A big data driven approach to extracting global trade patterns

Giannis Spiliopoulos[*], Dimitrios Zissis[* †] and Konstantinos Chatzikokolakis[*]
*MarineTraffic, London, United Kingdom
Email:{giannis.spiliopoulos,konstantinos.chatzikokolakis}@marinetraffic.com
[†] Department of Product & Systems Design Engineering, University of the Aegean, Greece
Email:dzissis@aegean.gr

Abstract
Unlike roads, shipping lanes are not carved in stone. Their size, boundaries and content vary over space and time, under the influence of trade and carrier patterns, but also infrastructure investments, climate change, political developments and other complex events. Today we only have a vague understanding of the specific routes vessels follow when travelling between ports, which is an essential metric for calculating any valid maritime statistics and indicators (e.g trade indicators, emissions and others). Whilst in the past though, maritime surveillance had suffered from a lack of data, current tracking technology has transformed the problem into one of an overabundance of information, as huge amounts of vessel tracking data are slowly becoming available, mostly due to the Automatic Identification System (AIS). Due to the volume of this data, traditional data mining and machine learning approaches are challenged when called upon to decipher the complexity of these environments. In this work, our aim is to transform billions of records of spatiotemporal (AIS) data into information for understanding the patterns of global trade by adopting distributed processing approaches. We describe a four-step approach, which is based on the MapReduce paradigm, and demonstrate its validity in real world conditions.

Keywords: Big Spatiotempotal Data, AIS, global shipping routes, K-Means Clustering, Apache Spark

1 Introduction

Since ancient times, trade has been conducted mostly by sea. Captains of their times sought out the safest, but also fastest routes connecting major trading sea ports. As early as 515BC, the sailor Scylax, made the first recording of the Mediterranean voyages or sailing instructions, which described safe passages between Mediterranean ports (later known as Hellenic Periploi*),* listing ports and coastal landmarks with approximate distances and routes between them. Throughout history, vessels have regularly set their shipping courses so as to take advantage of the prevailing winds and ocean currents, leading to the definition of major shipping trade routes, which are mostly in use until today. Such popular trade routes include the routes crossing the Pacific Ocean, the Atlantic Ocean routes and the Indian ocean routes, but unlike roads, these shipping routes are not carved in stone. The size of the corridor, its content and connections, can vary greatly over space and time, under the influence of trade and carrier patterns, but also due to infrastructure investments, climate change, political events and other complex international events. For instance, global warming is having a major effect on shipping routes; as new routes such as the Arctic Ocean shipping route north of Russia (which has cut thousands of miles off the journey from China to the European ports) are emerging [1], while icebergs are disrupting the traditional shipping lanes off the Canadian coast [2]. Similarly, investments in port terminals or canal expansions have widespread effects on routes and trading patterns; such as the recent expansion of the Panama Canal, which influenced a range of stakeholders regarding shipping rates, peripheral port capacity and port investments.

The importance of a well-developed understanding of the maritime traffic patterns and trade routes is critical to all seafarers and stakeholders. From a security perspective, it is necessary for understanding areas of high congestion, so that smaller vessels can avoid collisions with bigger ships. Moreover, an understanding of vessel patterns at scale can assist in the identification of anomalous behaviors and help predict the future location of vessels. Additionally, by combining ship routes with a model to estimate the emission of vessels (which depends on travel distance, speed, draught, weather conditions and characteristics of the vessel itself), emissions of e.g. CO_2 and NOx can be estimated per ship and per national territory [3]. From an economic side, stakeholders selecting to deploy a ship on a particular route need to find the optimal mix between a number of variables such as the shortest path between two ports, cost of route, expected congestion, travel time, size of vessel and capacity and many more. According to the findings of a recent Eurostat funded project, current problems in methods of calculating official maritime statistics include, (i) Distance travelled per ship is now based on an inaccurate average distance matrix for ports, (ii) Missing Information on travel routes for goods to estimate unit prices for transit trade statistics [3].

While in the past, maritime surveillance had suffered from a lack of data, current tracking technology has transformed the problem into one of an overabundance of information. Progressively huge amounts of structured and unstructured data, tracking vessels during their voyages across the seas, are becoming available, mostly due to the Automatic Identification System (AIS) that vessels of specific categories are required to carry. The AIS is a collaborative, self-reporting system that allows marine vessels to broadcast their information to nearby vessels and on-ground base stations. Vessels equipped with AIS transceivers periodically broadcast messages that include the vessel identifying information, characteristics, and destination together with other information coming from on-board equipment, such as current location, speed, and heading. AIS data slowly becoming available, provides almost global coverage as data collection methods are not restricted to a single country or continent, providing an opportunity for in depth analysis of patterns at a global scale which was previously unavailable [3].

However, current Information & Communication Technology (ICT) and traditional data mining approaches are challenged when called upon to decipher the complexity of these environments and produce actionable intelligence. AIS geospatial data-sets are very large in size, containing billions of records, and skewed, as specific regions, can contain substantially more data than others, making processing and storage with conventional methods highly challenging. As such traditional techniques and technologies have proven incapable of dealing with such volumes of loosely structured spatio-temporal data.

In our approach, we exploit a massive volume of historical AIS data to estimate trade routes in a data-driven way, with no reliance on external sources of information. We present a four phased approach which is based on the MapReduce distributed programming paradigm, and demonstrate its effectiveness and validity in real world conditions.

Our work presents novelties on three fronts:

- Distributed computation: We present an architectural prototype which is validated by efficiently processing billions of AIS message (>500Gb) within a few hours. To the best of our knowledge no previous work has successfully analyzed AIS datasets of this size and coverage within the time scale of our solution (less than 3 hours).
- Algorithmic Accuracy: We discuss our algorithmic approach to generating accurate trade routes by overcoming many of the known accuracy issues of AIS in a distributed fashion by adopting a MapReduce approach.
- Domain Specific: We uncover global maritime trade routes which can be used as a method of anomaly detection, investigation but also understanding and predicting variations in trade patterns and the effect of events.

The rest of the manuscript is organized as follows: Section 2 shortly presents previous work in this domain, while Section 3 describes our approach and Section 4 presents the preliminary results. Finally, section 5 presents the conclusion and briefly outlines shortcoming of this work and future improvements.

2 Related Work

AIS data has been used as valid method for extracting valuable information regarding vessel behavior, operational patterns and performance statistics for a number of years now. As Tichavska, Cabrera, Tovar and Arana point out, AIS data has been used for a variety of applications including, optimization of radio propagation channel techniques, real-time statistical processing of traffic information, improving ship traffic management and operations, sustainable transport solutions and many more [4]. Specifically, for route definition and motion pattern extraction, AIS is considered a valid source of data, used as a framework for trajectory forecasting and anomaly detection. Most published works can be categorized per the methods the authors follow which are either (i) grid based or (ii) methods of using vectorial representations of traffic. In (i) grid based approaches, the area of coverage is split into cells which are characterized by the motion properties of the crossing vessels to create a spatial grid . In the second category, vessel trajectories are modeled as a set of connected waypoints. Thus, vessel motions in large areas (e.g., at a global scale) can be managed thanks to the high compactness of the waypoint representation [5][6].

Towards this direction, in their recent work, Ristic, La Scala, Morelande and Gordon perform statistical analysis of vessel motion patterns to extract motion patterns which are then used to construct the corresponding motion anomaly detectors using adaptive kernel density estimation [7]. Mazzarella, Arguedas, & Vespe apply a Bayesian vessel prediction algorithm based on a Particle Filter (PF) on AIS data [8]. Zhang, Goerlandt, Kujala, Wang, & Nikitakos apply hierarchical and other clustering methods to learn the typical vessel sailing pattern within the waters of Xiamen Bay and Chengsanjiao, China [9]. Pallotta, Vespe and Bryan, present the TREAD (Traffic Route Extraction and Anomaly Detection) methodology, which relies on the DBSCAN algorithm for automatically detecting anomalies and projecting current trajectories and patterns into the future[10].

As the amount of available AIS data grows to massive scales though, researchers are realising that computational techniques must also contend with acquiring, storing, and processing the data. Applying traditional techniques to AIS data processing can lead to processing times of several days, if applied to global data sets of considerable size. In addition to this, many traditional approaches assume that the underlying data distribution is uniform and spatially continuous. This is not the case for global AIS data, as it is often to have large geographical coverage gaps, message collisions or erroneous messages especially when processing large areas [11,12]. This problem is mostly evident when dealing with extended geographical areas and "big" datasets. In their majority, previous research efforts have focused on limited geographical areas (e.g. a specific coastal area or sea port) and smaller datasets (e.g. several thousands of AIS messages/ GBs) [13–15], often overstepping the problems AIS data quality altogether. In their work [5], authors present a two-step method to achieve a balance between computational time and performance; first performing data simplification by applying the Douglas-Peucker(DP) algorithm before processing the simplified trajectories with Kernel Density Estimation.

Grid-based methods have been considered effective only for small area surveillance and the computational burden was regarded as its limitation when increasing the scale[10]. Therefore a "vectorial" representation of traffic was proposed [16] to allow implementation at a global scale, including waypoint objects and route objects. However, these methods had only been implemented in limited geographical areas (e.g. a 200 × 160 km area in the North Adriatic Sea and similar), and limited information was given about performance [16,17]. However, in their work Lin Wu, Yongjun Xu, Qi Wang, Fei Wang and Zhiwei Xu, demonstrate the ability of a grid-based method for computing shipping density, fast enough to be performed at a global scale (less than 56 hours). In this work it took 56 hours to produce all the global monthly ship density, traffic density and AIS receiving frequency maps, from August 2012 to April 2015; 33 months of data[17].

To date, very few works, apply the advancements that have been made on the big data front to AIS data processing. In 2008, the MapReduce programming approach was described by Google engineers, in which data-parallel computations are executed on clusters of unreliable machines by systems that automatically provide locality-aware scheduling, fault tolerance, and load balancing while and shortly after in 2011, the Hadoop implementation by Yahoo's engineers was made publicly available under an Apache License [10][18]. Although certainly not a panacea[19], Hadoop introduced millions of programmers and scientists to parallel and distributed computation, starting the "big data wave". In their work [20],Wang et al. attempt to tackle the big data issue caused by the AIS data for anomaly detection purposes. They implement a two-step process, where they firstly use an unsupervised technique, based upon the Density-Based Spatial Clustering of Applications with Noise considering Speed and Direction (DBSCAN-SD) incorporating non- spatial attributes, such as speed and direction, to label normal and abnormal position points of vessels based on the raw AIS data. Secondly, they train a supervised learning algorithm designed with the MapReduce paradigm running on Hadoop using the labelled data generated in from the first step. The authors support that the distributed approach is capable of outperforming the promise of traditional GIS applications.

Following Hadoop's success numerous frameworks and open source tools appear on the Big Data ecosystem. Apache Spark, originally designed by researchers at the University of Berkeley, was developed in response to limitations in the MapReduce/Hadoop cluster computing paradigm, which forces a particular linear dataflow structure on distributed programs (acyclic data flow model). Many iterative machine learning algorithms, as well as interactive data analysis tools, reuse a working set of data across multiple parallel operations. Spark processes data in-memory and has been shown to be capable of outperforming Hadoop by 10x in iterative machine learning workloads[18]. In our previous work, we confirmed the potential benefits of applying such techniques to large AIS data processing [21,22]. In [21,23] we presented an adaptation of the well-known KDE algorithm to the map-reduce paradigm to estimate a seaports extended area of operation from AIS data. This work is complementary, in that it addresses the problem of estimating global trade routes in an adaptive, scalable, and unsupervised way, based on k-means clustering applied in a distributed fashion. Similarly, [24] Salmon and Ray present their work on designing a hybrid approach based on the Lambda architecture [19] for both real-time and archived data approaches to processing maritime traffic data.

3 Approach

As described in the previous sections, the aim of this work is to calculate the global trade routes from large amounts of AIS data. Out of the 64 different types of AIS messages that can be broadcast by AIS transceivers (as defined by the ITU 1371-4 standard), our work focuses on the 6 most relevant ones, which account for approximately 90% of AIS typical scenarios [19]. Types 1, 2, 3, 18, and 19 are position reports, which include latitude, longitude, speed-over-ground (SOG), course-over-ground (COG), and other fields related to ship movement; type 5 messages contain static-and voyage information, which includes the IMO identifier, radio call sign, name, ship dimensions, ship and cargo types. In all messages, each vessel is identified by its Marine Mobile Service Identifier (MMSI) number. Data is received through the MarineTraffic system and for the purposes of this work we use a dataset of approximately 5 billion messages (i.e., 525 GB) recorded from January to December 2016 (Table 1). In the rest of this section we present the methodology followed to transform the raw data collected into meaningful information and thus useful for data analysis. We provide a detailed analysis of each step of the approach and explain thoroughly the effect our actions had to the considered dataset.

Dataset Statistics	
Time Period	January-December 2016
Positions count	>5 Billion
Port calls count	>3 Million
Number of unique vessels	>200K
Ports covered in dataset	>3K

Table 1. Original Dataset Statistics

3.1.1 Distributed Processing

For the distributed processing tasks, we rely on a HDInsight Azure Spark (2.1.0 version) cluster made up by: 6 worker nodes (D4v2 Azure nodes), each one equipped with 8 processing cores and 28 GB RAM; and 2 head nodes (D12 v2 Azure nodes), each one equipped with 4 processing cores and 28 GB RAM, summing up to a total of 56 computing cores and 224 GB RAM. This setup has been sufficient to cover the processing requirements of all the complex computations of our methodology.

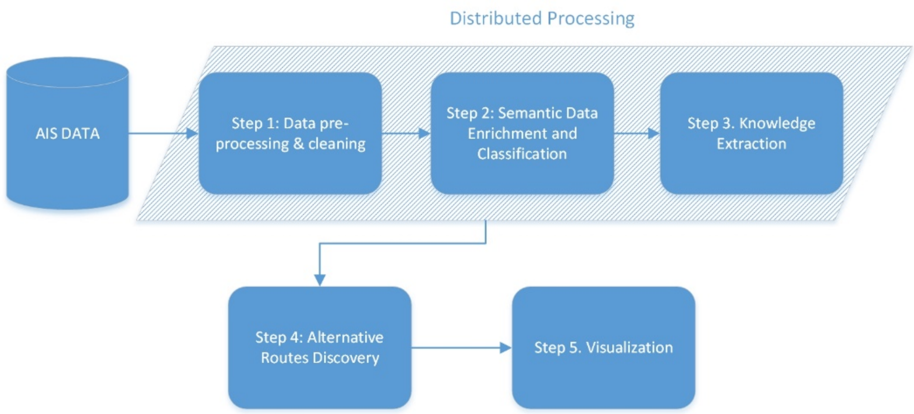

Figure 1. Four step distributed approach

3.1.2 Step 1: Trajectory Data Cleaning And Preprocessing

AIS messages do not provide any trustworthy voyage information with respect to departure and destination ports. In fact, the only relevant data collected from AIS are

type 5 messages which include the destination port information. However, this is manually given by the ship's crew and prone to errors, inconsistencies etc. Thus, it is fundamental to discover such knowledge (i.e., departure and destination ports) from the AIS positional data to perform route analysis. We algorithmically, calculate port destination and port departure for each AIS message on spark. The purpose of such correlation is to assign accurately the departure port, the destination port based on the timestamp of each AIS message and the departure time and arrival time respectively.
In addition to route assignment, for each message we compute the time elapsed (measured in minutes) since the vessel's departure based on the reported timestamps (i.e. from AIS messages). The time elapsed field enables us to include the time dimension in our analysis on AIS messages and group them into time-aligned routes with the same <departure, destination> port and vessel type.

We focus our work only on AIS messages originating from cargo and tanker vessels, as we are not interested in smaller vessels such as fishing boats, tugs, etc. that do not follow a repetitive pattern and may not be representative of global trade routes. Similarly, for this study, we filter out AIS messages originating from passenger vessels as these follow different patterns e.g. visiting different ports than cargo vessels and messages with recorded speed less than 0.5 knots which is considered as the lower bound for vessels moving underway using their engines. The resulting dataset of this preprocessing steps results in approximately 1 billion enriched AIS messages and approximately 28K distinct vessels.

3.1.3 Step 2: Semantic Data Enrichment and Classification (Distributed- Map Phase):

The preprocessing step uses a dataset of approximately $2*10^4$ ports resulting in a large number of port-to-port combinations (i.e., practically almost any port may be connected with any other port in the world resulting in $4*10^8$ combinations), which are further increased as we take into account also the ship-type. However, it makes sense to apply any algorithmic approach on data on each route separately. Thus, in order to take advantage of the parallelization ability of spark we map each enriched message to a key-value pair. The key uniquely identifies the route per vessel type and it is generated as the concatenated unique identifier of departure – destination pair and the vessel type, while the value is the enriched message itself.

3.1.4 Step 3: Knowledge Extraction from Trajectories (Distributed- Reduce phase):

Up to this point we have enriched each record of our dataset with additional voyage related information. To prepare our data for further processing, we organize all records on lists based on the key defined in the map phase. This is performed by a reduce-by-key procedure and produces a set of key-valued pairs organized in a set of rows equal to the distinct number of keys (i.e., 368473 unique routes per ship type). Each reduced set contains on average 2000 points per key, which can be further processed by a single node. Given that at this time, spark does not support nested map-reduce processes, we select to process multiple routes simultaneously by distributing their keys to multiple nodes, instead of splitting each set computation to multiple nodes. In order to capture a unique route for all the points included in each set we perform the well-known k-means clustering technique using WEKA, an open source software collection of machine learning algorithms for data mining tasks. The features selected for the k-means clustering are the following:

- latitude,
- longitude,
- relative timestamp.

This enables our solution to detect clusters of vessels' positions based on both location and time with respect to the departure timestamp, enhancing route perception with average elapsed time (or within a time range) from departure for vessels of the same type with the same route that sail in close by trajectories.
The number of clusters per route K is selected dynamically using the formula below.

$$K = \max\left(\min\left(\left(\frac{N}{10}\right), cmax\right), 1\right)$$

where,

- cmax is the maximum number of clusters (set to 100 for our evaluation),
- N is the #points per route.

The arbitrary selection of the maximum number of centroids cmax is equal to 100. The cmax selection is based on the degree of compression of information that we would like to achieve, while preserving some of its original granularity for the global dataset. Finally, before moving to the next step, we assign to each centroid the number of distinct vessels that have at least one point within the cluster.

3.1.5 Step 4: Discovering Alternative Routes

During the previous steps, it was assumed that same vessel types sailing on the same route will follow similar trajectories. However, this is not always accurate, as various factors such as weather conditions, draught, etc. may vastly differentiate the vessel's course. In such cases we observed that the outcome of the clustering phase suffered from fluctuating events and the produced trajectories had continuous changes of vessel's course. In most cases the actual data indicated the existence of (at least) two different courses for the same route. To identify these courses, we applied a trajectory splitting algorithm based on the following predefined set of rules
We examine all route points using a three points sliding observation window. Each point is assigned to a different course if any of the empirical rules evaluated in the following algorithm is valid.
For each course of a route detected do the following:
For each three sequential points p_{i-1}, p_i, p_{i+1} evaluate the following rules:

1. $\#vessels\ in\ p_{i-1} = 1\ ||\ \#vessels\ in\ p_i = 1$
2. $D_{i-1,i} + D_{i,i+1} > 2 * D_{i-1,i+1}$
3. $D_{i-1,i} > D_{th}$

4. $T_{i-1,i} > T_{th}$
5. $V_{i-1,i} > V_{th}$

where,

- $D_{i,j}$ is the distance between p_i, p_j
- $T_{i,j}$ is the time variance between p_i, p_j
- $V_{i,j}$ is the vessel's speed so as to reach p_j from p_i
- D_{th} is the distance threshold set to 1000km, which is a relaxed constraint for positional data received from two consecutive AIS messages.
- T_{th} is the time threshold that equals to 24h, which is a relaxed threshold compared to the AIS message received frequency.
- V_{th} is the speed threshold, which is set to 40 knots, as cargo ships and tankers typically have speed less than 35 knots.

Although multiple sailing courses are linked with the same route, they are completely different paths from departure port to destination port, and thus, when visualizing results we treat each sailing course as a different route. In the following section we analyze the evaluation results of our methodology and provide insights on some interesting new routes such as China to east USA.

4 Results

In this section, we present some preliminary results of the approach described in the previous section. All results presented below have been produced using the cluster setup presented in section 3.1.1 and the execution times correspond to the results of the spark processing with 24 executors having two cores each and kryo serializer option enabled to minimize the serialization cost.

As previously discussed in Section 3 our approach is a four stepped approach. In the preprocess step, we parsed the original dataset of 5 billion records and filtered out the messages that do not originate from cargo or tanker vessels based on the AIS vessel type recorded. Then, positional data were combined with recorded departures and destinations to create a smaller dataset with enriched data of 0.79 billion records. Processing efforts were split into 5 jobs, 12 stages and 3727 tasks, having a total computational cost of 1.3 hours. The map phase was executed through 1 job, 1 stage and 267 tasks and lasted almost 10 minutes. The reduce phase including the splitting process was executed in 1 jobs, 1 stage and 267 tasks and had a 10 minutes computational cost. The entire process from the initial import of raw data to end results extraction lasted approximately two hours.

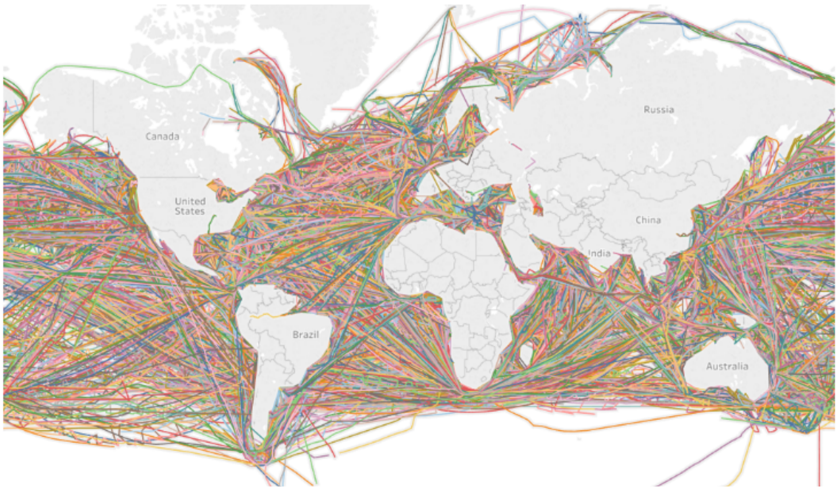

Figure 2. Routes extracted for cargo vessels for 2016.

The resulting dataset was stored in a single 850MB csv file including information for all the vessels. The total number of routes detected, after splitting was 440,854 represented with 10,847,328 points (Figure 2). The result represents a significant part of information in terms of routes detected, i.e. the number of routes detected is greater than the distinct number of routes per vessel type, while the compression rate is 73:1 in terms of positions. It should be noted that since the dataset used is recorded from January to December 2016, some vessel routes may start or end in the middle of the sea because the ship's voyage is not entirely in this period (i.e. ship's voyage may start before 2016, or may end in 2017).

Through our approach, we have been able to control the degree of compression of the resulting dataset based on the selection of K, being capable to create visualizations of sets of global routes. Regarding the domain, it is interesting to view that we can validate most popular trade routes with accuracy and new routes are uncovered, such as the artic route above Russia and new routes that pass over North America to link east USA with China have been detected (Figure 3). Finally, the addition of time in the clustering process for K-means introduces an average voyage time estimation on each centroid detected, this estimate could be used a baseline for future works on optimal route selection for the maritime industry.

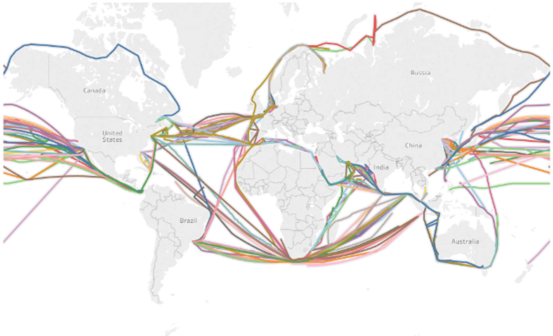

Figure 3. Overview of the routes connection some of the top 50 ports around the globe.

5 Conclusion and Future Work

This article focused on the challenges of analyzing huge amounts of vessel tracking data produced through the AIS. The novelty of the method is in the direction of adopting a map reduce approach to distribute the computational burden across a cluster of commodity machines to perform the computation in approximately 2 hours' time. Preliminary results presented in the previous section confirm the validity of the adopted approach. Future work, will be focused on improvements of the algorithmic approach to improve the accuracy of the identified routes.

6 Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 732310 and by Microsoft Research through a Microsoft Azure for Research Award

7 References

[1] B. Savadove, China Begins Using Arctic Shipping Route That Could "Change The Face Of World Trade," Bus. Insid. (2013).

[2] Euronews, Arctic icebergs reach Canadian coast, disrupting sea lanes and fishing, Euronews. (2017).

[3] Tessa de Wit, A. Consten, M. Puts, C. Pierrakou, M. Bis, A. Bilska, et al., ESSnet Big Data Deliverable 4.2 Deriving port visits and linking data from Maritime statistics with AIS-data, 2017.

[4] F. Cabrera, N. Molina, M. Tichavska, V. Arana, Design of a low cost prototype of automatic identification system (AIS) receiver, in: 2015 1st URSI Atl. Radio Sci. Conf. (URSI AT-RASC), IEEE, 2015: pp. 1–1. doi:10.1109/URSI-AT-RASC.2015.7303000.

[5] Y. Li, R.W. Liu, J. Liu, Y. Huang, B. Hu, K. Wang, Trajectory compression-guided visualization of spatio-temporal AIS vessel density, in: 2016 8th Int. Conf. Wirel.

Commun. Signal Process., IEEE, 2016: pp. 1–5. doi:10.1109/WCSP.2016.7752733.

[6] M. Fiorini, A. Capata, D.D. Bloisi, AIS Data Visualization for Maritime Spatial Planning (MSP), Int. J. E-Navigation Marit. Econ. 5 (2016) 45–60. doi:10.1016/j.enavi.2016.12.004.

[7] N.G. B. Ristic, B. La Scala, M. Morelande, Statistical analysis of motion patterns in ais data: Anomaly detection and motion prediction, IEEE, 2008.

[8] F. Mazzarella, V.F. Arguedas, M. Vespe, Knowledge-based vessel position prediction using historical AIS data, in: 2015 Sens. Data Fusion Trends, Solut. Appl., IEEE, 2015: pp. 1–6. doi:10.1109/SDF.2015.7347707.

[9] W. Zhang, F. Goerlandt, P. Kujala, Y. Wang, N. Nikitakos, An advanced method for detecting possible near miss ship collisions from AIS data, Ocean Eng. 124 (2016) 141–156. doi:10.1016/j.oceaneng.2016.07.059.

[10] G. Pallotta, M. Vespe, K. Bryan, Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction, Entropy. 15 (2013) 2218–2245. doi:10.3390/e15062218.

[11] J. Poļevskis, M. Krastiņš, G. Korāts, A. Skorodumovs, J. Trokšs, Methods for Processing and Interpretation of AIS Signals Corrupted by Noise and Packet Collisions, Latv. J. Phys. Tech. Sci. 49 (2012) 25–31. doi:10.2478/v10047-012-0015-3.

[12] M. Yang, Y. Zou, L. Fang, Collision and Detection Performance with Three Overlap Signal Collisions in Space-Based AIS Reception, in: 2012 IEEE 11th Int. Conf. Trust. Secur. Priv. Comput. Commun., IEEE, 2012: pp. 1641–1648. doi:10.1109/TrustCom.2012.109.

[13] N. Willems, H. van de Wetering, J.J. van Wijk, Visualization of vessel movements, Comput. Graph. Forum. 28 (2009) 959–966. doi:10.1111/j.1467-8659.2009.01440.x.

[14] N. Willems, H. van de Wetering, J.J. van Wijk, Evaluation of the Visibility of Vessel Movement Features in Trajectory Visualizations, Comput. Graph. Forum. 30 (2011) 801–810. doi:10.1111/j.1467-8659.2011.01929.x.

[15] G. Di Battista, J.-D. Fekete, H. Qu, IEEE Computer Society., Institute of Electrical and Electronics Engineers., IEEE Computer Society. Technical Committee on Visualization and Graphics., IEEE Pacific Visualization Symposium 2011 : proceedings : Hong Kong, China, March 1-4, 2011, IEEE, 2011.

[16] V. Arguedas, M. Vespe, G. Pallotta, Automatic generation of geographical networks for maritime traffic surveillance, in: Inf. Fusion (FUSION), 2014 17th Int. Conf., IEEE, 2014.

[17] L. Wu, Y. Xu, Q. Wang, F. Wang, Z. Xu, Mapping Global Shipping Density from AIS Data, J. Navig. 70 (2017) 67–81. doi:10.1017/S0373463316000345.

[18] M. Zaharia, M. Chowdhury, M.J. Franklin, S. Shenker, I. Stoica, Spark: Cluster computing with working sets, (2010).

[19] M. Stonebraker, Hadoop at a Crossroads?, n.d.

[20] X. Wang, X. Liu, B. Liu, E.N. de Souza, S. Matwin, Vessel route anomaly detection with Hadoop MapReduce, in: 2014 IEEE Int. Conf. Big Data (Big Data), IEEE, 2014: pp. 25–30. doi:10.1109/BigData.2014.7004464.

[21] L.M. Millefiori, D. Zissis, L. Cazzanti, G. Arcieri, A distributed approach to estimating sea port operational regions from lots of AIS data, in: 2016 IEEE Int. Conf. Big Data (Big Data), IEEE, 2016: pp. 1627–1632. doi:10.1109/BigData.2016.7840774.

[22] L. Millefiori, D. Zissis, L. Cazzanti, G. Arcieri, Computational Maritime Situational Awareness Techniques for Unsupervised Port Area, Nato Unclassified Reports, Science And Technology Organization Centre For Maritime Research And Experimentation, La Spezia, Italy, 2016.

[23] L.M. Millefiori, D. Zissis, L. Cazzanti, G. Arcieri, Scalable and Distributed Sea Port Operational Areas Estimation from AIS Data, in: 2016 IEEE 16th Int. Conf. Data Min. Work., IEEE, 2016: pp. 374–381. doi:10.1109/ICDMW.2016.0060.

[24] L. Salmon, C. Ray, Design principles of a stream-based framework for mobility analysis, Geoinformatica. 21 (2017) 237–261. doi:10.1007/s10707-016-0256-z.