



Navigation Decision Support: Discover of Vessel Traffic Anomaly According to the Historic Marine Data

A. Daranda, G. Dzemyda

Andrius Daranda*

VU Institute of Data Science and Digital Technologies
Vilnius University, Lithuania

*Corresponding author: andrius.daranda@gmail.com

Gintautas Dzemyda

VU Institute of Data Science and Digital Technologies
Vilnius University, Lithuania
gintautas.dzemyda@mif.vu.lt

Abstract

During the last years, marine traffic dramatically increases. Marine traffic safety highly depends on the mariner's decisions and particular situations. The watch officer must continuously observe the marine traffic for anomalies because the anomaly detection is crucial to predict dangerous situations and to make a decision in time for safe marine navigation. In this paper, we present marine traffic anomaly detection by the combination of the DBSCAN clustering algorithm (Density-Based Spatial Clustering of Applications with Noise) with k-nearest neighbors analysis among the clusters and particular vessels. The clustering algorithm is applied to the historic marine traffic data – a set of vessel turn points. In our experiments, the total number of turn points was about 3 million, and about 160 megabytes of computer store was used. A formal numerical criterion to compare anomaly with normal traffic flow case has been proposed. It gives us a possibility to detect the vessels outside the typical traffic pattern. The proposed method ensures the right decisions in different oceanic scale or hydro meteorology conditions in the detection of anomaly situation of the vessel.

Keywords: marine anomaly detection, marine traffic, spatial data, DBSCAN, clustering, k-nearest neighbors, regression.

1 Introduction

Marine traffic generates multidimensional and dynamic navigation data. The enormous amount of marine traffic data creates significant challenges to detect anomalies for situation awareness. Due to data complexity and specificity, unique methods are necessary for proper processing and resolving marine navigation data in different contexts. Usually, the human is an essential part of the marine industry despite excessive mechanization and technical advancements. The human in the marine field plays a vital role in controlling vessels and making decisions to ensure vessel safety.

However, many studies show that the main reason for marine accidents is human error [8], [2], [3], [15]. Many accidents happen because the vessel watch officer decides to take an unconventional route. The best-known example is the Costa Concordia disaster. The vessel ran aground and overturned after hitting an underwater rock. Also, human error could be caused by overworking, wrong managerial decisions, insufficient knowledge, lack of maintenance of standards, etc. [10]. Therefore, it is crucial to detect marine traffic anomaly and alert the watch officer about it as soon as possible. So, the early detection of such abnormalities is the essential key to ensure the safety of marine traffic, security, and protection of the environment.

2 The marine traffic

It is essential to get full navigational information about marine traffic. Therefore, the Maritime Safety Committee (MSC) of the International Maritime Organization (IMO) adopted and required the use of the Automatic Identification System (AIS) in the vessels [20]. This system ensures the exchange of data among vessels, AIS base stations, and satellites. The AIS highly enhances the safety and efficiency of marine traffic and complements situational awareness and assessment [17], [1], [5]. However, due to marine traffic complexity, it is not easy to decide about the navigational situation without a full analysis of the situation. The detection of anomaly navigation situation is highly required for all marine traffic participants [13]. However, the use of AIS alone is not enough [13]. The anomaly navigation situation could appear not only due to human error but also it may be a reason of piracy or robbery at sea, mi-grant smuggling, and organized crime within the fishing, oil spilling, and oil bunkering [13], [16]. Usually, such activities are trying to hide or mask by fake vessel AIS data [4], [14].

The AIS data could be divided into kinematic (vessel position – latitude & longitude, speed (SOG), course (COG, HDG), rate of turn (ROT)) and static (vessel name, maritime mobile service identity (MMSI), vessel type, dimension, the port of destination) information. AIS broadcasts the navigational data every 2-10 seconds (depending on vessel speed) [12] from every vessel in the vicinity throughout Self-Organized Time Division Multiple Access (STDMA) [9] vessel-to-vessel datalink. This broadcast creates an enormous volume of spatial data. For example, the vessels near Danish coasts (including Greenland) generate about 60 GB data every month. In this case, if we want to process one-year data, we have to deal with 720 GB.

Consequently, the maritime authorities need methods to detect and prevent illegal activities. Likewise, the watch officer in the vessel should be warned about unusual activity near passing vessels. The abnormal vessel behavior should take notice and take appropriate action in time to avoid collision or disaster.

Many techniques have been proposed for marine anomaly detection. All methods can be divided into three main categories [11]: supervised, unsupervised, and semi-supervised anomaly detection. Another way to separate anomaly detection is short-term and long-term. The short-term anomaly may be detected in coastal waters, e.g., using self-organizing map application for processing of AIS data [18]. Long-term marine anomaly detection is necessary for open seas or oceans [19].

3 DBSCAN algorithm for clustering the turn points

This paper proposes the method of detecting the potential anomaly situations based on the Density-based spatial clustering of applications with noise (DBSCAN) clustering algorithm [6] applied to the historic AIS marine data. Partitioning methods (k -means, PAM clustering) are suitable only for compact and well-separated clusters, but they are not designed to find concave clusters that are non-linearly separable. The results of these methods are negatively affected by the noise in the data and outliers. DBSCAN algorithm separates noise data. Moreover, unlike the k -means algorithm, DBSCAN itself computes the number of the clusters. DBSCAN algorithm was created for the clustering of the large volume of spatial data.

DBSCAN algorithm has two control parameters: a radius of the neighborhood with respect to some point (ϵ), and the minimum number of neighbors (data points) within radius ϵ (MinPts). DBSCAN

classifies data points into three groups: core points, reachable points, and noise or outliers. A point q is a core point if at least minPts points are within distance (ϵ) from it, including this point itself. A part of reachable points is reachable directly. A point q is directly reachable from p if point p is within distance ϵ from core point q . Points of the cluster may be reachable directly from core points, only. A point q is reachable from p if there is a path of points p_1, \dots, p_n with $p_1 = p$ and $p_n = q$, where each p_{i+1} is directly reachable from p_i . Note that all points on the path are the core points, with the possible exception of q . All points that are not reachable from any other point are outliers or noise points. Each cluster contains at least one core point.

In order to use the DBSCAN algorithm for anomaly detection, the vessel positions (latitude & longitude of the vessel) were extracted from historic marine data. The historic marine traffic data allows getting the experience of vessel navigation to the particular port of destination.

The vessel turn points only were analyzed and clustered by DBSCAN. They are filtered from the whole data. It is assumed that the turning point is the position where the vessel changes course for more than 4 degrees. Moreover, the use of a turning point is useful for filtering the vessel yaw and pitch motion. The turn points to the same port of destination are different for every vessel. It depends on the planned route, different navigational, and hydro meteorology conditions. Likewise, all vessels usually go slightly differently for the same planned trajectory every time. However, there are regions of turning, and the clustering may be applied to disclose these regions. To imagine the volume of data, e. g. in our research, the turn point data set contain information about 1350 vessels whose destination port was Rotterdam. The total number of turn points was about 3 million, and about 160 megabytes of computer store was used. As a result, we have obtained 450 clusters.

Figure 1 shows an example of the clustered turn points by DBSCAN and the general trajectory of vessels.

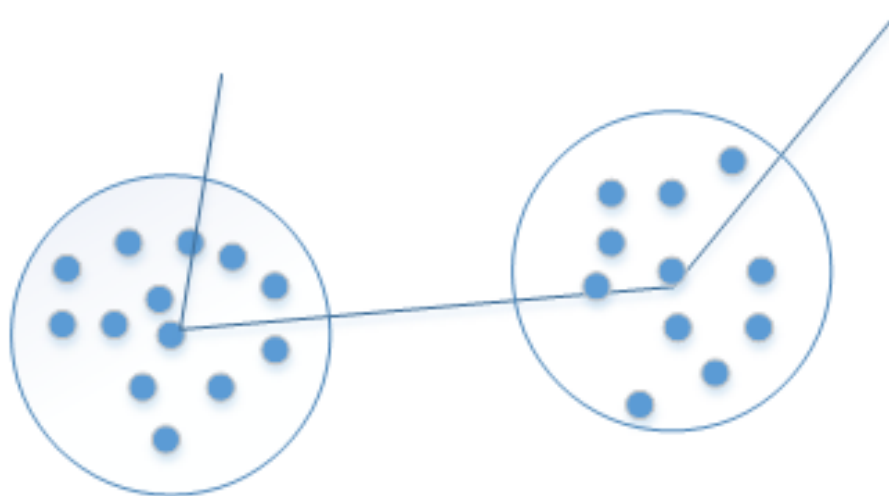


Figure 1: DBSCAN algorithm applied to for clustering the turn points

The calculated clusters must be filtered by removing turning points that are outliers, defined by DBSCAN. The clustered turn points make the recognized path to the port of destination, as shown in Figure 2. However, sometimes DBSCAN forms the out-laying clusters which do not belong to the typical path to the port of destination. The reason for such outlying clusters may be e.g. the mistake in setting the port of destination in the AIS. Such clusters and their groups are put in boxes in Figure 2. They are removed from further analysis because we need the data about proper marine traffic flow to the ports of destination.

Further, the spatial data should be filtered by the port of the destination. All vessels find a different path to the same port of destination, but the moderate trail can be easily recognized. The extracted remaining clusters of turning points reflect the navigation experience and regular marine traffic flow. This experience depends on the port of destination but could be used from any initial port of the voyage if this route is on the way.

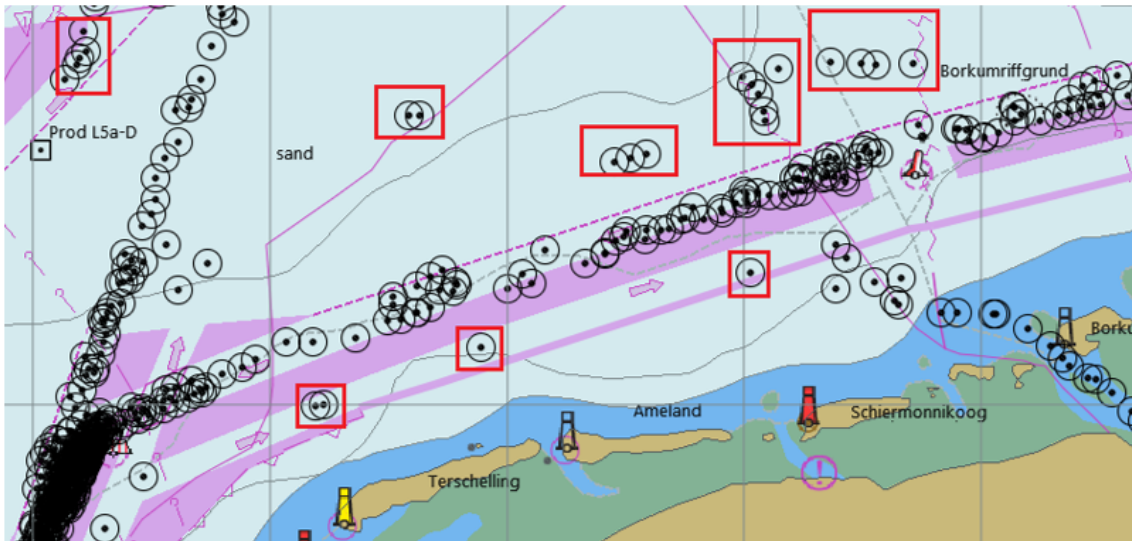


Figure 2: The result of DBSCAN algorithm

4 Anomaly detection

As a result of the analysis and filtering above, we get the marine traffic to some particular port of destination. The example of the clustered traffic flow is shown in Figure 3. Points denote the centers of turning point clusters. Then such centers are numbered and marked by symbol \times . These points are chosen in the set at random for further analysis. Let us consider a particular vessel. The vessel that is not in the normal clustered flow (e.g. somewhere outside the traffic flow that is given in Figure 3 could be classified as suspicious. There are various reasons for being not in the normal flow. For example, the navigator forgot to change the port of destination in the AIS settings. However, being not in the marine traffic flow could mean illegal fishing, smuggling, environmental pollution, or other criminal activity.

As an example, let us analyze the same marine traffic flow with some anomaly, see Figure 3. There are four cases of marine traffic anomaly. These cases are depending on the distance to the normal traffic flow. In this case, we assume that the outlier vessel proceeds to the same port of destination as the marine traffic flow. The reason for such an assumption may be grounded by information from the AIS.

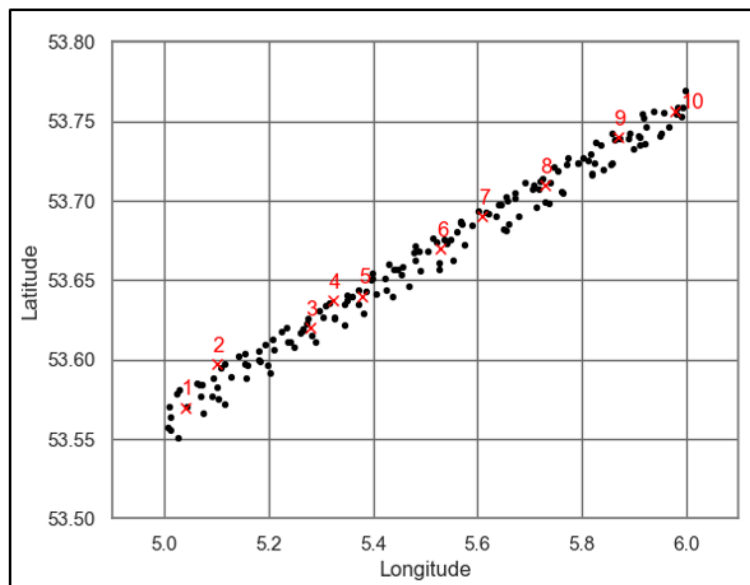


Figure 3: The clustered marine traffic flow – the centers of turning point clusters

Let us analyze the data in Figure 3 more in detail using k -nearest neighbors comparison. For each point, we find k -nearest points and compute the average distance among this point and the remaining k points. Afterward, the results are averaged for all points of the traffic flow. As a result, we get one average value for a fixed k .

In Figure 5, we present an average distance among points for different $k = 1, \dots, 8$. The average distance increases depending on the number of neighbors. Some regression may approximate this dependence. Linear, second-order polynomial and logarithmic functions were examined. In fact, we see the linear dependence with an R -squared statistical measure near 1.

The k -nearest neighbors from $k = 1$ to 8 for each of 14 marked points from Figure 3 and Figure 4, were calculated and dependencies on k of the average distance were determined. We noticed differences among the dependencies on the average distance for the particular point from the normal traffic flow and the anomaly cases. The examples of dependencies of the average distance among the neighboring points are presented in Figure 6 and Figure 7.

Table 1 summarizes the results: variances of estimates, when k runs from 1 to 8, i.e., each variance was evaluated using eight numbers. In the anomaly cases, the variances are much smaller than in the normal traffic flow. Moreover, Table 1 leads to the criteria to detect the vessels outside the typical marine traffic pattern:

1. the distance between the abnormally acting vessel and the nearest neighboring point from the normal traffic flow obtained by DBSCAN is much bigger than the average distance among the points of the normal traffic flow;
2. the variance of the average distance between neighboring points obtained for $k = 1, \dots, 8$ is much higher for the vessels from the normal traffic flow as compared with the abnormally acting vessel;
3. the variance decreases with growing the distance between the vessel and the normal traffic flow.

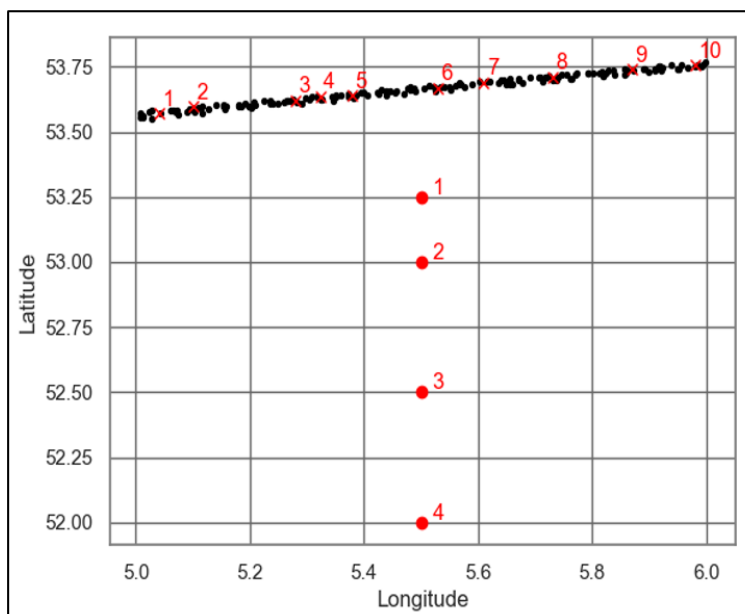


Figure 4: The marine traffic flow with anomaly cases

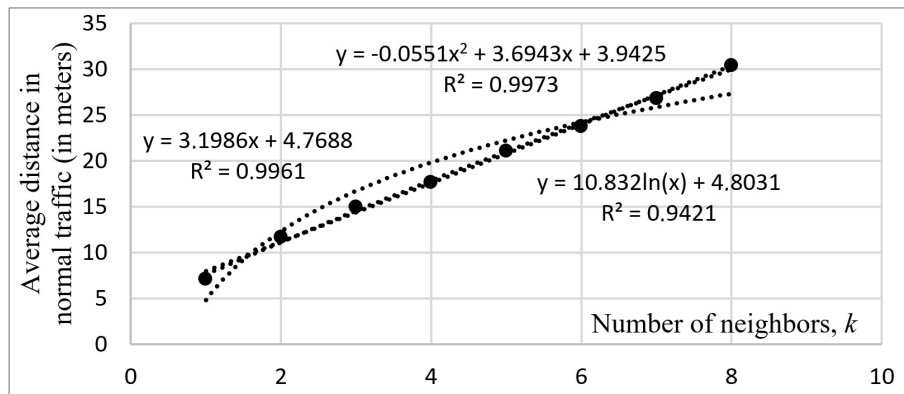


Figure 5: The average distance among the neighboring points in normal traffic in dependence on k

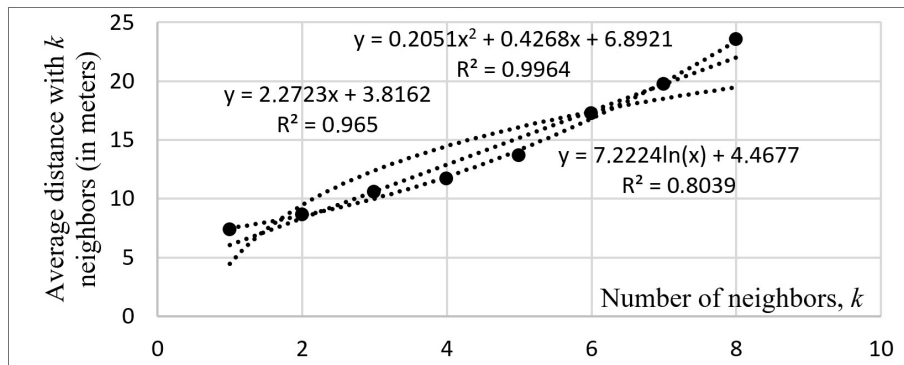


Figure 6: The average distance among the neighboring points in normal traffic in dependence on k for case No. 7

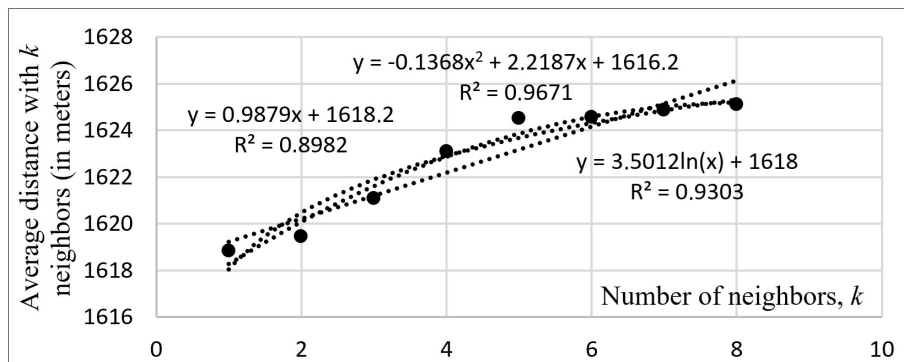


Figure 7: The average distance among the anomaly case No. 4 and the neighboring points in normal traffic in dependence on k

Table 1: Variances of estimates

Points		Variance
Points in traffic	1	72.71
	2	52.83
	3	25.49
	4	52.51
	5	38.72
	6	28.09
	7	139.36
	8	22.72
	9	70.82
	10	84.57
Points out of traffic	1	15.65
	2	10.16
	3	8.00
	4	5.70
Average for normal traffic		53.92

5 Navigation decision support

The traditional navigational methods like Radar and AIS ensure getting navigational data, but humans always must make a decision. The proposed anomaly detection could be used in the navigation decision support. There are different groups of users for such a decision support system (DSS): particular vessel, marine authorities that control marine traffics in narrow channels and ports, rescue services, environmental monitoring, and law enforcement authorities. Our method is fast because we need to apply the DBSCAN to a large amount of spatial data once. Further calculations are based on sufficiently small data containing coordinates of turn point clusters that are results of clustering by DBSCAN. This data is sufficient to determine whether the particular vessel is the outlier of the normal marine traffic. Depending on the DSS, the warning should be sent to the officer-in-charge.

Besides, we may need the real-time coordinates and ports of destinations of particular vessels that are in the area of interest if the DSS is devoted to control the particular vessel and to suggest the proper decision for its navigation officer. These coordinates come from the Automatic Identification System (AIS). In this case, anomaly detection could be a part of the technical solution in DSS.

6 Conclusion

The anomaly detection is crucial to predict dangerous situations and to make a decision in time for safe marine navigation. In this paper, we present marine traffic anomaly detection by the combination of DBSCAN clustering algorithm (Density-Based Spatial Clustering of Applications with Noise) with k-nearest neighbors analysis applied to the vessel turning point data. The data extracted from the historic marine traffic data serves as the reference point when evaluating the abnormal behavior of the vessel. A formal numerical criterion to compare anomaly with normal marine traffic flow case has been proposed. It gives us a possibility to detect the vessels out-side the typical traffic pattern.

This method could be beneficial for modeling marine traffic flow. Besides, this method could be used to detect marine traffic anomaly on many vessels at once and in real-time. Also, the proposed method is useful in different hydro meteorology conditions. Therefore, the presented approach provides a solution for marine surveillance and marine traffic situational awareness. Our experiments show the possibility of processing a large amount of spatial data (up to several terabytes). As a result, a sufficiently small set of cluster centers was aggregated using DBSCAN for further analysis and decisions. Further research should be addressed to the evaluation of further behavior of the vessel and its attempts to return to the normal marine traffic.

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare no conflict of interest.

References

- [1] Brusch, S.; Lehner, S.; Fritz, T.; Soccorsi, M.; Soloviev, A.; Van Schie, B. (2011). Ship surveillance with TerraSAR-X, *IEEE Transactions on Geoscience and Remote Sensing*, 49, 1092–1103, 2011.
- [2] Celik, M.; Cebi, S. (2009). Analytical HFACS for investigating human errors in shipping accidents, *Accident Analysis and Prevention*, 41, 66–75, 2009.
- [3] Chen, S.T.; Wall, A.; Davies, P.; Yang, Z.; Wang, J., Chou, Y.H. (2013). A Human and Organisational Factors (HOFs) analysis method for marine casualties using HFACS-Maritime Accidents (HFACS-MA), *Safety Science*, 60, 105–114 (2013).
- [4] Daranda, A.; Andziulis, J.S. (2015). Fake vessels identification in the AIS, *In: Transport Means 2015 Proceedings*, 248–252, 2015.
- [5] Eriksen, T.; Høye, G.; Narheim, B.; Meland, B.J. (2016). Maritime traffic monitoring using a space-based AIS receiver, *In: International Astronautical Federation - 55th International Astronautical Congress*, 5276–5289, 2004.
- [6] Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise, *In: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, 226–231, 1996.
- [7] Fournier, M.; Casey Hilliard, R.; Rezaee, S.; Pelot, R. (2018). Past, present, and future of the satellite-based automatic identification system: areas of applications (2004–2016), *WMU Journal of Maritime Affairs*, 17, 311–345, 2018.
- [8] Fujii, Y.; Shiobara, R. (1971). The analysis of traffic accidents, *Journal of Navigation*, 24, 534–543, 1971.
- [9] Gaugel, T.; Mittag, J.; Hartenstein, H. et al. (2019). In-depth analysis and evaluation of Self-organizing TDMA, *In: 2013 IEEE Vehicular Networking Conference, VNC, Boston*, 79–86, 2013.
- [10] Goerlandt, F.; Goite, H.; Valdez Banda, O.A.; Höglund, A.; Ahonen-Rainio, P.; Lensu, M. (2017). An analysis of wintertime navigational accidents in the Northern Baltic Sea, *Safety Science*, 92, 66–84, 2017.
- [11] Hodge, V.J.; Austin, J. (2018). A survey of outlier detection methodologies, *Artificial Intelligence Review*, 22, 85–126, 2004.
- [12] International Telecommunication Union (ITU) (2014). *Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile frequency band M Series Mobile, radiodetermination, amateur and related satellite*, 2014.
- [13] Jin, M.; Shi, W.; Lin, K.C.; Li, K.X. (2019). Marine piracy prediction and prevention: Policy implications, *Marine Policy*, 108, 2019.
- [14] Longépé, N.; Hajduch, G.; Ardianto, R. et al. (2018). Completing fishing monitoring with spaceborne Vessel Detection System (VDS) and Automatic Identification System (AIS) to assess illegal fishing in Indonesia, *Marine Pollution Bulletin*, 131, 33–39, 2018.

- [15] Mazaheri, A.; Montewka, J.; Kujala, P. (2013). Correlation between the ship grounding accident and the ship traffic – A case study based on the statistics of the Gulf of Finland, *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, 7, 119–124, 2013.
- [16] Prabowo, A.R.; Bae, D.M. (2019). Environmental risk of maritime territory subjected to accidental phenomena: Correlation of oil spill and ship grounding in the Exxon Valdez’s case, *Results in Engineering*, 4, 100035, 2019.
- [17] Tsou, M.C. (2016). Online analysis process on Automatic Identification System data warehouse for application in vessel traffic service, *Proceedings of the Institution of Mechanical Engineers Part M: Journal of Engineering for the Maritime Environment*, 230, 199–215, 2016.
- [18] Venskus, J.; Treigys, P.; Bernatavičienė, J.; Tamulevičius, G.; Medvedev, V. (2019). Real-time maritime traffic anomaly detection based on sensors and history data embedding, *Sensors (Switzerland)*, 19, 3782, 2019.
- [19] Wang, Y.; Han, L.; Liu, W.; Yang, S.; Gao, Y. (2019). Study on wavelet neural network based anomaly detection in ocean observing data series, *Ocean Engineering*, 186, 2019.
- [20] [Online] *IMO Resolution MSC 74 (69), Annex 3, Recommendation on Performance Standards for an Universal Shipborne Automatic Identification System (AIS)*, [http://www.imo.org/en/KnowledgeCentre/IndexofIMOResolutions/Maritime-Safety-Committee-\(MSC\)/Documents/MSC.74\(69\).pdf](http://www.imo.org/en/KnowledgeCentre/IndexofIMOResolutions/Maritime-Safety-Committee-(MSC)/Documents/MSC.74(69).pdf), Accessed on 10 December 2019.



Copyright ©2020 by the authors. Licensee Agora University, Oradea, Romania.

This is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License.

Journal’s webpage: <http://univagora.ro/jour/index.php/ijccc/>



This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).

<https://publicationethics.org/members/international-journal-computers-communications-and-control>

Cite this paper as:

Daranda, A.; Dzemyda, G. (2020). Navigation Decision Support: Discover of Vessel Traffic Anomaly According to the Historic Marine Data, *International Journal of Computers Communications & Control*, 15(3), 3864, 2020.

<https://doi.org/10.15837/ijccc.2020.3.3864>