Context information analysis from IMM filtered data classification

David Sánchez Pedroche, Daniel Amigo Herrero, Jesús García Herrero, José Manuel Molina López
Group GIAA, University Carlos III of Madrid, Spain
Email: { davsanch, damigo, jgherrer } @inf.uc3m.es, molina@ia.uc3m.es

ABSTRACT

On this paper the authors study a set of maritime trajectory classifiers that classify the type of the ship or the kind of movement that is doing. All with the objective of deciding the viability of using the variables set introduced to the algorithm and the classification result as context information useful for a maritime vigilance system that could estimate the ship trajectories with the help of those classification results. To address the classification problem a decision tree and a support vector machine algorithms have been tested and compared. The set of variables to evaluate is obtained from the movement estimation of an Interacting Multiple Model (IMM) filter and its mode probabilities changes. Through the experimentation is evaluated the viability of the problem with the proposed variables, concluding those that could be useful for the proposed problem.
Keywords: AIS, anomaly detection, classification, data mining, IMM filter, maritime surveillance, trajectory reconstruction

1. INTRODUCTION

Currently there is a large amount of technologies that allow the localization and tracking of moving targets with the objective of using the resulting values in higher level systems like the surveillance systems.
For its better operation the tracking algorithms need a parameter adjustment in which the knowledge about the operational context could involve a great improvement. Context can be defined as any information surrounding a situation that is helpful to understand it, predict its evolution but it is not part of the situation [1]. For instance, in this domain the ship category, intended type of motion, or operational rules applied in a certain area is contextual information which can be helpful to estimate and predict the ship trajectories.
This paper proposes a context information analysis in the maritime traffic field with the objective of evaluate the capability of classify the movement or the type of a ship obtaining context information about the targets that could improve their tracking.
Between all the areas in which a context analysis for tracking improvement could be used this study is going to focus into the maritime surveillance area, this area is vital to ensure the safety and security of the maritime traffic and to prevent illegal activities [2].

To this study, due to the ubiquity of AIS-equipped ships worldwide, a set of AIS data is filtered to obtain a smoothed trajectory estimation that could be used by a context learning process, that process aims to provide useful knowledge to describe the behavior of ships and estimate more precisely their positions and cinematic states.
That behavior could be useful as context information for specific estimation systems that use the context to obtain better trajectory information to the maritime surveillance systems.
The first approach of the study [3] was the utilization of a binary tree classification algorithm to classify two possible multivalue classes, the ship type and the ship maneuver movement, all with the objective of create a useful context information source to the problem by the classification results.
With this approach also is possible to study the different filter configurations and the usefulness of the different used variables.
However, the conclusions also shown the necessity of improve the classification success rates to achieve a usable context information source that will provide class information useful for a theoretical position estimation system that works specifically adapted to the classes.

To achieve that objective, in this paper a binary class (an instance is classified as the class or not as the class) analysis of the predominant classes of the previous study is made to obtain new information about the classes and the variables of the used environment.

This paper is organized as follows: In section 2 the state-of-art methods in classification of maritime vehicles tracks are analyzed. In section 3 is explained the source of the data used for the investigation and in section 4 is related the implemented system and in 5 results of the work are shown. Finally, the conclusions and perspectives for future works are presented in section 6.

2. RELATED WORK

There are numerous studies on AIS trajectory data. Specifically, in the area of maritime surveillance there are previous works addressing the search of traffic patterns to enhance Situational Awareness in maritime domain, especially to organize (cluster), reconstruct and classify trajectories, including prediction of activities and anomaly detection.

The work [4] processes AIS messages with deep learning framework (recurrent neural networks with latent variables) to address real aspects such as noisy data and irregular time- sampling for tasks of trajectory reconstruction, anomaly detection and vessel type identification.

The work in [5] has a proposal for a representation of routes as spatial grids built with AIS data to model the navigational patterns. It is extended in [6] to perform trajectory classification and anomaly detection in a system named as Traffic Route Extraction and Anomaly Detection (TREAD) based on extraction of frequent routs to classify real-time trajectories and trigger anomaly detection.

A survey of techniques proposed for mining trajectory data in multiple domains is provided in [7], focusing on data preparation, preprocessing, management and mining tasks (pattern mining, outlier detection, and trajectory classification), while a specific survey of maritime anomaly detection is provided in [8], distinguishing available data, methods, systems and user aspects. In [9], an analysis of AIS trajectory clustering is presented, with appropriate distance measures and dimensionality reduction. Aspects related with efficiency and scalability are dealt in [10], with data organization based on quad trees and modelling with Gaussian Mixtures.

Other approach is proposed in [11] and [12], where the computed dynamic parameters of a IMM filter are processed to perform segment classify missiles based on their dynamics and to divide Air Traffic Control data in homogenous segments and reconstruct the trajectories for evaluation purposes, respectively.

In [3] and in this paper, the objective is to analyze the ability to separate multiple categories of ships based on a IMM filter kinematic output and his filter information.

3. DATA SOURCE

This study needs a reliable dataset from which obtain enough information to create classifiers that could learn correctly. For this task, a fundamental element would be obtaining pre-labeled data to use supervised classification algorithms, providing to the classifier a class to learn how to classify it. Also, is required to have a measurement of the position of the vessel in each instant, allowing the IMM filter to estimate and generate new reliable information for the classifier.

The selected AIS data is a repository provided by Danish Maritime Authority [13], in which millions of raw AIS contacts detected on the coast of Denmark and surrounding areas are available every day from 2006 to the present. This source provides a practically unlimited amount of information, enough to generate a useful dataset for the objective of the study. Also, in the raw AIS plots, two possible classes for the classification process are detected: a “ship type” category that could be useful to obtain information about the ship based on data about its movement, and the “navigational status” category which is useful to classify the type of maneuver that a ship is doing in a specific moment.

4. DEVELOPED SYSTEM

The approach of this study aims to use the information of a state estimator to obtain trajectories from the sensor measurements and use them in a context analysis.

To achieve that the first step implies a strong preprocessing of the raw data, to clean it and extract only the useful information for the analysis and also preparing it as a viable input for the state estimator.

This approach uses an Interacting Multiple Model (IMM) filter as the state estimator for its well proven performance in problems of trajectory processing and smoothing by the reduction of the influence of the atypical sensor measurements.

The output of the estimation is prepared by a segmentation of trajectories to allow a comparable classification process (some trajectories have thousands of plots while others only have around one hundred) that provides new information useful for the context analysis.

4.1 Data preprocessing

The process starts with a file that contents all the plots of one day, around ten million of AIS plots. The first step consists in dividing the file in multiple files, each of one containing the information of one specific ship using the MMSI identifier.
Later, the MMSI contacts that doesn't provide useful information for these concreted objectives are discarded. These ones are contacts sent by a base station, and contacts that does not contain any of the classes for all the plots.

4.2 Track division

Once the tracks of a ship are obtained, the following process is to provide a correct input for the IMM filter. As the dataset used is filled by real information, some parts present some inconsistencies like big time gaps between two plots of the same vessel. These inconsistencies would make the filter return bad results that would compromise the classification outcomes.
Therefore, the solution for this task is to divide all the plots for every vessel into different tracks, each of one containing a minimum number of sequential plots with a maximum time gap between them. Also, this minimum whole track must be labeled with the same maneuver type to simplify the classification problem. This solution ensures that the IMM filter is capable of estimate correctly all the tracks and later be a correct and useful inputs to the classifier.

With the trajectories obtained as a result of these algorithm is possible to proceed to the next stage, where the IMM filter is going to process these trajectories to obtain some new information and to smooth the possible outliers in the position measurement.

4.3 Interacting Multiple Model (IMM)

The IMM filter is a tracking technique that allows adaptation to the target pattern of movement by the combination of the state estimation of multiple filter models into one common estimation.
On this study, the IMM filter use one mode for the representation of linear movement and another for the representation of the target maneuvers. Both implemented with an Extended Kalman filter (EKF) with the only difference of their Q matrix.

4.4 Trajectories segmentation

To allow the classification needed to the context analysis is necessary to transform the filtered trajectories into a useful input for the classification algorithm. To achieve this the first step is the segmentation of trajectories in equal size blocks to make them comparables into the classification.
An analysis of the classes values shows that there are instances with non-instanced values as "Reserved for future amendment [HSC]", "-" or "Other" that are cleaned with the minority classes with less than 0.0005% instances.
As a final step, is necessary to extract the information of each segment into data usable by the classification algorithm, using inputs that represent all the segment information and not only specific plots.
To achieve that the inputs are statistical values like average, mode, standard deviation, maximum, minimum and the 3 quartiles. All of them applied over the movement information of the segment:

- Speed of the target.
- Speed variation within the segment.
- The length of the movement.
- The heading variation.
- The time duration of the movement.

The attributes of the classifier instances would be like "the average speed of the target along the segment"
Also, as the IMM filter allows the location of maneuver movements with the analysis of mode probabilities (μ_j) a group of descriptors are defined to categorize the changes in the type of movement:

- Descriptor 1: Linear movement probability over 0.9 (the other one less than 0.1).
- Descriptor 2: Linear movement probability between 0.9 and 0.6 (the other mode between 0.1 and 0.4)
- Descriptor 3: Both probabilities between 0.6 and 0.4.
- Descriptor 4: Maneuver movement probability between 0.9 and 0.6 (the other one between 0.1 and 0.4).
- Descriptor 5: Maneuver mode probability over 0.9 (the other one less than 0.1).

With the information of the descriptor, is possible to explore the changes between type of movement in the segment by counting them, providing information about the maneuver movement model activation.
Figure 1 show the changes between descriptor with the variation of the linear movement mode probability. The information of the descriptors would be the number of measures in each descriptor (in the figure there are only two measures for the descriptor 3) and changes between descriptors (in the figure there are two changes from descriptor 2 to descriptor 1).

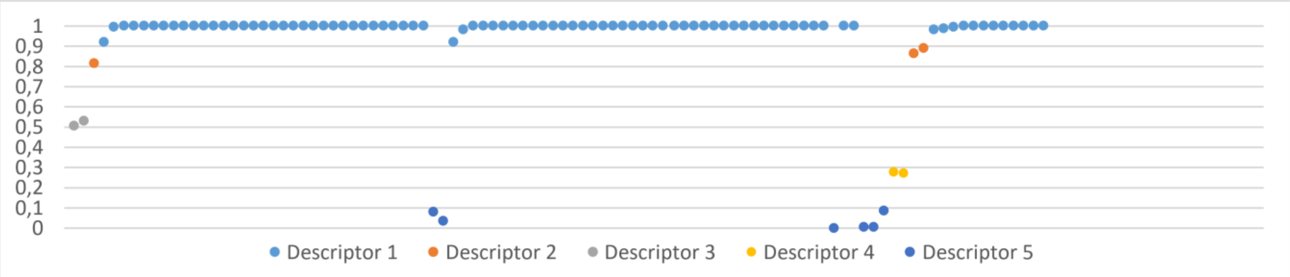

Figure 1. Descriptors over the linear movement probability

5. EXPERIMENTS AND RESULTS

In this paper four predominant classes from the ship type and the maneuver (four each one) are selected to made binary classifiers (classified as the class or not classified as the class) with a decision tree algorithm. The success rate comes from each segment (the classification results) or from the trajectory (the mode class of all the segments).

Table 1 shows the success rate of those classifiers to the ship type and maneuver classes.

Table 1. Classifiers success rates

SHIP TYPE CLASSIFIERS			MANEUVER CLASSIFIERS		
Classifier	TEST SEGMENT SUCCESS RATE	TEST TRAJECTORY SUCCESS RATE	Classifier	TEST SEGMENT SUCCESS RATE	TEST TRAJECTORY SUCCESS RATE
isCargo	73.6183 %	77.7693 %	isEngagedInFishing	75.9482 %	77.0614 %
isFishing	73.3423 %	73.4934 %	isRestrictedMovement	96.6902 %	96.6807 %
isPassenger	81.1167 %	85.1827 %	isSailing	94.8493 %	94.0119 %
isTanker	84.6864 %	86.0532 %	isUsingEngine	71.3486 %	68.5958 %

That results seems quite satisfactory in some of the classes but with the analysis of the confusion matrix is possible to observe an unbalancing problem, as the classifiers prefer to classify most of the variables as no members of each minoritarian class, obtaining false negatives instead of true positives.
As an example, is possible to compare the "*isRestricted*" class with the "*isUsingEngine*" class, the first one obtains the best results of all the classifiers whereas the second one obtains the worst results. In Table 2, the confusion matrices of the restricted movement and the is using engine classes are shown.

Table 2. Confusion matrices for example classifiers

Real Class \ Assigned Class	ISNOT	IS
isNotRestrictedMovement	113819	2424
isRestrictedMovement	1474	54
isNotUsingEngine	10082	13217
isUsingEngine	20526	73946

The results clearly show that for the restricted movement classifier is preferable to classify most of the variables as *NotRestrictedMovement*, not finding the *restrictedMovement* class (which is the objective of the classifier).

In addition, the classifier "Using Engine" (which is the one with most of the instances) find is class but with poorer results.

To balance the classifiers is possible to reduce the "ISNOT" instances to the same number of the "IS" instances, taking as an example the *isRestrictedMovement* class, the balanced classifier will work only with around 1528 instances for other classes.
Also, is necessary to ensure the presence of all the classes, avoiding taking all of the predominant class. To achieve this is possible to obtain the percentage of presence of each class and calculate their instances.
In Table 3, there are the results for the balanced classification.

Table 3. Success rates of the balanced classifiers

SHIP TYPE CLASSIFIERS			**MANEUVER CLASSIFIERS**		
Classifier	TEST SEGMENT SUCCESS RATE	TEST TRAJECTORY SUCCESS RATE	Classifier	TEST SEGMENT SUCCESS RATE	TEST TRAJECTORY SUCCESS RATE
isCargo	71.0087 %	86.2541 %	isEngagedInFishing	63.4969 %	88.5977 %
isFishing	83.1454 %	93.1127 %	isRestrictedMovement	60.5168 %	98.5556 %
isPassenger	62.5543 %	89.8125 %	isSailing	61.6474 %	97.6468 %
isTanker	66.7262 %	94.7293 %	isUsingEngine	71.5179 %	72.8047 %

Is important to remark that the trajectory success rate with the balanced classification have a lot of improvement because the balanced data have an important reduction in the number of trajectories and as a result obtain better results which are not the objective of this part of the study.
The new results have less success rate in comparison with the not balanced classifier, but as is shown in Table 4, the confusion matrices are better at the class prediction which is the objective of this classification.
In other words, with the new confusion matrix our classifier reduces the false negatives (improving the true positives that are the objective of those classifiers) but in exchange there is an increase of false positives obtaining lowest success rates.

Table 4. Balanced classifier for maneuver class

Real Class \ Assigned Class	ISNOT	IS
isNotRestrictedMovement	869	660
isRestrictedMovement	547	981

The results of these classifiers show the same conclusion of the previous multivalue classification, being a difficult task for the classification to divide the instances into different classes.
To analyze that is possible to view the attributes changes through all the classes, by the creation of some characteristic vectors for each class (calculating the average of each attribute between all the instances classified as a specific class).
With the results of the classifiers is possible to aggregate all the instances classified as a class making a representative vector which dimensions are the average of the attributes of the instances set.
Also, to ensure that all the dimensions of the vectors are equivalent, the classification is made with normalized values, making all the attributes comparable.
In Figure 2 could be seen that the vectors of the four ship type classes have the same variations with the balanced classifier, whereas in Figure 3 the results of the maneuver classifier are shown.

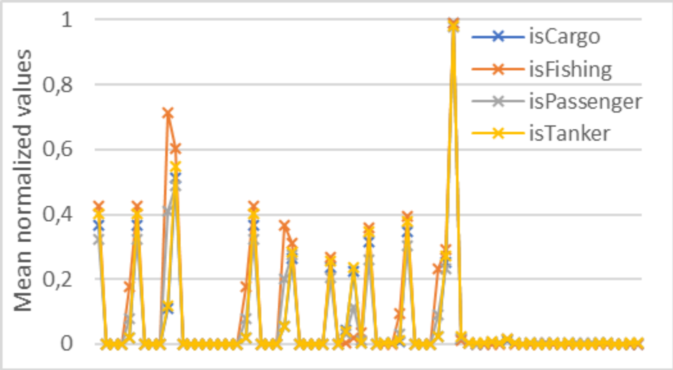

Figure 2. Ship type classes representative vectors

As it can be seen, most attributes have close to zero value and all the vectors shown really little variations between classes stating that the attributes hardly change between classes.

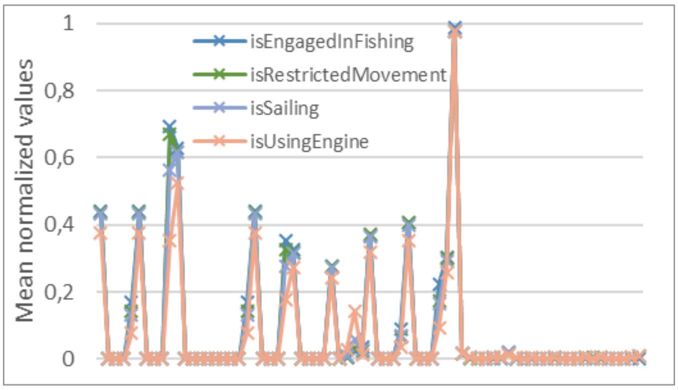

Figure 3. Maneuver classes representative vectors

An observation of the attributes shows that all of those with more variation comes from the movement variables of the direction variation or the time.
The figures from Figure 4 to Figure 7 show an example of those variables among the instances classified as the different classifiers, being possible to observe the real variation of those variables in the different classifiers.

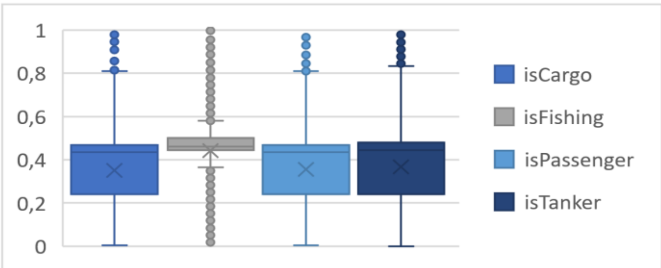

Figure 4. Ship type classes time duration

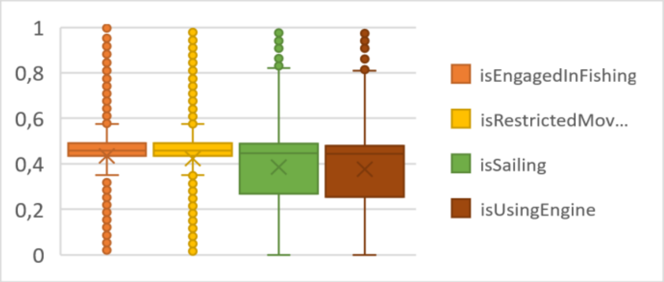

Figure 5. Maneuver classes time duration

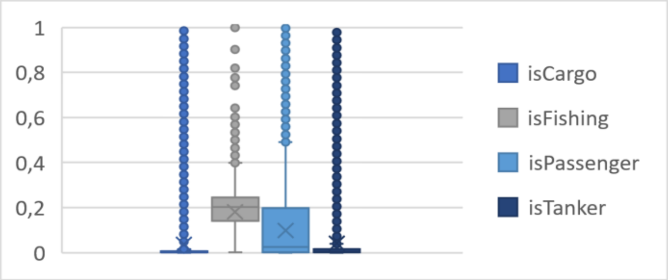

Figure 6. Ship type classes direction variation

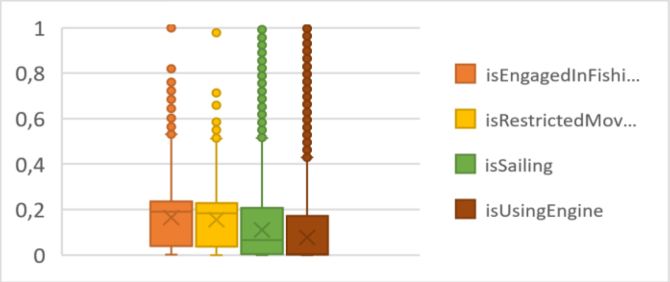

Figure 7. Maneuver classes direction variation

These results show that between these attributes (those with variation between classes the difference is minimal), the ship type classes barely differ between the *isTanker* and the *isCargo* and only the *isFishing* show enough differences.
Between the maneuver classes only the time create a difference and only with two groups whose members are also difficult to differentiate.
To a mayor improvement of the classification a new algorithm of support vector machines has been tested also as binary classifiers (classified as the class or not classified as the class), showing the results some improvement in the success rates as it can be seen in Table 5 and in Table 6.

Table 5. Support vector machines classifiers success rates

SHIP TYPE CLASSIFIERS			**MANEUVER CLASSIFIERS**		
Classifier	TEST SEGMENT SUCCESS RATE	TEST TRAJECTORY SUCCESS RATE	Classifier	TEST SEGMENT SUCCESS RATE	TEST TRAJECTORY SUCCESS RATE
isCargo	79.1626%	88.2629%	isEngagedInFishing	75.1039%	89.8508%
isFishing	83.6765%	92.6727%	isRestrictedMovement	58.0635%	98.4695%
isPassenger	62.6371%	89.1429%	isSailing	68.3383%	97.9434%
isTanker	63.7878%	91.6109%	isUsingEngine	80.3062%	81.8251%

Table 6. SVM classifier confusion matrix

Real Class \ Assigned Class	ISNOT	IS
isNotRestrictedMovement	688	841
isRestrictedMovement	441	1087

6. CONCLUSIONS AND PERSPECTIVES

As a conclusion of the obtained results, most of the introduced attributes not give enough differentiation to achieve an optimal classification. The experiments show that only the time duration of each segment, and the direction variation seems to have enough variability between the different classes to consider it useful to the proposed classification problem.

With that fact in mind is obvious that is necessary to find in future works other variables that have the required differentiation between classes to achieve an optimal classification that could be useful for the maritime surveillance systems as information of the context of operation.

Also, the SVM algorithm has clearly superior results than the decision trees algorithm so in future works is possible to study different classification algorithms to a major improvement of the success rate obtained from the algorithms shown on this paper.

7. ACKNOWLEDGEMENT

This work was funded by public research projects of Spanish Ministry of Economy and Competitivity (MINECO), reference TEC2017-88048-C2-2-R.

8. REFERENCES

[1] J. Llinas, L. Snidaro, J. García, and E. Blasch, "Context-Enhanced Information Fusion," in *Context Enhanced Information Fusion*, Springer, 2015, pp. 3–23.

[2] M. A. McNicholas, B. Wilson, and S. D. Genovese, *Maritime Security: An Introduction*. 2016.

[3] D. Amigo, D. Sánchez Pedroche, J. García, and J. M. Molina, "AIS trajectory classification based on IMM data," *Int. Conf. Inf. Fusion*, 2019.

[4] D. Nguyen, R. Vadaine, G. Hajduch, R. Garello, and R. Fablet, "A Multi-task Deep Learning Architecture for Maritime Surveillance using AIS Data Streams," 2018.

[5] V. F. Arguedas, G. Pallotta, and M. Vespe, "Maritime Traffic Networks : From Historical Positioning Data to Unsupervised Maritime Traffic Monitoring," vol. 19, no. 3, pp. 722–732, 2018.

[6] G. Pallotta, M. Vespe, and K. Bryan, "Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction," *Entropy*, vol. 15, no. 6, pp. 2218–2245, 2013.

[7] M. Riveiro, G. Pallotta, and M. Vespe, "Maritime anomaly detection: A review," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 5, 2018.

[8] Y. U. Zheng, "Trajectory data mining: An overview," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, p. 29, 2015.

[9] H. Li, J. Liu, R. W. Liu, N. Xiong, K. Wu, and T. H. Kim, "A dimensionality reduction-based multi-step clustering method for robust vessel trajectory analysis," *Sensors (Switzerland)*, vol. 17, no. 8, 2017.

[10] A. Graser and P. Widhalm, "Modelling massive AIS streams with quad trees and Gaussian Mixtures," *21st Int. Conf. Geogr. Inf. Sci. (AGILE 2018)*, pp. 1–5, 2018.

[11] K. Yoo, J. Chun, and J. Shin, "Target Tracking and Classification for Missile Using Interacting Multiple Model (IMM)," *2018 Int. Conf. Radar, RADAR 2018*, no. Imm, pp. 1–6, 2018.

[12] J. García, J. A. Besada, J. M. Molina, and G. De Miguel, "Model-based trajectory reconstruction with IMM smoothing and segmentation," *Inf. Fusion*, vol. 22, pp. 127–140, 2015.

[13] Danish Maritime Authority, "AIS data." [Online]. Available: https://www.dma.dk/SikkerhedTilSoes/Sejladsinformation/AIS/Sider/default.aspx#.