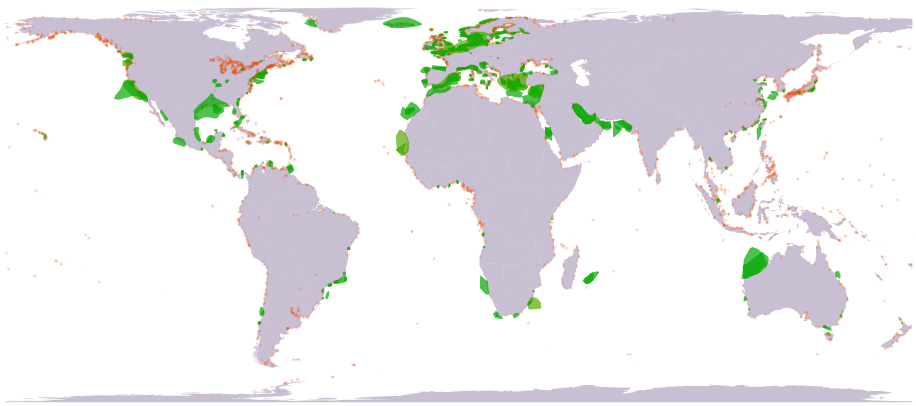

Figure D.2: Approximate AIS coverage (green), global ports (red).

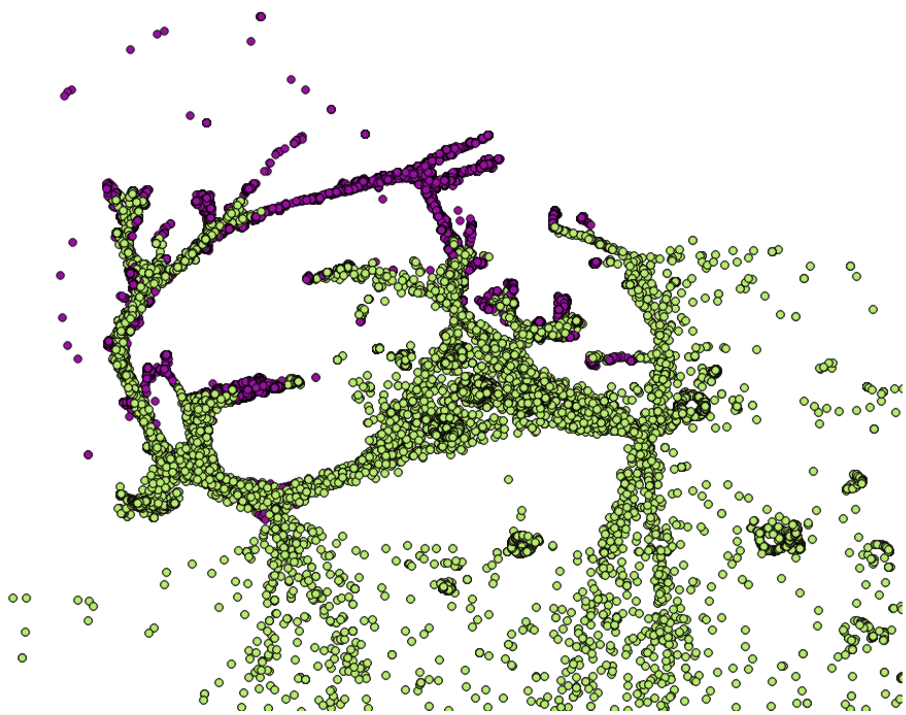

Figure D.3: Long Beach Harbor, California. Points shown in purple are 'on land', but most of these observations are actually within the harbor.

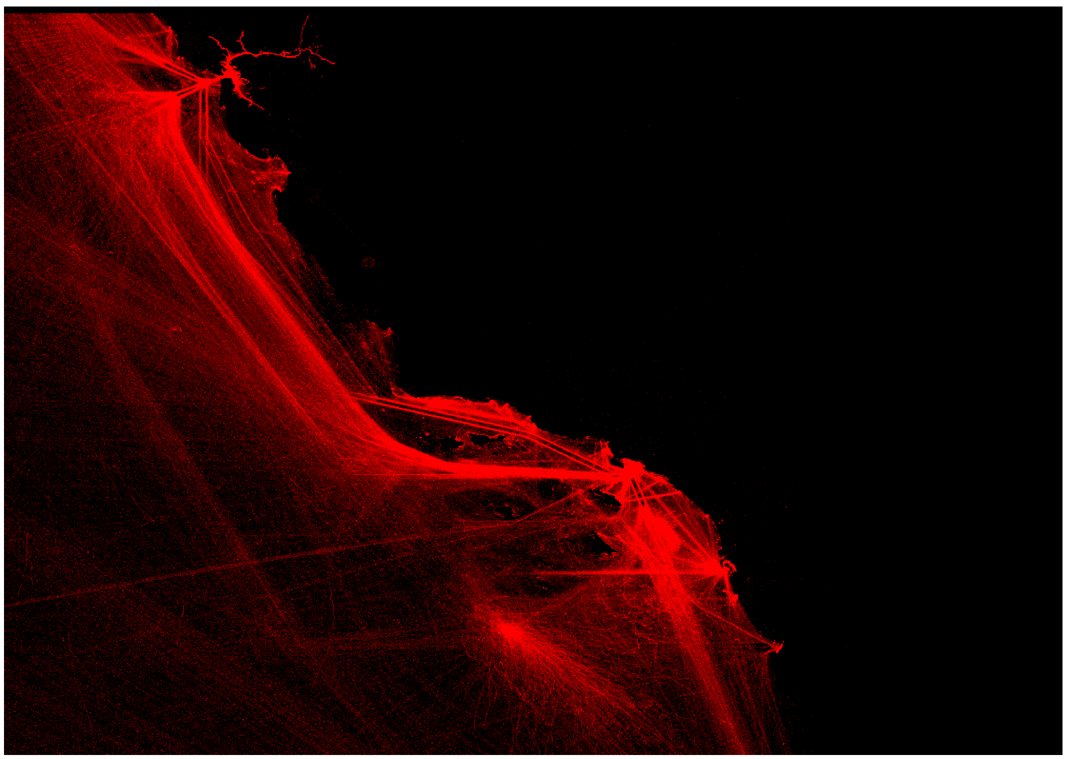

Figure D.4: AIS observations, Southern California Bight. Nov 2010–Dec 2011.

Note the ballast water exchange point lower left.

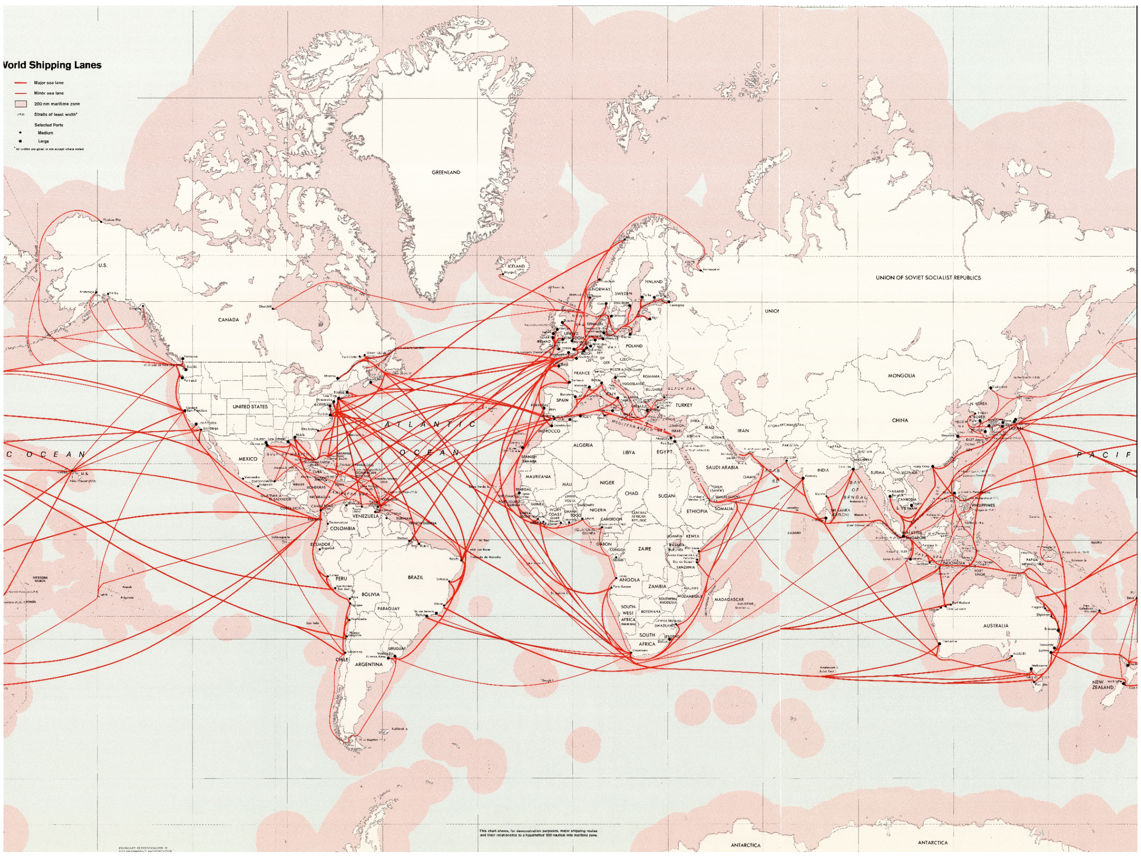

Figure D.5: "World Shipping Lanes" map produced by the Central Intelligence Agency, 1973.

Bibliography

Bivand R (2011). "Introduction to representing spatial objects in R."

Boersma P, Parrish JK (1999). "Limiting abuse: marine protected areas, a limited solution." *Ecological Economics*, **31**(2), 287–304. doi:10.1016/S0921-8009(99)00085-3.

Bowman AW, Azzalini A (1997). *Applied Smoothing Techniques for Data Analysis*. Oxford University Press, Oxford.

Bresenham J (1965). "Algorithm for computer control of a digital plotter." *IBM Systems journal*, **4**(1), 25–30.

Brohan P, Allan R, Freeman J, Waple A, Wheeler D, Wilkinson C, Woodruff S (2009). "Marine observations of old weather." *Bulletin of the American Meteorological Society*, **90**(2), 219–230.

Burrough P, Frank A, Foundation ES (1996). *Geographic Objects With Indeterminate Boundaries*. GISDATA series. CRC Press.

Canning D, Fay M (1993). "The effects of transportation networks on economic growth."

Carson-Jackson J (2012). "Satellite AIS–Developing Technology or Existing Capability?" *Journal of Navigation*, **65**(02), 303–321.

Christen P (2012). *Data matching concepts and techniques for record linkage, entity resolution, and duplicate detection*. Springer, Berlin; New York.

Cogan J (2009). "Representation and Power in International Organization: The Operational Constitution and Its Critics." *American Journal of International Law*, **103**, 209.

Cohen W, Ravikumar P, Fienberg S, *et al.* (2003). "A comparison of string distance metrics for name-matching tasks." In *Proceedings of the IJCAI-2003 Workshop on Information Integration on the Web (IIWeb-03)*, pp. 73–78.

Corbett JJ, Winebrake JJ, Green EH, Kasibhatla P, Eyring V, Lauer A (2007). "Mortality from ship emissions: a global assessment." *Environmental science & technology*, **41**(24), 8512–8.

Costello C, Ovando D, Hilborn R, Gaines S, Deschenes O, Lester S (2012). "Status and Solutions for the Worlds Unassessed Fisheries." *Science*, **338**(6106), 517–520.

Currier KE, Goodchild MF, Li L, Hurt I, Mooney P (2012). "The Social Construction of Feature Types in Volunteered Geographic Information." Unpublished Draft.

De Smith M, Goodchild M, Longley P (2007). *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*. Troubador Publishing.

Ducruet C, Notteboom T (2012). "The worldwide maritime network of container shipping: Spatial structure and regional dynamics." *Global Networks*.

Elwood S, Goodchild M, Sui D (2011). "Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice."

Equasis (2012). "The world merchant fleet in 2011." *Technical report*, Equasis. URL `http://www.equasis.org/EquasisWeb/restricted/PublicStatistic?fs=ShipSearch`.

Eyring V, Isaksen IS, Berntsen T, Collins WJ, Corbett JJ, Endresen O, Grainger RG, Moldanova J, Schlager H, Stevenson DS (2009). "Transport impacts on atmosphere and climate: Shipping." *Atmospheric Environment*, (x), 1–37. ISSN 13522310. doi:10.1016/j.atmosenv.2009.04.059.

Federal Geographic Data Committee (2008). "Geographic Information Framework Data Content Standard Part 0: Base Document. FGDC-STD-014.0-2008."

Fletcher J (2010). "JCOMM Voluntary Observing Ship Scheme." URL `http://www.bom.gov.au/jcomm/vos/vos.html`.

Food and Agriculture Organization (2012). "Vishing Vessels." URL `http://www.fao.org/fishery/topic/1616/en`.

Freeman JE (2010). "NCDC: Volunteer Observing Ship Climate Project." URL `http://www.ncdc.noaa.gov/oa/climate/vosclim/vosclim.html`.

Goodchild M (1992). "Geographical information science." *International Journal of Geographical Information Systems*, **6**(1), 31–45.

Goodchild M (1999). "Cartographic futures on a digital earth." *The Map Reader*, pp. 141–146.

Goodchild M (2007). "Citizens as sensors: the world of volunteered geography." *GeoJournal*, **69**(4), 211–221.

Goodchild M, Li L (2012). "Assuring the quality of volunteered geographic information." *Spatial Statistics*, pp. 110–120.

Guimera R, Mossa S, Turtschi A, Amaral L (2005). "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles." *Proceedings of the National Academy of Sciences*, **102**(22), 7794–7799.

Halpern B, Warner R (2002). "Marine reserves have rapid and lasting effects." *Ecology letters*, **5**(3), 361–366.

Halpern BS, Longo C, Hardy D, McLeod KL, Samhouri JF, Katona SK, Kleisner K, Lester SE, O'Leary J, Ranelletti M, Rosenberg Aa, Scarborough C, Selig ER, Best BD, Brumbaugh DR, Chapin FS, Crowder LB, Daly KL, Doney SC, Elfes C, Fogarty MJ, Gaines SD, Jacobsen KI, Karrer LB, Leslie HM, Neeley E, Pauly D, Polasky S, Ris B, St Martin K, Stone GS, Sumaila UR, Zeller D (2012). "An index to assess the health and benefits of the global ocean." *Nature*, **488**(7413), 615–20. doi:10.1038/nature11397.

Halpern BS, Walbridge S, Selkoe KA, Kappel CV, Micheli F, D'Agrosa C, Bruno JF, Casey KS, Ebert C, Fox HE, Fujita R, Heinemann D, Lenihan HS, Madin EMP, Perry MT, Selig ER, Spalding M, Steneck R, Watson R (2008). "A global map of human impact on marine ecosystems." *Science New York NY*, **319**(5865), 948–952.

Hatch L, Fristrup K (2009). "No barrier at the boundaries: implementing regional frameworks for noise management in protected natural areas." *Marine Ecology Progress Series*, **395**, 223–244. doi:10.3354/meps07945.

International Association of Marine Aids to Navigation and Lighthouse Authorities (2004). "1028 On the Automatic Identification System (AIS) Volume 1, Part I, Operational Issues."

International Maritime Organization (1980). "International convention for the Safety of Life at Sea (SOLAS), Chapter V Safety of Naviation, Regulation 19."

International Maritime Organization (2009). "Prevention of air pollution from ships second IMO GHG study 2009." *Technical report*, International Maritime Organization.

International Telecommunications Union (2010). "Recommendation ITU-R M.1371-4." *Technical report*.

Jensen A, Silber G, Calambokidis J (2004). *Large whale ship strike database*. US Department of Commerce, National Oceanic and Atmospheric Administration.

Jones MT (2012). "Keynote: 2012 Joint Warfighting Conference." `http://www.youtube.com/watch?v=Dk3IIOimGwE`.

Jurczyk P, Lu J, Xiong L, Cragan J, Correa A (2008). "FRIL: A tool for comparative record linkage." In *AMIA Annual Symposium Proceedings*, volume 2008, p. 440. American Medical Informatics Association.

Kaluza P, Kölzsch A, Gastner MT, Blasius B (2010). "The complex network of global cargo ship movements." *Journal of the Royal Society, Interface / the Royal Society*, **7**(January), 1093–1103. doi:10.1098/rsif.2009.0495.

Karney C (2012). "Algorithms for geodesics." *Journal of Geodesy*, pp. 1–13.

Keller RP, Drake JM, Drew MB, Lodge DM (2011). "Linking environmental conditions and ship movements to estimate invasive species transport across the global shipping network." *Diversity and Distributions*, **17**(1), 93–102. doi:10.1111/j.1472-4642.2010.00696.x.

Lack D, Cappa C, Langridge J, Bahreini R, Buffaloe G, Brock C, Cerully K, Coffman D, Hayden K, Holloway J, *et al.* (2011). "Impact of Fuel Quality Regulation and Speed Reductions on Shipping Emissions: Implications for Climate and Air Quality." *Environmental science & technology*, **45**(20), 9052–9060.

Lagueux K, Zani M, Knowlton A, Kraus S (2011). "Response by vessel operators to protection measures for right whales Eubalaena glacialis in the southeast US calving ground." *Endangered Species Research*, **14**(1), 69–77. doi:10.3354/esr00335.

Leppäkoski E, Gollasch S, Gruszka P, Ojaveer H, Olenin S, Panov V (2002). "The Baltic a sea of invaders." *Canadian Journal of Fisheries and Aquatic Sciences*, **59**(7), 1175–1188.

Lloyd's Register-Fairplay Ltd (2010). "World shipping encyclopaedia." URL `http://www.ihsfairplay.com/Maritime_data/WSE/WSE.html?product=WSE`.

Lubchenco J, Petes LE (2010). "The interconnected Biosphere: Science at the oceans tipping Points." *Oceanography*, **23**(2), 115.

McKenna MF, Katz SL, Condit C, Walbridge S (2012). "Response of Commercial Ships to a Voluntary Speed Reduction Measure: Are Voluntary Strategies Adequate for Mitigating Ship-Strike Risk." *Coastal Management*, **40**(October), 634–650.

Merrifield M, McClintock W, Burt C, Fox E, Serpa P, Steinback C, Gleason M (2012). "MarineMap: a web-based platform for collaborative marine protected area planning." *Ocean & Coastal Management*.

Miller H, Bridwell S (2008). "A field-based theory for time geography." *Annals of the Association of American Geographers*, **99**(1), 49–75.

Miller H, Han J (2009). *Geographic data mining and knowledge discovery*. CRC.

Moore MJ, McLellan WA, Daoust PY, Bonde RK, Knowlton AR (2007). *Right whale mortality: a message from the dead to the living*, pp. 358–379. Harvard University Press.

Mueller H (2011). "Sailwx.info." URL `http://www.sailwx.info`.

National Geospatial-Intelligence Agency (2012). *World Port Index*. Twenty second edition. National Geospatial-Intelligence Agency.

Notteboom T (2004). "Container shipping and ports: an overview." *Review of network economics*, **3**(2).

Open Geospatial Consortium Inc (2008). *OGC KML 07-147r2*, 2.2.0 edition.

Open Geospatial Consortium Inc (2010). "Simple features specification for SQL."

Organisation for Economic Co-operation and Development (2010). *Globalisation, Transport and the Environment*. Organisation for Economic Co-operation and Development, Paris. ISBN 9789264079199.

PostgreSQL Global Development Group (2012). "PostgreSQL 8.4.14." `http://www.postgresql.org`.

Rabus B, Eineder M, Roth A, Bamler R (2003). "The shuttle radar topography missiona new class of digital elevation models acquired by spaceborne radar." *ISPRS Journal of Photogrammetry and Remote Sensing*, **57**(4), 241–262.

Ramsey P, *et al.* (2005). *PostGIS manual*.

Raymond ES, Schwehr K (2012). "The Toils of AIS: A Case Study in Application Protocol Design and Analysis." Unpublished Draft, URL `https://gitorious.org/toils-of-ais/toils-of-ais`.

Rodrigue J, Comtois C, Slack B (2009). *The geography of transport systems*. 2nd edition. Taylor & Francis.

Ruiz G, Rawlings T, Dobbs F, Drake L, Mullady T, Huq A, Colwell R (2000). "Global spread of microorganisms by ships." *Nature*, **408**(6808), 49–50. doi:10.1038/35040695.

Schwehr KD, Foulkes JA, Lorenzini D, Kanawati M (2011). "Global Coastal and Marine Spatial Planning (CMSP) from Space Based AIS Ship Tracking." *AGU Fall Meeting Abstracts*, p. B1278.

Slater J, Garvey G, Johnston C, Haase J, Heady B, Kroenung G, Little J (2006). "The SRTM data finishing process and products." *Photogrammetric engineering and remote sensing*, **72**(3), 237–247.

Tetreault B (2002). "Use of the Automatic Identification System (AIS) for Maritime Domain Awareness (MDA)." *Proceedings of OCEANS 2005 MTS/IEEE*, pp. 1–5. doi:10.1109/OCEANS.2005.1639983.

Tilman D, Kareiva PM (1997). *Spatial ecology. The role of space in population dynamics and interspecific interactions*. Princeton University Press, Princeton & Oxford, USA.

Univeristy of the Aegean (2010). "Live Ships Map - AIS." URL `http://www.marinetraffic.com/ais/`.

Vanderlaan A, Corbett J, Green S, Callahan J, Wang C, Kenney R, Taggart C, Firestone J (2009). "Probability and mitigation of vessel encounters with North Atlantic right whales." *Endangered Species Research*, **6**(March), 273–285. doi:10.3354/esr00176.

Vries G, Malaisé V, Someren M, Adriaans P, Schreiber G (2009). "Semi-automatic ontology extension in the maritime domain." In *Proceedings of the twentieth Belgian-Dutch Conference on Artificial Intelligence*, pp. 265–272. University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.

Wang C, Corbett J, Firestone J (2007). "Modeling energy use and emissions from North American shipping: application of the ship traffic, energy, and environment model." *Environmental science & technology*, **41**(9), 3226–32.

Wessel P, Smith W (1996). "A global, self-consistent, hierarchical, high-resolution shoreline database." *Journal of Geophysical Research*, **101**, 8741–8743.

Westwood J, Parsons B, Rowley W (2001). "Global ocean markets." *Oceanography*, **14**(3), 83.

Winkler W (1990). "String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage." Proceedings of the Section on Survey Research Methods (American Statistical Association).

Woodruff S, Worley S, Lubker S, Ji Z, Eric Freeman J, Berry D, Brohan P, Kent E, Reynolds R, Smith S, *et al.* (2010). "ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive." *International Journal of Climatology*, **31**(7), 951–967.

Worm B, Hilborn R, Baum J, Branch T, Collie J, Costello C, Fogarty M, Fulton E, Hutchings J, Jennings S, *et al.* (2009). "Rebuilding global fisheries." *science*, **325**(5940), 578–585.

Wright D, Goodchild MF (1997). "Data from the deep: Implications for the GIS community." *International Journal of Geographical Information Science*, **11**(6), 523–528.

Wu Y, Miller H (2002). "Computational tools for measuring space-time accessibility within transportation networks with dynamic flow." *Journal of Transportation and Statistics*, **4**(2/3), 1–14.