Maritime Surveillance, Vessel Route Estimation and Alerts using AIS Data

Spyridon Patmanidis[1], Iasonas Voulgaris[2], Elena Sarri[1], George Papavassilopoulos[1], George Papavasileiou[2]

Abstract— **This paper focuses on estimating the route of a vessel and detecting abnormal behaviors based on AIS data. Linear Filtering using ARMA models is proposed in order to create trajectory forecasts. Furthermore, various alert criteria are set expressing different abnormalities. The software application created to test the proposed methodology is presented. AIS data collected from Saronikos Bay were used and the results show that the program could successfully monitor the bay in real time and report abnormalities in vessel behavior.**

Keywords – Estimation, Linear Filtering, ARMA, AIS Vessel Monitoring, Alert

I. INTRODUCTION

Today's rapid technological advancement enables vessels to be equipped with electronic devices, which transmit real time data containing positioning and speed information. The data can be processed accordingly and help surveillance authorities to monitor naval traffic and vessel interaction. Naval traffic surveillance is essential, both for vessels' security and passengers' safety, as well as for illegal activity detection (e.g. smuggling, illegal immigration, etc.). Nowadays, maritime surveillance becomes crucial since major immigration volumes appear.

Automatic Identification System (AIS) [1] is an electronic system, which can provide real time data, such as geographical position, course and speed details, etc., that can be used to monitor naval traffic. The problem is that the amount of data AIS produces is huge. As a result, automated procedures are required in order to filter the data and make them useful for use.

In literature, there is a variety of techniques proposed to solve the problem of detecting abnormal vessel behavior. Mascaro et al. [2] proposed the use of Bayesian Networks, trained with real data collected by AIS and producing models at two different time scales - both for the track as a whole and moment to moment, in order to learn abnormal vessel behavior. Johanson and Falkman [3] also proposed the use of Bayesian Networks for detection of anomalous vessel behavior due to their ability to include expert knowledge into the model and their simplity of understanding and interpretation. Their approach is implemented on synthetic data. However, Bayesian Networks have the disadvantage that AIS data need to be preprocessed before used. Another, widely used technique is the route model creation using Genetic Programming, which has the disadvantage of increased complexity. Kowalska [4] proposed the use of a Bayesian Model in combination with a Machine Learing Technique - Active Learing - in order to compute the model. Will et al. [5] present a state-of-the-art non-parametric regression model based on Gaussian Processes, constructed from AIS. Kd-Tree is also used in this work in order to decrease the complexity. Finally, another popular solution is the combination of AIS with other surveillance equipment, such as special radars for example [6], [7].

It is known, that on Euclidean space the shortest distance between two points is the straight line. Vessels on the sea, are moving in straight line in some projected space if possible, and their route shows long linear segments. Linear Estimators can therefore be used to forecast a vessel's route. Additionally, AIS data consistency can be tested by using simple kinematic equations. This work focuses on estimating a vessel's future route based on past positioning information and setting alert cases in order to expose abnormal vessel behavior and AIS data forgery. A software application has also been developed to test the proposed algorithm. The simulations conducted show, that the algorithm can be applied to real practical problems and give robust results. This method also gives the advantage of real time naval monitoring.

This work introduces an intergrated tool, which can assist the naval authorities detect abnormalities in vessel behavior. It proposes basic alert criteria and ways to detect these alerts. It also shows that Linear Estimators can be used effectively to forecast a vessel's route. The Linear Estimator proposed is the ARMA model, which is a fundamental algorithm in the field of forecasting [8], [9]. This algorithm can be found in numerous variations, and has many applications in the fields of economics [10], weather prediction [11], [12], electric power load prediction [13], etc.

Section II presents the method used to forecast the vessel's future position and various alert criteria for the traffic monitoring and data integrity. Section III describes the software application created to test the proposed methodology to forecast the trajectory of a vessel and detect the various alerts. In section IV the simulation results are presented followed by a short discussion on them.

II. METHODOLOGY

The first part of this section, presents the method used in order to forecast the future positions of a vessel based on

[1]S. Patmanidis, E. Sarri, G. Papavassilopoulos are with the Department of Electrical and Computer Engineering, National Technical University of Athens, 15780 Athens, Greece, spatmanid@gmail.com; elena@netmode.ntua.gr; yorgos@netmode.ece.ntua.gr

[2]I. Voulgaris and G. Papavasileiou are with Interoperability Systems International ISI Hellas S.A, 16451 Argyroupoli, Greece iasonas.voulgaris@isihellas.com; george.papavasileiou@isihellas.gr

previous positions. On the second part, there are definitions of the inconsistencies considered alert cases.

A. Vessel Route Estimation

AIS data contain the vessel's geographical position (Latitude and Longitude). Using Time-series Analysis on these data, it is possible to create a route model and based on this model the future position can be forecasted. More specifically, Linear Filtering/ARX models [14] are used in order to forecast the vessel's future position. These models are special cases of autoregressive moving average (ARMA) models [15], commonly used on Time-series Analysis. A short analysis on ARMA models is presented below. A Deterministic ARMA model, is a model format in which the output vector is expressed as a linear combination of past outputs, y(t), and past inputs, u(t):

$$A_0 y = -\sum_{j=0}^{n_1} A_{\mathrm{j}} y(t+j) + \sum_{j=0}^{m_1} B_{\mathrm{j}} u(t-j-d), t \geq 0 \quad (1)$$

where A_0 is square and nonsingular, and d represents a time delay.

The term in the past values of y is called the autoregressive component and the terms in u is called the moving-average component. Using q^{-1} as the backward shift operator, the model (3.1) can be expressed as

$$A(q^{-1})y(t) = B(q^{-1})u(t), t \geq 0 \quad (2)$$

where

$$A(q^{-1}) = A_0 + A_1 q^{-1} + ... + A_{\mathrm{n1}} q^{-\mathrm{n1}}$$

$$A_0 \ nonsingular$$

$$B(q^{-1}) = (B_0 + B_1 q^{-1} + ... + B_{\mathrm{m1}} q^{-\mathrm{m1}}) q^{-\mathrm{d}}$$

A DARMA model is equivalent to an observable state-space model with arbitrary initial state. Also, it can describe the input-output properties of a general state-space model (which is not necessarily completely observable or completely controllable) having arbitrary initial state [15].

A natural generalization of a general deterministic linear system that can be described by a DARMA model to the stochastic case is to add an independent noise input [15]. This leads to the stochastic autoregressive moving average model with auxiliary input:

$$A(q^{-1})y(t) = B(q^{-1})u(t) + C(q^{-1})w(t) \quad (3)$$

where $\{w(t)\}$ is a white noise sequence and $C(q^{-1})$ is a filter of the form

$$C(q^{-1}) = I_0 + c_1 q^{-1} + ... + c_{\mathrm{n}} q^{-\mathrm{n}}$$

The notation ARMA is introduced for the model 3, when $u(t) \equiv 0$, that is,

$$A(q^{-1})y(t) = C(q^{-1})w(t) \quad (4)$$

The full model 3 will be called the stochastic autoregressive moving average model with auxiliary input, or ARMAX model for short.

The variation of ARMA applied in the specific application is the Linear Filtering/ARX model. AIS transmits data on specific time moments. These data contain the information needed to create the ARX models [14]. More specifically, we need the geographical position (Latitude and Longitude), as well as the time frame the measurement took place. However, Time-series analysis requires measurements on constant time frames and AIS does not support this kind of data. Because of AIS standards, AIS data are not suitable for Time-series analysis. They have to be processed to extract the information needed (position on specific time moments). Therefore, resampling at constant time intervals T is needed. As soon as, the time interval T is set, the problem needed to be solved can be defined as: Knowing vessel's position on time moments T, $2T$... , kT, where $k \in \mathbb{N}$, estimate of the positions for the future moments $(k+1)T$, ... ,$(k+n)T$ is desired, where $n \in \mathbb{N}$. The k positions come after applying linear interpolation on AIS data between T and kT time.

The above problem can be solved using the ARX model. The model's mathematic description is the following:

$$\hat{Y}_{\mathrm{k+1}} = \sum_{i=m}^{k} A_{\mathrm{i}} Y_{\mathrm{i}} \quad (5)$$

where Y_{k} is the vessel's position vector on time moment kT (2x1 vector, where $y_{1,1}$ contains the Latitude and $y_{2,1}$ contains the Longitude), $m < k$ where $m \in \mathbb{N}$, and is the index of a past time moment. A_{i} is a 2x2 matrix whose elements belong in $\mathbb{R}$. The least-squares method is used to compute $A_{\mathrm{i,(j,k)}}$ matrices. After that, the positions on time moments $(k+n)T$ can be estimated using the formula:

$$\hat{Y}_{\mathrm{k+n}} = \sum_{i=0}^{n-2} A_{\mathrm{k-i}} \hat{Y}_{\mathrm{k-i+n-1}} + \sum_{i=0}^{k-m-n+1} A_{\mathrm{k-i-n+1}} Y_{\mathrm{k-i}} \quad (6)$$

B. Alert Criteria

This part defines the alert criteria, whose violation entails abnormal behavior. These alerts can be used by marine authorities for effective supervision. The alert criteria are the following:

- Divergence of the estimated position from the real position (Figure 1, left). If the estimated position lies outside a circle centered at the real position, with radius based on the vessel type, then the speed of vessel will change. This change may indicate a problem or some sort of abnormality that has to be checked.
- Trajectory changes within a certain angular range (e.g. $3\pi/4, \pi$) in a given time frame (e.g. 5-10 minutes) (Figure 1, right). Rapid change on the vessel's course may indicate abnormal behavior.
- Based on vessel type, which means different permitted max velocity, a vessel is forced not to exceed that velocity. If a vessel moves with speed exceeding the maximum permitted, an alert is signaled.

Knowing that AIS measurements contain data both for the location (Latitude and Longitude) and the movement (SOG – Speed Over Ground, COG – Course Over Ground) of

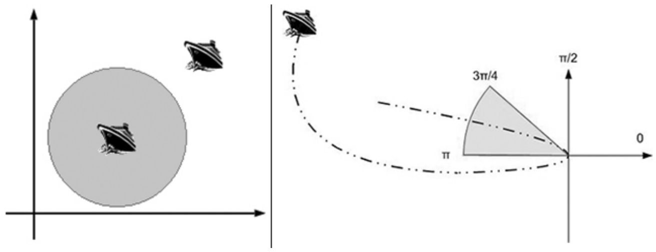

Fig. 1. Divergence and Angular Range Alert Criteria

Step 1: Load Data
Step 2: Resample Data
Step 3: Check Alert Criteria
Step 4: Report Alerts
Step 5: Create ARX Model and Forecast Future Positions
Step 6: Return to **Step 1**

Fig. 2. Pseudo-code of the algorithm

the vessel, deciding if these data are robust or they have been forged is essential. Using location and movement data separately, it is possible to detect forgery if one of these two data categories have been altered, or if both have been forged but their results don't match. Consequently, we can implement the following alert criterion:

- Based on velocity and location (Latitude and Longitude), if $v \cdot t >> s = \|y_{\mathrm{i}} - y_{\mathrm{i\text{-}1}}\|$ an alert is triggered. y_{i}, $y_{\mathrm{i\text{-}1}}$ are the positions of the vessel on time moments i and $i-1$, v is the vessel's moving speed and t is the time interval between time moments i and $i-1$. Considering that the velocity remains fixed as the vessel, moves violation of the above equation requires an alert.

In order to compute the distance between two consecutive locations, the Vincenty algorithm [16] is used. Vincenty's formulae are two related iterative methods used in geodesy to calculate the distance between two points on the surface of a spheroid, developed by Thaddeus Vincenty (1975). They are based on the assumption that the figure of Earth is an oblate spheroid, and hence they are more accurate than methods such as great-circle distance which assume a spherical Earth.

- Based on vessel type, different maximum speed is feasible. If a vessel moves faster than the feasible velocity, an alert is signaled.
- The lack of AIS data for certain periods also signal alerts. AIS is designed to send data regularly, therefore missing data for a big period of time means abnormality.

III. APPLICATION

In this section, we describe the program created in order to monitor vessels' route and detect anomalous behavior based on data collected by AIS. The program was implemented on Matlab environment and uses real AIS data transmitted from cargo vessels. Numerous tests using different databases have been conducted to test the integrity of the results.

The input data are formed as a SQLite database. Every two minutes the program loads the received data. After a simple preprocess to get the suitable form, the data are loaded on a Matlab structure which contains information for every ship. That preprocess is the resampling step (Step 2, Figure 2), which is necessary because AIS does not transmit information on constant time steps and Time-Series Analysis requires data on constant time steps. To get proper data form to create the ARX models resampling the data is required.

If the ship is not anchored, the program compares the route of the ship reported by AIS with the one estimated before the update and checks if any of the anomalies presented on section II occurs. Then, it uses the position data of the previous minutes in order to create a new ARX model (see section II) and based on that model, computes the vessels' route for a future time interval.

More specifically, the creation of the ARX model requires a certain amount of past data. By default the program created, waits to collect a source input containing information representing a time interval of ten minutes (one measurement for each minute). If there are not enough source data for a vessel, an alert is reported and steps 2 to 5 (Figure 2) are skipped for this particular vessel until the program collects enough data.

On Step 3 (Figure 2), the program checks the conditions set as alerts. If any of the alert conditions are met, the discrepancy is reported on Step 4 (Figure 2). Every alert type has its own unique key number and it is reported with the vessels Maritime Mobile Service Identity (MMSI) [1] plus information of the time and the location.

Finally, ARX models created (Step 5, Figure 2) for every vessel, forecast the vessel's route, in order to compare the forecasted route with the real one the AIS will report on the next two minutes.

The algorithm of Figure 2 sums up the program created to monitor the naval behavior of vessels. The default values of the past data were chosen after repeated trial testing to meet the needs of the specific naval application. Using the program for different applications, will require different parameterization.

In the next section we present and discuss the results from the naval monitor using our program.

IV. SIMULATION RESULTS & DISCUSSION

In this section, we present the experimental results of the application presented on section III. The program loads the available information, checks the alert conditions, reports the alerts and forecasts the vessel's route for the next two minutes. When, after two minutes, enough data has been received by AIS, the program loads the data and the procedure is repeated.

The following figures present the simulation results for a vessels. Figure 3 maps vessel's trajectory resulting from AIS data (circles - Data) and the trajectory resulting from the resampling procedure (crossed - rData). It is easy to notice, that the resampled positions fit perfectly to the real route and there is no divergence. The same result applies on all vessel routes tested, so we conclude that resampling does not affect the data integrity.

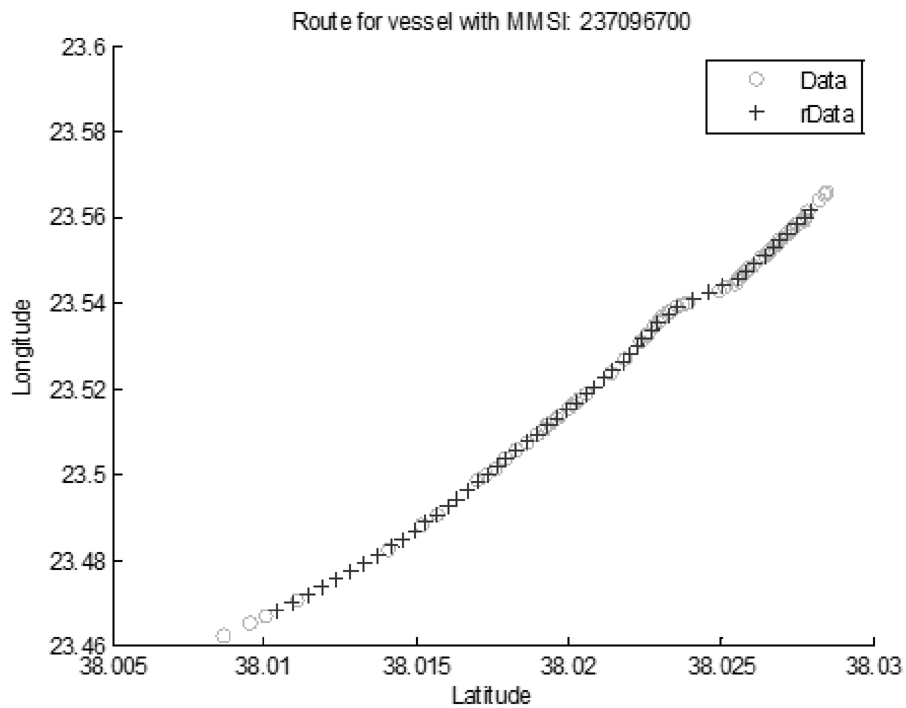

Fig. 3. Vessel's real and resampled route

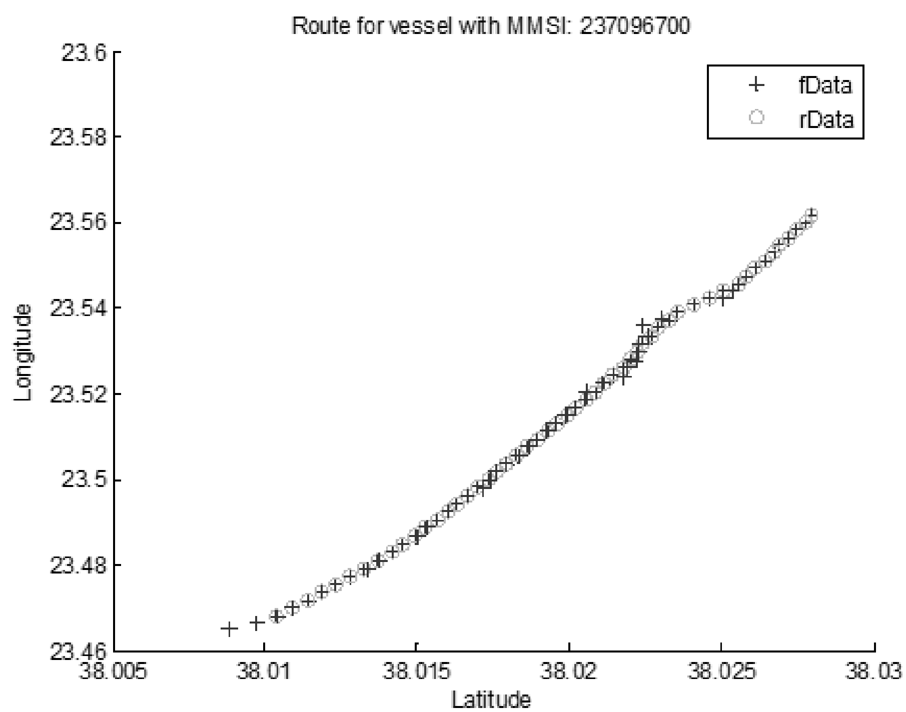

Fig. 4. Vessel's resampled and forecasted route

Figure 4 maps vessel's trajectory resulting from the resampling (circular - rData) and the positions estimated by the ARX model (crosses - fData). Absence of crosses means that there are not enough data in order to initialize the ARX model, which triggers an alert reported to the user. Simulation results showed that the models created, succeeded in forecasting the future position of the vessels on the majority cases. Exceptions appeared in cases of lack of data transmission for significant periods of time, the speed of the vessel had variations (large values for acceleration or deceleration) and the vessel declined a lot from linear direction. These behaviors are not considered normal for a vessel. Failing to forecast those trajectories is desired, since further investigation is needed by the authorities and an alert is created.

It is important to point out the consistency of the program. The program has been tested on hundreds of vessels and was almost completely able to detect anomalies. It was also able to forecast the future position of the vessel and detect forgery on data added manually by the testers. The forecasted positions showed very small divergence from the real ones (about 50-80 meters), within a tested vessel's length. False forecasting was detected only when there was data loss or when the vessels were having big turns on their movement. Although, the forecast on these cases fail to match the real route, this behavior is desired. On the first case, data absence is always an alert because AIS is obliged to send data on regular time moments, while on the second case, a vessel's route diverging from linear direction can, in some cases, be abnormal and in others a desired cause, of course. But, the program is always able to detect this divergence and inform the user of the system to investigate further.

Summarizing, the contribution of this work, is an applied tool created to monitor vessel behavior, in terms of producing forecasting and other introduced alerts. This tool has passed numerous tests and has been proved that it can be used effectively in order to detect alerts and abnormal vessel behavior.

As future work, the use of naval maps containing information on when a vessel has to change its route (for example avoiding crashing to islets) is proposed. Combining this information with the positioning information from AIS and forecasting can reduce resulting alerts sufficiently, when the change on the moving direction is mandatory. Furthermore, the combination of AIS with specialized monitoring equipment, for example radars, can be implemented in order to provide more accurate results.

V. CONCLUSIONS

Simulation results showed that the models used are able to estimate the route of a vessel. When significant divergence from linearity has been observed, the system recognized this change and warned the user in order to check for anomalous behavior. After the alert, the model adapted to the new route. As a result, the use of Linear Filter/ARX models has proved to be suitable for solving the problem of forecasting the future position of a vessel in a short time interval.

In addition, we saw that the detection of forgery in the AIS data could be attained without the use of additional systems (e.g. radars), except cases when forgery was conducted by specialists who know how to forge all kinetic data. Apparently, the absolute data robustness would require the use of additional monitoring systems apart from AIS. Nevertheless, this technique could be applied successfully leading the user to safer results. Summarizing, we could see that the proposed methods for maritime surveillance and vessel route estimation gave concrete and robust real-time results.

REFERENCES

[1] B. Tetreault, "Use of the Automatic Identification System (AIS) for Maritime Domain Awareness", in Proc. of OCEANS 2005, Washington DC, USA, September 2005, Vol. 2: 1590-1594.

[2] S. Mascaro, B. Kevin and A. Nicholson, "Learning Abnormal Vessel Behavior from AIS Data with Bayesian Networks at Two Time Scales", Technical Report, Clayton School of Information Technology, Monash University, 2010.

[3] F. Johansson and G. Falkman, "Detection of vessel anomalies - a Bayesian network approach", in Proc. of the 3rd International Conference on Intelligent Sensors, Sensor Networks and Information, Melbourne, 2007, pp. 395-400.

[4] K. Kowalska and L. Peel, "Maritime Anomaly Detection using Gaussian Process Active Learning", in 15th International Conference on Information Fusion, Singapore, 2012.

[5] J. Will, L. Peel and C. CLaxton, "Fast Maritime Anomaly Detection using Kd-Tree Gaussian Processes", in 2nd IMA Maths in Defence Conference, Shrivenham, 2011.

[6] M. Butler, "Project Polar Epsilon: Joint Space-Based Wide Area Surveillance and Support Capability", in IEEE International Geoscience and Remote Sensing Symposium, Seoul, 2005.

[7] F. Katsilieris, P. Braca and S. Coraluppi, "Detection of malicious AIS position spoofing by exploiting radar information", in Proc. of the 16th International Conference on Information Fusion, Istanbul, 2013.

[8] G.E.P. Box, G.M. Jenkins and G.C. Reinsel, Time Series Analysis, Forecasting and Control, 1994, Prentice Hall.

[9] G. P. Zhang, "Time series forecasting usinga hybrid ARIMA and neural network model", Neurocomputing, vol. 50, pp. 159-175, 2003.

[10] M. Rout, B. Majhi, R. Majhi and G. Panda, "Forecasting of currency exchange rates using an adaptive ARMA model with differential evolution based training", in Journal of King Saud University - Computer and Information Sciences, vol. 26, pp. 7-18, January, 2014.

[11] J. L. Torres, A. Garca. M. De Blas, A. De Francisco, "Forecast of hourly average wind speed with ARMA models in Nevarre (Spain)", in Solar Energy, vol 79, pp. 65-77, July, 2015.

[12] P. Burlando, R. Rosso, L. G. Cadavid, J. D. Salas, "Forecasting of short-term rainfall using ARMA models", in Journal of Hydrology, vol. 144, pp.193-211, April, 1993.

[13] A. J. Conejo, M. A. Plazas, R. Espinola, A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models", in IEEE Transactions on Power Systems, vol. 20, pp. 1035-1042, May, 2005.

[14] L. Ljung, Matlab & Simulink System Identification Toolbox User's Guide, Natick, MA, USA: Mathworks, 2014.

[15] G. C. Goodwin and K. S. Sin, Adaptive Filtering Prediction and Control, New York, NY, USA: Dover Publications, 2009.

[16] Vincenty's formulae, 3 December 2014. [Online]. Available: http://en.wikipedia.org/wiki/Vincenty%27s_formulae.