engineering proceedings

MDPI

Proceeding Paper

AI-Driven Estimation of Vessel Sailing Times and Underwater Acoustic Pressure for Optimizing Maritime Logistics †

Rosa Martínez Álvarez-Castellanos *, Jose Antonio García Gambín and Ivan Felis Enguix

Centro Tecnológico Naval y del Mar, 30320 Fuente Álamo, Murcia, Spain; email1@email.com (J.A.G.G.); email2@email.com (I.F.E.)

* Correspondence: rosamartinez@ctnaval.com; Tel.: +34-968-19-75-21

† Presented at the 10th International Electronic Conference on Sensors and Applications (ECSA-10), 15–30 November 2023; Available online: https://ecsa-10.sciforum.net/.

Abstract: This paper presents an innovative AI-based approach to estimate vessel sailing times in port surroundings. Leveraging historical vessel data, including ship characteristics and weather conditions, the model employs preprocessing techniques to enhance accuracy. Additionally, an underwater acoustic propagation model studies underwater noise pressure, aligning with environmental goals. The dataset, covering January to December 2022 in the Port of Cartagena, Spain, undergoes analysis, revealing intriguing patterns in ship routes. Employing various ML models, the study selects Random Forest as the most accurate, achieving an R2 of 0.85 and MSE of 0.145. The research showcases promising accuracy, aiding port optimization and environmental impact reduction, advancing maritime logistics with AI.

Keywords: port call optimization; artificial intelligence; machine learning; deep learning; maritime logistics; vessel dwell time; environmental impact

Citation: Álvarez-Castellanos, R.M.; Gambín, J.A.G.; Enguix, I.F. AI-Driven Estimation of Vessel Sailing Times and Underwater Acoustic Pressure for Optimizing Maritime Logistics. *Eng. Proc.* **2023**, *56*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s): Name

Published: 15 November 2023

Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Today, a large number of ships and nautical elements are active at sea. According to UNCTAD [1], approximately 80% of world trade is transported by sea, and this number is expected to further increase in the coming years [1]. In addition, shipping companies have been reporting over the years that disruptions and deviations from the initial plan occur frequently, resulting in most cases in delays [2]. These delays contribute to poor port optimization, disruptions in the market chain, and increased pollution, mainly greenhouse gas emissions and underwater radiate noise, due to prolonged idle times of vessels awaiting port calls. In fact, in April 2018, the IMO adopted the Initial Strategy for the reduction of GHG emissions from shipping which sets key ambitions, including cutting annual greenhouse gas emissions from international shipping by at least half by 2050, compared with their level in 2008.

This strategy goes in line with the Zero-Emission waterborne Transport, the Horizon Europe partnership that aims to deliver and demonstrate zero-emission solutions for all major ship types and services before 2030. Ports over the world are starting to implement R&D tools to optimize their own performance and to be able to partner with these strategies. One example is the Port of Rotterdam, which has developed and implemented tools based on the Just In Time (JIT) arrival criterion, optimizing the speed of each vessel throughout its journey and reducing CO_2 emissions by 14%.

In this paper, we present an innovative approach that incorporates artificial intelligence (AI) models, specifically machine learning (ML), and preprocessing techniques, to estimate the sailing time of vessels in port surroundings. All of this is accomplished by leveraging historical vessel data, such as ship characteristics, movement patterns, weather conditions, and port-specific factors (docks and areas of action). Also, by implementing underwater acoustic propagation model to each ship in its route, direct aspects related to the underwater noise pressure in the port context are studied. This study aligns with the MSFD, in particular regarding Descriptor 11 [3], searching a balance between optimizing economical marine activities with good environmental status.

2. Study Area: Cartagena Port

This study encompasses two port docks of Cartagena Port (sited in Murcia, Spain) specialized in different traffics: the Cartagena Dock (sports marinas, cruise ships and container cargo) and the Escombreras Dock (specialized in liquid and solid bulk traffic and storage activities) which was recently expanded and has resulting in the performance of an hegemonic role of the port in these traffics throughout the Spanish Southeast. The port receives important flows that cross the Mediterranean, having dense networks with the Maghreb, the French and the Italian coasts. The waters of the area are also furrowed by the local professional fishing boats and by the maritime traffic that connect the Atlantic Ocean and the Mediterranean Sea. Moreover, Cartagena receives part of the maritime passenger traffic that connects the Peninsula with the Balearic Islands. Besides, more and more cruise lines are calling at the port (240,000 cruise tourists on 170 ships in 2019), being one of the national ports that grows the most in this sense.

3. Methodology

Hence, in the present document, we propose a methodology focus on the data analysis, emphasizing on preprocessing, to enhance further predictions with machine learning models, as well as underwater acoustic propagation models to assess the underwater radiated noise of ships in the port surroundings with the aim of reducing their impact.

3.1. Data Analysis

The dataset derived from a Shiplocus (Multi-application platform for port management and maritime traffic exploitation (GMV)) account provided by the APC (Port Authority of Cartagena) through LIFE PortSounds (LIFE PortSounds. Reducing the impact of underwater noise on the marine environment of the Port of Cartagena (LIFE2020)) project. It consisted in 472 MB (1,585,941 × 32), of vessels and trajectories' relevant parameters of the selected area (Table 1), which corresponds to the Impact Zone (IZ) of the APC. A preliminary data analysis was conducted for computational purposes, thus undefined and incomplete data was removed, as well as irrelevant columns, maintaining '*Latitude*', '*Longitude*', '*MMSI*', '*Name*', '*Date*', '*Vessel type*', '*SOG*', '*COG*', '*Length*', '*Cargo*' and '*Registered Owner*' parameters. In addition, 'SOG' (Speed over the ground) is used to remove data coming from vessels moving at abnormal speeds, such as very low speeds (1.5 knots) or physically impossible speeds, given by the expression (1). Hence, dataset obtained after the preliminary analysis consists of 113 MB (58,632 × 11).

$$v_{max} = 2.8\sqrt{L} \quad (1)$$

where L refers to the vessels length.

Table 1. Impact Zone meshgrid coordinates.

Points	Lat	Lon
1	37°37,930′ N	01°10,613′ O
2	37°37,930′ N	00°33,988′ O
3	37°21,783′ N	01°10,613′ O
4	37°21,783′ N	00°33,988′ O

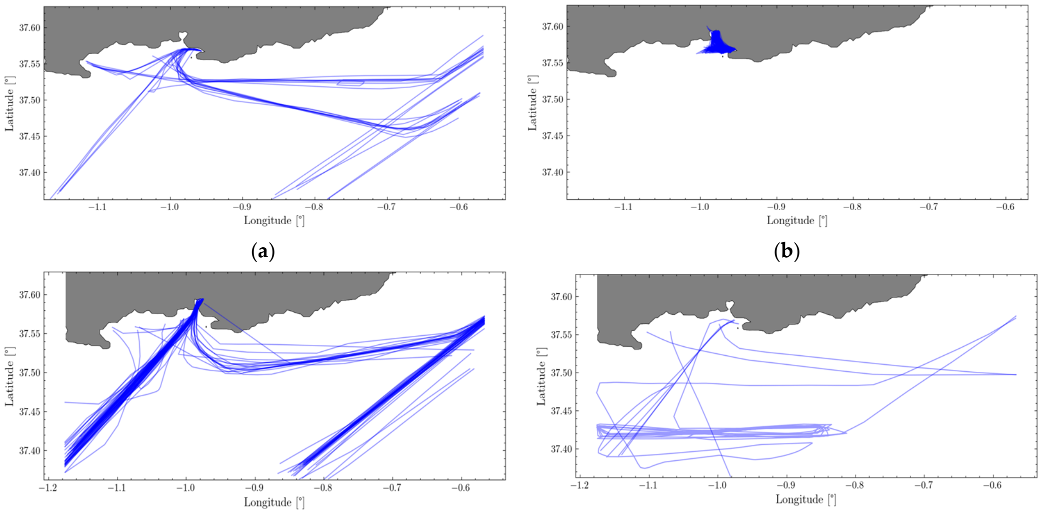
A **route** was defined as the **union of successive AIS messages** from a vessel, where successive messages are defined as those between which **no more than 5 h have elapsed**.

Therefore, AIS messages remaining on the dataset after the data processing steps were used to the crafting of routes.

Given the dataset (consisting in a concatenation of AIS points), routes were transformed into the following features: '*MMSI*', '*Time spent*', '*Vessel type*', '*Length*', '*Mean SOG*', '*Cargo*', '*Owner*', '*Arrival date*', '*Start point*', '*End point*' *and* '*Passing through*', where the new columns are:

- '*Time spent*': The duration of the whole route
- '*Arrival date*': The timestamp where the route begins
- '*Start point*': A sectorization of the area was performed and key areas were defined, so start point defines the key area where the route begins.
- '*End point*': As with the start point, the end point is given by the key area where the route ends.
- '*Passing through*': Coded as 'YES' or 'NO' whether a vessel is ending its itinerary in the port area or is just passing through the area but will not end its itinerary.

As it can be seen in Figure 1, vessels show great similarities in their routes over the year, except for the tugs, which are always moving close to the docks and show a big uncertainty in the duration of the routes, ranging from 1 h to 22 h. This is due to tugs being vessels that inhabit in the port and are designed primarily for towing and pushing other vessels in harbours, canals, and other confined waterways. As tugs are vessels from the port, they were excluded from the routes dataset. Also, in Figure 1d, anomalous routes can be observed. These routes were filtered to avoid confusions on the model.



(**a**) (**b**)

(**c**) (**d**)

Figure 1. Routes performed by: (**a**) LPG Tanker vessel; (**b**) Tug; (**c**) Container Ship; (**d**) Chemical/Oil Products Tanker.

With this, a thorough exploratory analysis is conducted to analyze routes differences and similarities among vessel's type, to understand their behaviour, as well as to sectorize areas depending on the traffic density. Thus, the sectorization made it possible to understand the key areas (areas with the higher density of AIS points) and to filter out those routes that did not start or end in a key area. Finally, a curated dataset (Figure 2) is obtained to implemented in underwater acoustic pressure and machine learning analysis.

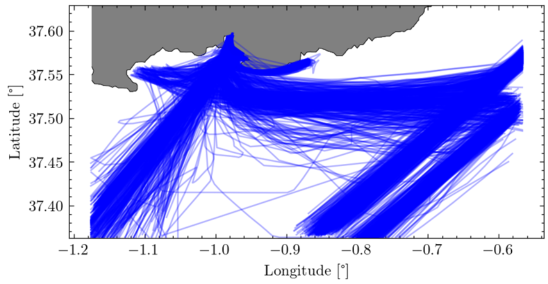


Figure 2. Set of routes after the clustering classification and the filtering of anomalous routes.

3.2. Underwater Acoustic Pressure

To assess the noise emitted by the vessels, **Ross** model [4] was applied to each route. **Ross** model takes into account physical vessel parameters, such as the speed and the length, and also parameters like the frequency to estimate the **Source Level** (SL), which is essentially the Sound Pressure Level (SPL) at 1 m distance of the acoustic source.

After the evaluation of the SL, a spherical loss model dependent only on the distance (for the sake of simplicity) was used to obtain a first approximation of the noise distribution over the area. Two cases are studied, one with a high number of vessels on the impact zone and one with just one vessel on the test site.

For the high number of vessels test, four ships (one cruise ship and three cargo ships) were selected from the data set. For these four ships, the SL was obtained at frequencies 62.5 Hz and 125 Hz, which are affected by cavitation noise. Once the SL was obtained, the spherical model was run in the area and the transmission losses were obtained. Finally, the SPL field that would be generated by each of the vessels was calculated and added coherently and 100% additively in linear units to obtain the worst case that could not occur in reality. The physical properties of the 4 ships selected are shown in the table below.

Table 2. Physical properties of the vessels modelled as sources.

Type	Speed (Knots)	Length
Cruise Ship	15	247
General Cargo Ship 1	10.7	108
General Cargo Ship 2	10.7	108
General Cargo Ship 3	7	90

For the one vessel case, the cruise ship from the first case is selected as the source to be modelled. The methodology is the same but this time no SPL maps need to be added, as there is only a single source on the area.

3.3. ML Models

Several machine-learning (ML) models were tested to predict the time spent for a vessel knowing its arrival date, starting and ending point and the passing through field. Such models were Gradient Boosting and Random Forest Regressor. Gradient Boosting (GB) is a machine learning algorithm that uses an ensemble technique to create a more accurate prediction model from multiple simpler models. The main idea behind GB is to combine several weak models to form one strong model [5]. Random Forest Regressor (RFR) is a meta estimator that fits a number of classifying decision trees on various subsamples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting [6]. For each model, hyperparameters optimization techniques were conducted and the results were compared to select the best model.

4. Results

4.1. Underwater Acoustic Pressure Assessment

As for the noise assessment, the Source Level of every route was studied and classified for every type of vessel.

As depicted in Figure 3a, the SL generated by the vessels at the 62.5 Hz frequency was mostly ranging between 150 and 165 dB. Also, for the 125 Hz component, the SL values were comprised between 110 and 170 dB, where most values were distributed between 140 and 155 dB.

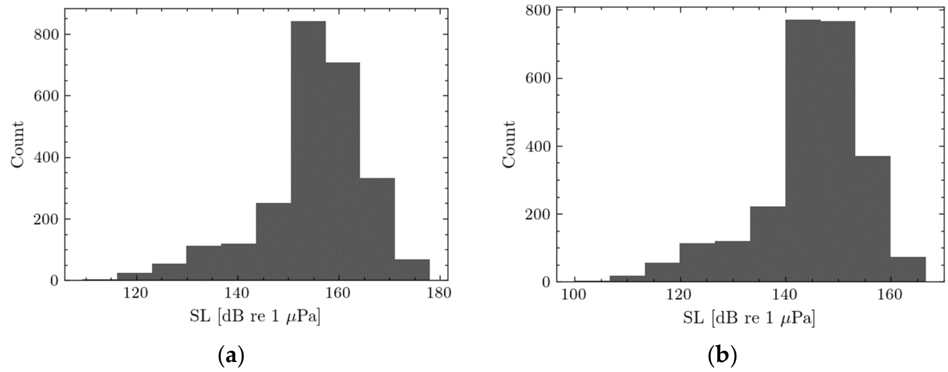


Figure 3. (**a**) SL histogram at 62.5 Hz; (**b**) SL histogram at 125 Hz.

In Figure 4, a SL breakdown per type of vessel and frequency is shown. As can be seen, the lowest values found are coming from the FSV (Fishing Support Vessel) type. Also, the biggest values are coming from the Cruise Ship and the LNG Tanker, whose median values are up to 165 dB in (a) and up to 155 dB in (b).

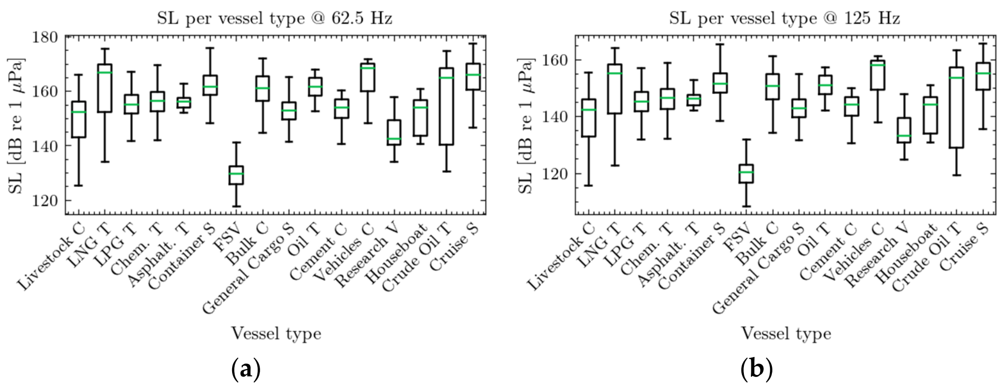


Figure 4. (**a**) Box and Whisker plot of SL values per vessel type at 62.5 Hz component; (**b**) Box and Whisker plot of SL values per vessel type at 125 Hz component.

Using the methodology described in Section 3.2, the SPL maps at 5 m depth for 63 and 125 Hz are obtained, both for one vessel case and four vessels case on the IZ of the Cartagena port (see Figure 5).

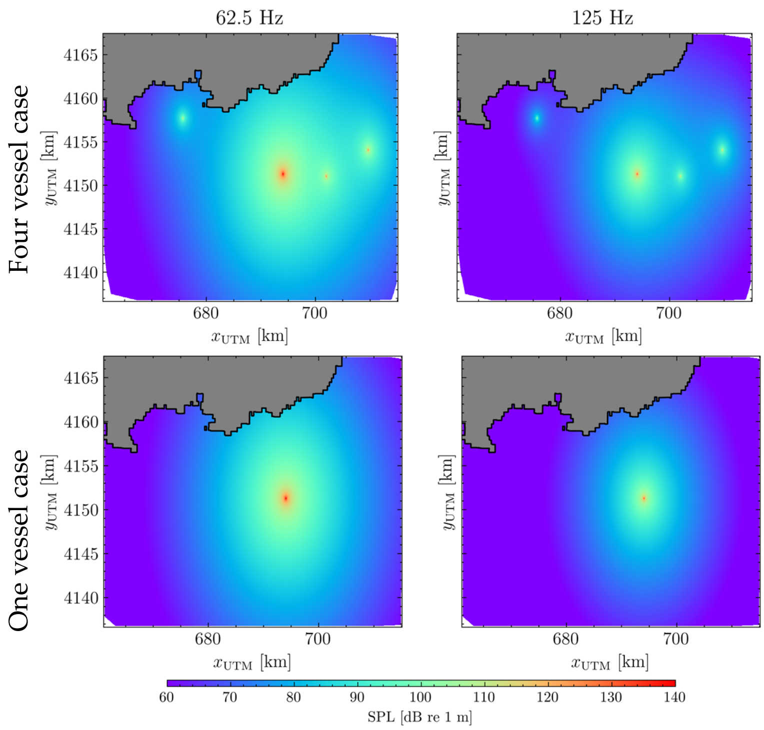


Figure 5. SPL map obtained at 5 m depth for the 4 vessels case (**above**) and for one vessel case (**below**) for 62.5 Hz and 125 Hz on the IZ of the project.

Highest SPL levels are found close to the source and range between 140 dB and 130 dB. In Figure 5, a slight difference between the two cases can be appreciated, as highest SPL in the distance is found for the 4 vessels case. For the four-vessel case, mean values of SPL for 62.5 Hz and 125 Hz, are, respectively, 75.8 dB/km^2 and 65.3 dB/km^2. For the one-vessel case, the values found are 72.6 dB/km^2 and 61.7 dB/km^2.

Between the one-vessel case and four-vessel case, SPL differences found are close to 4 dB/km^2. Also, it should be noted that the SPL levels found for the 125 Hz component are lower than for the 62.5 Hz, this due to the SL being lower for the 125 Hz.

It should be noted that these models were computed as a first approximation to understand underwater acoustic pressure, and the 100% coherent and additive addition of the acoustic waves carried out for the 4 vessels case will never occur in real terms.

4.2. ML Models

In Table 3 we can see the models metrics after the fine-tuning of the hyperparameters using a GridSearchCV algorithm, where it can be observed that RFR performs the best.

Table 3. Models summary and parameters.

Model	R^2_{train}	R^2_{test}	MSE
Gradient Boosting	0.884	0.8	0.198
Random Forest Regressor	0.974	0.85	0.145

5. Conclusions

This paper presents the study and descriptive analysis of an AIS dataset with the aim of creating an ML-driven tool for the optimization of waiting times in the port of Cartagena. In addition, a first approach to the visualization of the impact generated by the traffic from an acoustic point of view has been carried out. The machine learning model used was able to predict the transit time of vessels in the defined area with an MSE of 0.145.

The acoustic models, although built as a first approximation, showed differences between different frequencies and different numbers of coherent vessels.

Future work will focus on the use of more complex models such as MMPE (Miami Monterrey Parabolic Equation) for a better estimation of transmission losses in the area. In addition, these more complex models will be used over a time range of several hours to see the noise signature left by the traffic. On the other hand, more complex models such as NN will be used for a better prediction of the sailing time.

Author Contributions: Conceptualization and methodology, R.M.Á.-C. and I.F.E.; data curation, J.A.G.G.; formal analysis, J.A.G.G. and R.M.Á.-C.; validation, R.M.Á.-C. and I.F.E.; writing—original draft preparation, J.A.G.G.; writing—review and editing, R.M.Á.-C. and I.F.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Climate, Infrastructure and Environment Executive Agency (CINEA), grant number LIFE20 ENV/ES/000387.

Data Availability Statement: The data presented in this study are not publicly available due to privacy reasons.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Review of Maritime Transport 2022 | UNCTAD. Available online: https://unctad.org/rmt2022 (accessed on 22 February 2023).
2. Nikghadam, S.; Molkenboer, K.F.; Tavasszy, L.; Rezaei, J. Information sharing to mitigate delays in port: The case of the Port of Rotterdam. *Marit. Econ. Logist.* **2023**, *25*, 576–601.
3. Vighi, M.; Boschetti, S.T.; Hanke, G. *Marine Strategy Framework Directive Review and Analysis of EU Member States' 2018 Reports*; Publications Office of the European Union, Luxembourg, 2021.
4. McKenna, M.F.; Ross, D.; Wiggins, S.M.; Hildebrand, J.A. Underwater radiated noise from modern commercial ships. *J. Acoust. Soc. Am.* **2012**, *131*, 92–103.
5. Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurorobotics* **2013**, *7*, 21.
6. El Mrabet, Z.; Sugunaraj, N.; Ranganathan, P.; Abhyankar, S. Random Forest Regressor-Based Approach for Detecting Fault Location and Duration in Power Systems. *Sensors* **2022**, *22*, 458.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.