SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2024
ORGANIZATION AND MANAGEMENT SERIES NO. 208

DATA INTEGRITY ANALYSIS ON THE EXAMPLE OF AIS DATABASE

Adam KIERSZTYN[1*], Dariusz CZERWIŃSKI[2], Ernest CZERMAŃSKI[3], Aneta ONISZCZUK-JASTRZĄBEK[4], Konrad SMOLIŃSKI[5], Patrycja MIAZEK[6], Tomasz LASKOWICZ[7], Jakub JANKIEWICZ[8], Agnieszka RZEPKA[9], Radosław MIŚKIEWICZ[10]

[1] Department of Computational Intelligence, Lublin University of Technology; a.kiersztyn@pollub.pl, ORCID: 0000-0001-5222-8101
[2] Department of Applied Computer Science, Lublin University of Technology; d.czerwinski@pollub.pl, ORCID: 0000-0001-8631-1181
[3] Department of Maritime Transport and Seaborne Trade, University of Gdansk; ernest.czermanski@ug.edu.pl, ORCID: 0000-0002-2114-8093
[4] Department of Maritime Transport and Seaborne Trade, University of Gdansk; aneta.oniszczuk-jastrzabek@ug.edu.pl, ORCID: 0000-0002-4268-0011
[5] Department of Computational Intelligence, Lublin University of Technology; k.smolinski@pollub.pl, ORCID: 0009-0004-6155-2471
[6] Department of Computational Intelligence, Lublin University of Technology; p.miazek@pollub.pl, ORCID: 0009-0009-5980-8227
[7] Department of International Economics and Economic Development, University of Gdansk; tomasz.laskowicz@ug.edu.pl, ORCID: 0000-0003-2028-7291
[8] Department of Maritime Transport and Seaborne Trade, University of Gdansk; jakub.jankiewicz@ug.edu.pl, ORCID: 0000-0003-0408-2480
[9] Faculty of Management, Lublin University of Technology; a.rzepka@pollub.pl, ORCID: 0000-0003-4495-6066
[10] Institute of Management, University of Szczecin; radoslaw.miskiewicz@usz.edu.pl, ORCID: 0000-0003-2370-4783
* Correspondence author

Purpose: The purpose of this paper is to address the challenges of missing data and anomalies in the Automatic Identification System (AIS) database and to explore methods for data imputation and anomaly detection to optimize maritime traffic monitoring.
Design/methodology/approach: This research applies machine learning and data imputation techniques to AIS datasets to fill gaps and detect anomalies, aiming to enhance operational efficiency in maritime transportation.
Findings: The study finds that filling missing data improves vessel traffic monitoring systems, supports better asset management, and contributes to fuel consumption optimization in maritime operations.
Research limitations: Limitations include the reliance on available AIS data quality, with future research needed to integrate real-time environmental conditions and scale the methods to large datasets.
Practical implications: The findings offer practical solutions for improving maritime operations, leading to cost savings through optimized vessel management and reduced environmental impact.

Social implications: The research supports environmental sustainability by reducing emissions in maritime transport, influencing corporate responsibility and regulatory policies.
Originality/value: his paper offers innovative methods for AIS data imputation and anomaly detection, providing value to the maritime industry by improving decision-making and operational efficiency.
Keywords: data analysis, Automatic Identification System, maritime traffic monitoring, vessel traffic monitoring systems, fuel consumption optimization.
Category of the paper: Research paper.

Introduction

Filling in data gaps alongside anomaly detection has become one of the key challenges in the field of data analysis and machine learning (Little, Rubin, 2019; Karczmarek et al., 2020a, 2020b, 2021a; Kiersztyn et al., 2020a, 2020b; Łopucki et al., 2022). Many collected datasets contain missing values, which can result from various factors such as measurement errors, technical issues, or simply lack of information. These gaps can negatively impact the quality of analysis and modeling, leading to incomplete and inappropriate results. Data gap filling plays a crucial role in the process of data analysis and interpretation (Yu, Kim, 2019). There are numerous tech- niques and methodologies that can be applied to address this problem, depending on the type of data, available information, and analysis objectives (Kim, 2020; Yoon, Jordon, van der Schaar, 2018; Mishra, Singh, 2021; Zhang, Guo, Liu, 2019). Filling in missing data in the AIS (Automatic Identification System) database presents a significant challenge in the field of maritime traffic monitoring and management (Abbasi, Ghanbari, Manikas, 2019; Du et al., 2021; Wang et al., 2021; Han et al., 2020). The Automatic Identification System is widely used for collecting ship information in seas and oceans. AIS data is invaluable for various sectors such as navigation safety, route planning, port management, and marine environment. However, due to various factors like equipment failures, signal interference, deliberate transponder shutdowns, and other technical obstacles, the AIS database often contains gaps and missing information (Li et al., 2019, 2020; Chen et al., 2020; Xu et al., 2019; Yang et al., 2020; Wang et al., 2019; Liu, Lu, Zhan, 2021; Jin et al., 2020; Tang, Zhu, Ma, 2019). Filling in missing data in the AIS database is essential to ensure completeness and consistency of maritime traffic information. Missing data can lead to the loss of valuable vessel location, speed, course, and other parameters, which have serious implications for monitoring and management operations. In recent years, numerous research studies and scientific works have focused on developing effective methods and techniques for filling in missing data in the AIS database. This article provides an overview of various approaches and technologies applied in this area.

Economic justification of the significance of the problem

Vessel traffic monitoring systems provide support for a wide range of applications, including asset management, safety, and pollution compliance. Access to vessel position information, including GPS data, atmospheric information, and safety regulations, can help to avoid accidents but also to optimize vessel operating costs. The use of vessel traffic monitoring systems can bring tangible benefits to owners and operators, as they enable rapid response to changes in the environment, monitoring of technical condition, as well as faster and more efficient fleet management. Through their use, it is possible to reduce operating costs and increase safety. The economic benefits associated with choosing the optimal route include reduced damage to steering and hull propulsion systems, cargo and onboard systems, lower fuel consumption, or more timely arrival at the port of destination (Jurdziński, 2010).

In the area of ship operational optimization, the reduction of operating costs is most often indicated as the main objective. A basic solution that has been used for decades is slow steaming, i.e., ship operation based on reducing speed and thus progressively reducing fuel consumption (and thus fuel costs). This is the use of a function of fuel consumption and ship speed, which is similar to a logarithmic function in its course, and thus in the upper speed ranges allows a significant reduction in fuel consumption at the expense of a relatively small reduction in speed (Cariou, 2011).

From an economic point of view, lower fuel consumption generates lower voyage costs for the ship, but also results in lower greenhouse gas (GHG) emissions into the atmosphere. An important aspect that cannot be overlooked is the reduction of bunker fuel consumption costs, as bunker fuel consumption costs typically account for 50% (Notteboom, 2006) or even more than 60*%* (Golias et al., 2009*)* of a container ship's total operating costs. These include various types of navigation systems that optimize the sea voyage taking into account navigational and market conditions (e.g., fuel management, voyage weather planning, crew eco-driving training). The process of weather optimization of a sea route involves taking into account all historical data and forecasts for a given sea body of the future sea voyage, in order to best align the route with the main objective of minimizing energy (fuel) consumption, including, in particular, consideration of wind strength and direction and wave action. The influence of sailing speed on bunker fuel consumption in the area of shipping analysis has been written about (Golias et al., 2009; Christiansen et al., 2013; Meng et al., 2014a) including considering fleet deployment (Brouer et al., 2014), or in relation to speed and displacement (Álvarez, 2009).

In practice, this refers to the avoidance of storms, strong winds, and high waves, which increase the vessel's resistance to motion and result in either a reduction in speed at the same engine rpm or the need to increase engine rpm to maintain a constant cruising speed. Methods for selecting the optimum speed were described (Mulder, Dekker, 2014). Tests carried out over a full year on a specific vessel showed that by using a suitable weather optimization system it was possible to reduce fuel consumption by 4% over the year, with the potential to increase this reduction to 8%. An extension of this technology is to also take into account sea currents, which are quite well known and described by oceanographers. In extreme cases, this can help to increase speed by 3 kn while maintaining the same number of thruster revolutions, relative to traveling under the same conditions but without the assistance of a sea current in the direction of the voyage.

In contrast (Qi, Song, 2012), investigated the expected reduction in fuel consumption along a linear route when optimizing the voyage schedule of ships with uncertainty in port stays. (Wang, Meng, 2012) studied the problem of optimizing the sailing speed of a container ship considering container routing and handling.

The amount of fuel consumption on a ship is a function of its speed and the power required to achieve it. For each type of ship's power plant, fuel or ship size and type, this function will have different values and course. However, a common feature is that as speed increases, fuel consumption increases disproportionately faster. In other words, a unit increase in speed requires more units of fuel. While a reduction in speed has, as indicated, a non-linear effect on power demand and fuel consumption, from the point of view of shipping economics, a reduction in vessel speed has a linear correlation with capacity, where a vessel's efficiency decreases in proportion to the decrease in its operating speed. The implications of this are therefore clear and indicate that the more we reduce the speed of ships (to reduce CO2 emissions), the less capacity they will present a global annual basis.

The potential for reducing fuel consumption by managing the ship's trim is determined to be in the range of 1-4% (ABS Ship Energy Efficiency Measures Advisory, 2013). In a study carried out under real conditions on a container ship, fuel consumption was measured alternately at trim to stern – 60 cm and trim to bow – 60 cm. It turned out that despite the increase in displacement (by 255 tons – ballast water), the demand of this ship decreased by 2.6%, so that fuel consumption dropped from 63 to 61.5 tons per day. Under optimum conditions for this vessel, the maximum reduction in power required could be 2.8 per cent, translating into a reduction in fuel consumption of 3.77 tons of HFO per day. Considering that this vessel can make an average of 4.28 voyages per year carrying 37,200 TEUs, being at sea for an average of 282 days, the fuel saving potential could be 1,063.5 tons of fuel. Under non-SECA conditions for HFO fuel, this equates to fuel cost savings of USD 382,866 per year, translating into savings of approximately USD 10.30/TEU slot per voyage. For SECA conditions and the use of MGO fuel, the savings are much higher at USD 67,650 per year, or approximately USD 15.50/TEU slot (Czermański, 2019).

Vessel traffic monitoring systems can significantly improve the efficiency and productivity of the maritime industry. Firstly, they allow real-time data to be transmitted from remote locations, allowing resources and operations at sea to be monitored and managed from anywhere in the world. Secondly, they also offer enhanced communications security, ensuring that asset and operations data are protected from unauthorized access. Thirdly, the technology can help companies in the maritime industry optimize their operations to achieve greater efficiency and reduce operational costs, including fuel consumption. Vessel position data allows better planning of routes through specific areas. All this makes the maritime industry more efficient and productive.

Database Description

The analyzed database was purchased on a commercial basis from the S\&P Global data provider, within the IHS Markit. In the whole shape, the database consists of more than 225 indicators describing a ship, where data is (or should be at least) sourced from the vessel registration authority. For the purpose of the study, there were selected 4 constant characteristics of a ship: LRIMOShipNo, MMSI, ShipName, and CallSign.

LR/IMO Ship Number is a unique digit number assigned to a ship, remaining unchanged during the whole life of the ship, even in the case of rebuilding or Ship type conversion. This IMO number is assigned to the total or greater portion of the ship's hull, including the machinery space. The IMO identification number was adopted on 19th November 1987 in IMO Resolution A.600(15). The LR/IMO Number is never reassigned to another vessel. This number is also utilized in respect of SOLAS XI 1/3 and 1/5. The consequence of that is that we can assume the uniqueness of a ship hidden behind the IMO Number whenever registered/observed.

MMSI Number (Mobile Maritime Station Identifier) is a 9-digit number assigned to a ship towards identifying her via VHF radio communications. The first 3 digits denote the country of registry; therefore, by reflagging a ship, also the MMSI number is subject to be changed and updated. The consequence of that is, it cannot be assumed the unchangeability of the number in the database.

ShipName is another changeable and variable characteristic of a ship. It should be named in English format and in accordance with the registration authority up to date. The name of a ship can be changed independently from any other events, for any reason at any time. This also has a serious impact on the data reliability, and it cannot be included as fixed data.

CallSign is an alphanumeric identifier of a ship via radio communications and, similarly to the MMSI, is related to the flag of registration. Each flag authority possesses of a call sign range from that is selected a unique number to a ship. Therefore, as well as the MMSI, it should be noted that CallSign is variable by the reflagging of a ship.

Description of missing data

The database contains a range of columns enabling the identification of the ship for which a particular entry is made in the database. The following fields are available in the database: LRIMOShipNo, MMSI, ShipName, and CallSign. In theory, we are dealing with a significant data redundancy. However, upon careful analysis of the data for the Baltic Sea basin, interesting observations can be made. It turns out that many records have missing data, with corresponding values indicating the presence of data gaps in the respective fields.

Furthermore, for the MMSI number, there are values that do not meet the conditions imposed on the MMSI number. The analysis was conducted on entries registered from 2011 to 2022. The first entry was recorded on December 31, 2011, at 23:00:20, while the last one was on June 30, 2022, at 22:59:02. The data for the year 2011 only contains information about individual entries for the last hour of that year. Therefore, to increase the clarity of the analysis, these values were added to the data for the year 2012. In total, 348,591,048 records were analyzed for the Baltic Sea. The Figure 1 illustrates the number of entries for each year.

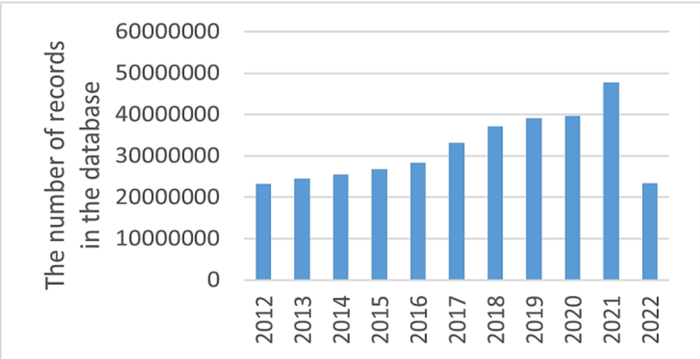

Figure 1. The number of records in the database for each year.

Table 1 presents the number of data gaps and incorrect values for each ship identification field.

Table 1.
Number of data gaps and incorrect values for ship identification fields

ID	Number of missing values
LRIMOShipNo	189607816
MMSI	537856
ShipName	2298526
CallSign	9945659

The largest number of data gaps is found in the LRIMOShipNo field. It turns out that for 189,607,816 records in this field, a value indicating a data gap was entered. Thus, the number of data gaps accounts for over 54% of all recorded entries. The distribution of data gaps across different years is also interesting.

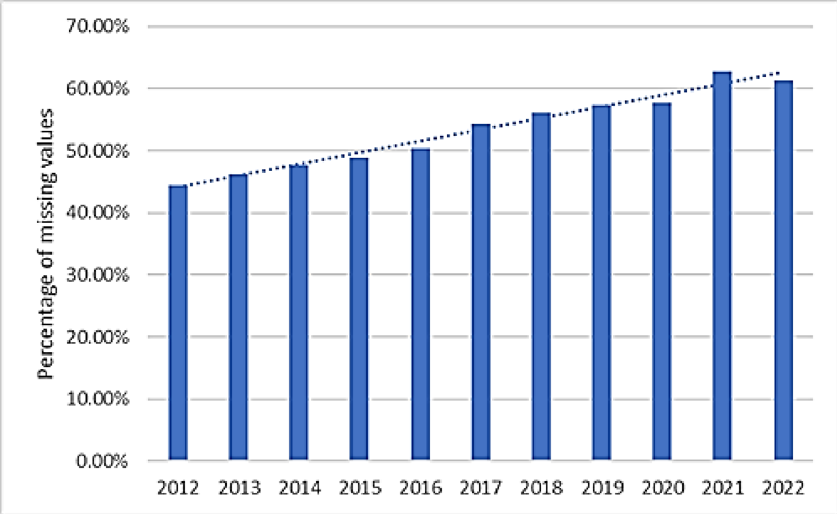

Figure 2. Percentage of data gaps for the LRIMOShipNo field.

Analyzing the results presented in Figure 2, we observe a consistent upward trend. The data for 2022 is incomplete, which may slightly distort the overall trend. In the case of other fields identifying individual records, a similar relationship is not observed (see Figure 3).

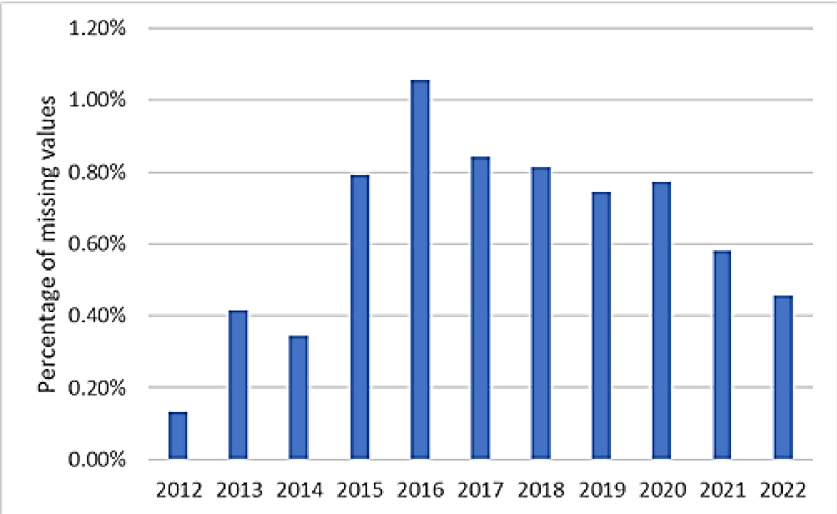

Figure 3. Percentage of data gaps for the SHIPNAME field.

At this point, a natural question arises: Is the distribution of data gaps in the LRIMO number influenced by certain factors? It seems reasonable to examine whether different months, days of the week, and hours exhibit the same level of data gaps or if there are higher percentages during certain time periods.

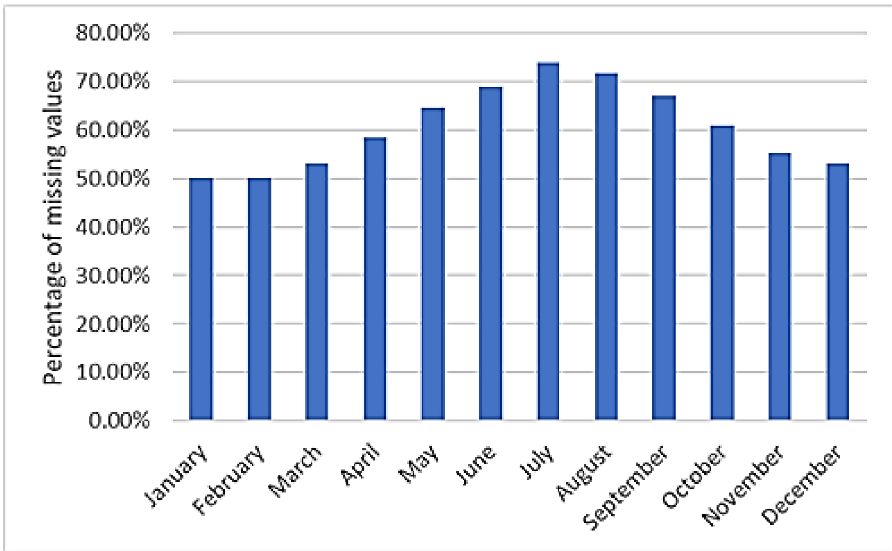

Figure 4. Percentage of data gaps in different months of 2021.

Analyzing the results presented in Figure 4, we can observe that the percentage of data gaps for the LRIMO identifier varies across different months. The highest percentage of data gaps occurs in the summer period, while the lowest is in the winter period. Furthermore, this difference is statistically significant, as confirmed by the conducted Fischer-Snedecor test. It is worthwhile to explain this fact and determine the reasons for the variation in data integrity across different months. When examining the data integrity level in terms of the influence of individual days of the week, the values obtained are presented in Figure 5.

Table 2.
Percentage of missing data for each type of vessel

ShipType	Number of occurrences	Number of gaps	Percentage of missing values
Anti Pollution	174092	88455	50.8093
Cargo	7246549	589947	8.1411
High Speed Craft	342685	160307	46.7797
Law Enforcement	470292	303333	64.4989
Medical Transport	16800	16798	99.9881
N/A	3697381	2074072	56.0957
Passenger	5021903	2887092	57.49
Pilot Boat	1441544	1260684	87.4537
Search And Rescue	1730106	1648592	95.2885
Tanker	2757768	286283	10.381
Tender	260860	251497	96.4107
Tug	2906347	900576	30.9865
Vessel	21634083	19461979	89.9598
Wing In Ground-effect	31368	27791	88.5967

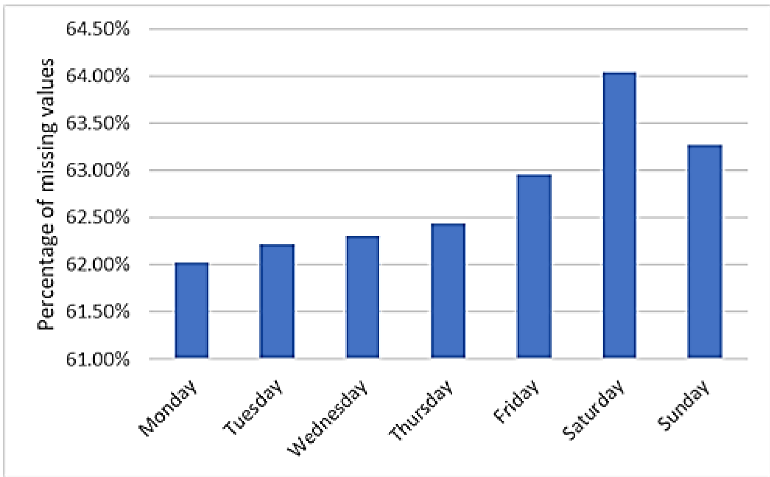

Figure 5. Percentage of data gaps on different days of the week.

In the case of the percentage of data gaps on different days of the week, we also observe a statistically significant difference. Interestingly, the highest number of data gaps is recorded on weekends, while the most consistent data comes from Mondays.

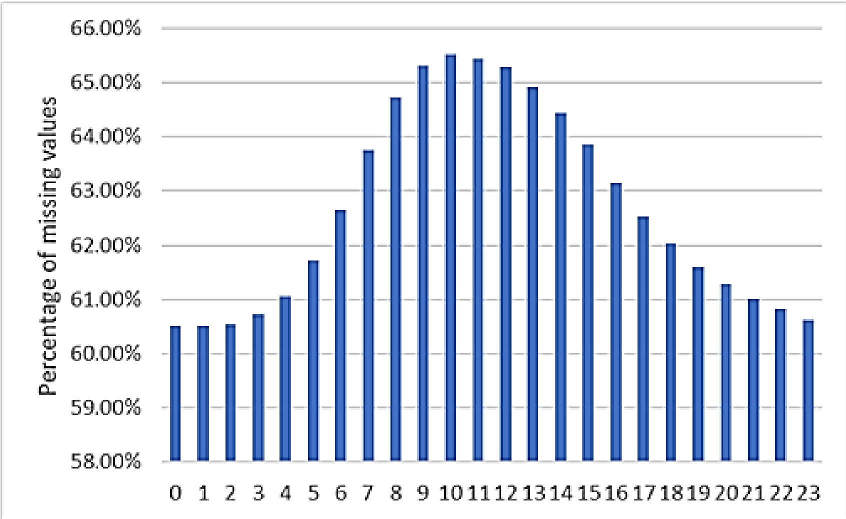

Figure 6. Percentage of data gaps broken down by hours.

Similarly to the previous analysis of the percentage of data gaps in the LRIMO column, broken down by hours (see Figure 6), statistically significant differences are observed. Additionally, it is worth noting that this distribution resembles a normal distribution, similar to the analysis based on the month of occurrence. In conclusion, it can be stated that the time of event registration statistically influences data integrity. Both the month, day of the week, and the hour of occurrence impact data quality. At this point, another natural question arises: Are there statistically significant differences for different types of ships and different ship operational statuses?

Table 3.
Percentage of gaps for each status

MovingStatus	Number of occurrences	Number of gaps	Percentage of missing values
Anchored	9768	3365	34.4492
Aground	1297701	162617	12.5312
Constrained by draught	63120	10122	16.0361
Engaged in fishing	1695834	1369200	80.739
Moored	9968212	3020910	30.3054
N/A	20609874	20139984	97.7201
Not under command	102833	55421	53.8942
Restricted manoeuverability	593835	214845	36.1792
Under way sailing	280263	151511	54.0603
Under way using engine	13110338	4829431	36.8368

Table 3 illustrates the impact of ship status on data quality. As expected, the highest percentage of missing data occurs in cases where the ship status is not provided. A significant percentage is also observed when the status is listed as "Engaged in fishing". Another interesting aspect is to examine how the position of the ship affects data integrity. To facilitate analysis, a random sample of 100,000 observations was taken from the data from the year 2021. The number of data gaps identifying each record was counted. In the selected sample, there were no cases where all fields were incomplete. It turns out that ships located far from the shore generally have complete data, while numerous data gaps occur for ships near the coast or inland (see Figure 7).

Figure 7. Percentage of data gaps broken down by hours.

The number of unique values for each identifier is different. This is evident from the values obtained for the most abundant year, 2021 (see Table 4).

Table 4.
Number of unique values for each identifier in 2021

ID	LRIMO	MMSI	ShipName	Call Sign
Number of different values	10598	81101	39299	41161

Impact of anomalies on filling data gaps

It might seem that in cases where missing data only occurs for certain ship identifying variables, it would be easy to fill in the missing entries based on available information. However, nothing could be further from the truth. It turns out that the lack of data integrity is not limited to the occurrence of missing data but also extends to inconsistencies in the designations. For example, we will limit ourselves to one selected ship whose MMSI number is 211002010. In the analyzed database, all LRIMO field values for this ship are -1000, which of course means no data. Therefore, it is not possible to complete the LRIMO values without collecting information from additional sources. In addition, it is worth noting that there is also a lack of consistency in the ShipName and CallSign fields. The distributions of individual variables are presented in tables 5 and 6.

Table 5.
Distribution of the ShipName variable for the analyzed vessel

ShipName	Count
-1000	1082
DELIVERANCE	3
IRMA	98
IRMA	1197
SII	4
SIRENITA	311
WINJA	1

Table 6.
Distribution of the CallSign variable for the analyzed vessel.

CallSign	1000	DGTF	WDJ3977	XXXXX
count	936	1677	1	82

The results presented in Tables 6 and 7 confirm major problems and lack of consistency in determining such key values as variables identifying the ship. This fact confirms the belief that other data contained in the database should also be approached with a great deal of uncertainty. Table 7 presents a description of anomalies for selected variables.

Table 7.
Selected types of anomalies for individual variables

Variable name	Anomaly description
Beam	=0
Length	=0
Speed	>90
Draught	=0
Heading	>360
Time diff	>3600
Time diff 2	<0

Values equal to zero were considered anomalies for the Beam, Length and Draft variables. In the case of the Speed variable, values greater than 90 knots were considered outliers, although it is safe to assume a much higher value. For the Heading variable, values outside the range of this variable were considered anomalies. Variables Time diff is the difference between the reporting time (MovementDateTime variable) and the time of saving to the database (ProcessedData). A delay of more than an hour was deemed to be an anomaly. In the case of the Time diff 2 variable, we are dealing with the difference between the reporting time and the expected time of arrival (ETA). If this difference is negative, i.e. the ship is late, we are dealing with an anomaly.

Analyzing the results presented in Table 8, it can be seen that in some cases the number of anomalies is significant and may affect the integrity of the data, and thus further analyses. In the case of Draft, nearly half of the records contain data that is suspected to be anomalous. In addition, it is worth noting that even in the case of a variable describing the destination port, we can encounter inconsistency. It is not uncommon for different notations for the destination port to be used, which can lead to errors.

Table 8.
The number of anomalies for individual variables and combinations of two variables

Variable Name	Number of Anomalies	Percentage of Anomalies
Beam	2,168,148	4.54%
Length	2,102,851	4.41%
Speed	8,798	0.02%
Draught	22,03,298	46.17%
Heading	38,950	0.08%
Time Diff	633	0.00%
Time Diff 2	11,173,346	23.41%
Beam & Length	2,065,820	4.33%
Beam & Speed	7,053	0.01%
Beam & Draught	1,630,572	3.42%
Beam & Heading	5,201	0.01%
Beam & Time Diff	88	0.00%
Beam & Time Diff 2	216,140	0.45%
Length & Speed	7,053	0.01%
Length & Draught	1,573,059	3.30%
Length & Heading	5,215	0.01%
Length & Time Diff	86	0.00%
Length & Time Diff 2	208,690	0.44%

Cont. table 8.

Speed & Draught	8,664	0.02%
Speed & Heading	47	0.00%
Speed & Time Diff	3	0.00%
Speed & Time Diff 2	92	0.00%
Draught & Heading	17,184	0.04%
Draught & Time Diff	476	0.00%
Draught & Time Diff 2	378,119	0.79%
Heading & Time Diff	1	0.00%
Heading & Time Diff 2	379	0.00%
Time Diff & Time Diff 2	58	0.00%

Conclusion

The in-depth analysis of data from the AIS database carried out above confirms that the lack of data integrity is a huge problem that both scientists and specialists in the field of data analysis face on a daily basis. The comprehensive list presented above confirms that both data gaps as well as numerous anomalies and outliers are not isolated phenomena. There are numerous missing data in the analyzed database. It is true that some of them can be supplemented based on the available information, but this is not always possible. In many cases, overlapping missing values in object-identifying variables render records useless. In addition, numerous outliers and anomalies present in the database make the data analysis process more difficult. An important factor affecting the possibility of data mining is the frequent inconsistency in designations and nomenclature.

Acknowledgements

The work was co-financed by the Lublin University of Technology Scientific Fund: FD-20/IT-3/002.

References

1. Abbasi, A., Ghanbari, A., Manikas, I. (2019). Data imputation in marine monitoring systems using machine learning techniques. *Journal of Marine Science and Engineering, 7(7),* p. 203.
2. Álvarez, J.F. (2009). Joint routing and deployment of a fleet of container vessels. *Maritime Economics & Logistics, 11,* pp. 186-208.
3. Brouer, B.D., Alvarez, J.F., Plum, C.E.M., Pisinger, D., Sigurd, M.M. (2014). Abase integer programming model and benchmark suite for liner-shipping network design. *Transportation Science, 48(2),* pp. 281-312.
4. Cariou, P. (2011). Is slow steaming a sustainable means of reducing CO2 emissions from container shipping. *Transportation Research Part D: Transport and Environment, 16(3),* pp. 263-273.
5. Chen, L., Jia, X., Yu, H., Lai, D. (2020). Missing data imputation for ship trajectories based on the combination of spline interpolation and Kalman filtering. *Advances in Mechanical Engineering, 12(9),* p. 1687814020954961.
6. Christiansen, M., Fagerholt, K., Nygreen, B., Ronen, D. (2013). Ship routing and scheduling in the new millennium. *European Journal of Operational Research, 228(3),* pp. 467-483.
7. Czermański, A. (2019). Morska żegluga kontenerowa a zrównoważony rozwój transportu. Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego Instytutu Transportu i Handlu Morskiego.
8. Du, Y., He, J., Yan, S., Tang, W. (2021). Research on ship trajectory reconstruction method based on incomplete data of automatic identification system. *Journal of Navigation, 74(1),* pp. 55-70.
9. Golias, M.M., Saharidis, G.K., Boile, M., Theofanis, S., Ierapetritou, M.G. (2009). The berth allocation problem: Optimizing vessel arrival time. *Maritime Economics & Logistics, 11,* pp. 358-377.
10. Han, S., Li, C., Yang, C., Ding, M. (2020). A novel method for filling incomplete AIS data based on extended Kalman filtering. *IEEE Access, 8,* pp. 138190-138199.
11. Jin, L., Zhao, J., Zhang, Y., Zong, C. (2020). Missing data imputation for ship trajectories using long short-term memory neural network. *Applied Sciences, 10(13),* p. 4504.
12. Jurdziński, M. (2010). Metody zmniejszenia zużycia paliwa w procesie eksploatacji statku. *Zeszyty Naukowe Akademii Morskiej w Gdyni, 67,* pp. 18-28.
13. Karczmarek, P., Kiersztyn, A., Pedrycz, W. (2020). *Fuzzy set-based isolation forest*. 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), July 2020, pp. 1-6.
14. Karczmarek, P., Kiersztyn, A., Pedrycz, W., Al, E. (2020). K-Means-based isolation forest. *Knowledge-based systems, 195,* p. 105659.

15. Karczmarek, P., Kiersztyn, A., Pedrycz, W., Czerwiński, D. (2021). Fuzzy c-means-based isolation forest. *Applied Soft Computing*, *106*, p. 107354.
16. Kiersztyn, A., Karczmarek, P., Kiersztyn, K., Pedrycz, W. (2020). Detection and classification of anomalies in large datasets on the basis of information granules. *IEEE Transactions on Fuzzy Systems*, *30(8)*, pp. 2850-2860.
17. Kiersztyn, A., Karczmarek, P., Łopucki, R., Pedrycz, W., Al, E., Kitowski, I., Zbyryt, A. (2020). *Data imputation in related time series using fuzzy set-based techniques*. 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), July 2020, pp. 1-8.
18. Kim, J. (2020). Missing not at random imputation with recurrent neural networks. *Journal of the American Statistical Association*, *115(530)*, pp. 461-476.
19. Li, A., Ma, X., Li, G., Zhang, X. (2020). AIS data imputation method based on convolutional LSTM network. *IEEE Access*, *9*, pp. 1372-1382.
20. Li, Y., Zhang, Z., Hu, Y., Sun, Y. (2019). Research on method of ship trajectory data interpolation based on AIS data. *Mathematical Problems in Engineering*, p. 2562451.
21. Little, R.J.A., Rubin, D.B. (2019). *Statistical analysis with missing data*. John Wiley & Sons.
22. Liu, M., Lu, Y., Zhan, J. (2021). A ship trajectory prediction method based on LSTM neural network using AIS data. *IEEE Access*, *9*, pp. 3508-3522.
23. Meng, Q., Wang, S., Andersson, H., Thun, K. (2014). Containership routing and scheduling in liner shipping: Overview and future research directions. *Transportation Science*, *48(2)*, pp. 265-280.
24. Mulder, J., Dekker, R. (2014). Methods for strategic liner shipping network design. *European Journal of Operational Research*, *235*, pp. 367-378.
25. Notteboom, T.E. (2006). The time factor in liner shipping services. *Maritime Economics and Logistics*, *8(1)*, pp. 19-39.
26. Qi, X., Song, D.-P. (2012). Minimizing fuel emissions by optimizing vessel schedules in liner shipping with uncertain port times. *Transportation Research Part E*, *48(4)*, pp. 863-880.
27. Tang, Y., Zhu, T., Ma, F. (2019). Trajectory reconstruction of incomplete AIS data based on improved particle filter algorithm. *Mathematical Problems in Engineering*, p. 4932180.
28. Wang, A., Meng, Q., Du, Y. (2015). Liner container seasonal shipping revenue management. *Transportation Research Part B*, *82*, pp. 144-160.
29. Wang, A., Ye, X., Tsui, K.L. (2017). *Time series imputation with generative adversarial nets*. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 755-764.
30. Wang, A., Zhou, J., Zhang, X., Li, Y. (2021). Trajectory-based ship movement prediction using AIS data and random forest. *Journal of Navigation*, *74(3)*, pp. 615-632.
31. Wang, S., Meng, Q. (2012). Sailing speed optimization for container ships in a liner shipping network. *Transportation Research Part E*, *48(3)*, pp. 701-714.

32. Wang, S., Thun, K. (2012). Minimizing CO2 emissions by optimizing the vessel speed and fleet operation in container shipping. *Transportation Research Part E*, *48(5)*, pp. 879-890.
33. Xu, Y., Cheng, J., Zhang, Y., Liu, X. (2019). Missing data imputation for AIS based on matrix completion and belief propagation. *Complexity*, p. 6212681.
34. Yang, B., Yang, C., Wang, Y., Zhao, C. (2020). A data recovery algorithm for incomplete AIS data based on multisource fusion. *Mathematical Problems in Engineering*, p. 8839494.
35. Yang, L., Zhang, J., Wang, W. (2020). A novel method for data reconstruction of AIS based on trajectory clustering. *IEEE Access*, *8*, pp. 146084-146095.
36. Yoon, J., van der Schaar, M. (2020). Missing data imputation with GANs: Theory and applications. *Journal of Machine Learning Research*, *21*, p. 1362.
37. Yoon, J., Jordon, J., van der Schaar, M. (2018). *GAIN: Missing data imputation using generative adversarial nets.* International Conference on Machine Learning, pp. 5672-5681.
38. Yu, H., Kim, J. (2019). Interpretable missing not at random models for decision making with missing data. *Journal of Machine Learning Research*, *20(17)*, pp. 1-32.
39. Zhang, Y., Guo, S., Liu, Y. (2019). A review on missing data imputation methods for environmental sensor networks. *Computers, Environment and Urban Systems*, *77*, p. 101364.
40. Zhang, Y., Hu, S., Li, Y. (2020). Missing value imputation based on deep learning and fuzzy systems. *IEEE Transactions on Fuzzy Systems*, *28(10)*, pp. 2341-2350.
41. Zhao, J., Li, X., Wang, S. (2020). AIS data imputation based on LSTM and KNN. *Mathematics*, *8(12)*, p. 2150.