Automatic Identification System (AIS) Data Reliability and Its Implications for Maritime Safety in Indonesia

Mohammad Danil Arifin[1*], Fanny Octaviani[2], Muswar Muslim[3], Danny Faturachman[4]
(Received: 17 September 2025 / Revised: 24 September 2025 /Accepted: 24 September 2025 / Available Online: 26 September 2025)

***Abstract*— The Automatic Identification System (AIS) is central to vessel monitoring, traffic management, and maritime safety, yet concerns remain regarding its reliability due to incomplete, inaccurate, or delayed reporting. This study assesses AIS data from the Indonesian maritime domain, focusing on four parameters: completeness, accuracy, consistency, and timeliness. AIS records data were preprocessed through data cleaning, filtering, and detection of missing values in static fields such as draught, beam, LOA, deadweight, and gross tonnage (GT). Statistical and spatial-temporal analyses using Python were applied to quantify missing data, identify anomalies, and evaluate reporting intervals. Results show high completeness (97.5%), although missing draught data (6.77%) limited under-keel clearance assessments, while small gaps in beam and LOA affected collision risk modeling and berth allocation. Accuracy was moderate, with invalid speed and course records observed, whereas consistency was excellent, with MMSI and ship names fully aligned. Timeliness proved weakest, with median reporting intervals (8,380 seconds) exceeding IMO standards, restricting real-time navigational use but remaining suitable for long-term monitoring. Overall, AIS in Indonesia is reliable for strategic traffic analysis but insufficient for operational safety management. Strengthening reporting compliance, integrating port and registry databases, and applying anomaly detection and satellite AIS are recommended to enhance maritime safety.**

Keywords—AIS data, data reliability, maritime safety

Corresponding Author: danilarifin.mohammad@gmail.com

I. INTRODUCTION

The Automatic Identification System (AIS) has evolved from a collision-avoidance aid into a ubiquitous data backbone for maritime safety, surveillance, port analytics, and policy. AIS continuously broadcasts vessel static attributes (e.g., MMSI, IMO, type, LOA, beam, draught, GT, DWT), dynamic kinematics (position, SOG, COG, heading), and voyage fields (destination, ETA). Those data feed Vessel Traffic Services (VTS), Search-and-Rescue (SAR), port call optimization, and risk modeling, increasingly with machine learning and big-data pipelines. As more authorities rely on derived indicators (e.g., congestion, emissions, risk heatmaps), data reliability (completeness, accuracy, consistency, timeliness) becomes a first-order safety parameter rather than a mere data hygiene concern. Recent literature emphasizes that missing or falsified AIS can degrade risk estimates, mislead anomaly detectors, and bias traffic statistics—ultimately raising operational safety risks [1] [2].

In parallel, AIS is expanding beyond terrestrial receivers via satellite AIS and migrating toward higher-throughput links such as VDES (VHF Data Exchange System). While these developments promise richer data, they also broaden the attack surface and expose reliability gaps (e.g., spoofing, interference, protocol misuse) that can ripple into safety-critical decisions if left unmanaged. Recent work proposes credibility monitoring for the AIS/VDES link layer and hybrid detection (interference + spoofing localization) to protect data integrity [3].

A. The Indonesian Imperative

Indonesia's archipelagic geometry, with major chokepoints (Malacca, Sunda, Lombok, Makassar), dense ferry routes, mixed fleets (from VLCCs to artisanal boats), and uneven terrestrial AIS coverage, makes AIS reliability uniquely consequential. Empirical studies specific to Indonesia are emerging (e.g., Makassar Port risk analysis that fuses FRAM with AIS trajectories), highlighting sociotechnical variability in VTS operations and the need to quantify risk from real traffic traces. Yet systematic, nationwide audits of AIS static-field completeness and their safety implications (grounding, collision, port throughput) remain sparse [4].

A second Indonesian challenge is the under-representation of certain vessels in AIS: coastal craft and small fishing boats often lack mandated transponders, biasing density and exposure maps in littoral zones. A recent multi-method study found that only ~43% of coastal traffic transmitted AIS, with strong variation by vessel type and season, warning that AIS-only situational awareness is incomplete in nearshore corridors [5].

Mohammad Danil Arifin, Department of Marine Engineering, Darma Persada University, East Jakarta, 13450, Indonesia. E-mail: danilarifin.mohammad@gmail.com

Fanny Octaviani, Naval Architecture Department, Darma Persada University, East Jakarta, 13450, Indonesia. E-mail: fanny_octaviani@yahoo.com

Muswar Muslim, Department of Marine Engineering, Darma Persada University, East Jakarta, 13450, Indonesia. E-mail: muslim.muswar@gmail.com

Danny Faturachman, Marine Engineering Department, Darma Persada University, East Jakarta, 13450, Indonesia. E-mail: fdanny30@yahoo.com

B. AIS Reliability

Reliability spans four dimensions, i.e., Completeness (no missing/nulls where expected); Accuracy (values agree with physics & ground truth); Consistency (identifiers and static fields remain stable), and Timeliness (report cadence matches maneuvering dynamics).

Large, real-world AIS sets routinely violate these. Static fields (draught, LOA, beam, GT, DWT) are often blank, stale, or implausible because they rely on manual entry and are weakly enforced during operations. Voyage fields (destination/ETA) are notoriously inconsistent. Dynamic tracks can contain gaps, clock drift, duplicated bursts, or non-physical kinematics that require denoising and interpolation. Researchers continue to propose cleaning, imputing, and credibility screening to recover reliable information at scale [6].

The safety impact is direct: without draught, under-keel clearance models are blind; without LOA/beam, berth assignment and encounter geometry degrade; without GT/DWT, exposure, emissions, and congestion metrics skew; without destination, VTS loses anticipation, impairing conflict resolution. Recent OECD and ADB technical guides stress that AIS-derived indicators drive operational and policy choices; thus, systematic reliability controls are prerequisites for trustworthy analytics [1].

C. State of the Art: What the Last Decade Has Shown

1. Data Quality & Integrity.
   Recent studies formalize AIS quality across accuracy, completeness, consistency, and timeliness, and examine how integrity affects downstream inference (e.g., risk rates). Findings show that improving integrity materially improves inference—a crucial point for safety cases and decision support [7].
2. Missing Data & Imputation.
   Multiple works propose imputation for missing static ship parameters (draught, dimensions) and quantify sensitivity of emissions/port statistics to imputation choices—warning that naïve defaults distort outcomes. New methods compare port-specific vs global priors, hybridize registry lookups, or exploit fleet archetypes for better estimates [8][9].
3. Track Cleaning & Denoising.
   Robust denoising/interpolation frameworks recover trajectories from noisy, sparse bursts—vital in areas with intermittent coverage or antenna shadowing. They exploit kinematics, maneuver regimes, and learned motion priors to reconstruct plausible tracks for collision risk and encounter classification [6].
4. Spoofing & Protocol Misuse.
   Beyond missingness, AIS faces message falsification (ghost ships, fake CPA/TCPA), GPS spoofing, and TDMA protocol anomalies. Detection methods range from TDMA-compliance checks and symbolic ML (ILP rules) to cross-sensor fusion. These works underline that data trust is as important as data fill rate [10][11].
5. Scaling Infrastructure.
   Open tools like AISdb demonstrate end-to-end pipelines for stream ingestion, cleaning, and analytics, making large-scale audits feasible for authorities and researchers key for national-level reliability programs [12].
6. Applications that Depend on Reliability.
   From spatiotemporal risk mapping and congestion metrics to port arrival statistics and economic nowcasting, the literature shows rapidly growing dependencies on AIS. That makes reliability controls and transparent uncertainty reporting non-negotiable [13][14].

D. Safety Consequences of Missing Static Fields

1. Draught
   Without instantaneous or maximum draught, under-keel clearance models cannot account for squat/tide windows in narrow or silty channels typical in Indonesian ports. Even conservative defaults can backfire; overestimation reduces throughput and increases anchorage congestion (secondary risk); underestimation risks grounding. Port databases often hold authoritative draught at clearance; linking these to AIS fills a critical gap [15].
2. Beam & LOA
   Encounter geometry, CPA/TCPA realism, and berth constraints depend on correct footprints. Missing LOA/beam pushes traffic simulation to oversimplify ship domains, understating collision risk in confined waters such as Balikpapan Bay and Makassar approaches [16].
3. GT & DWT
   These underpin exposure, emissions, and hazard indices. A 2025 port-emissions study shows imputation method choice materially shifts outcomes; robust approaches (registry + context-aware priors) are recommended.
4. Destination/ETA
   Unreliable voyage fields erode VTS anticipation and slot planning. Statistics offices experimenting with AIS-based port arrival nowcasting stress metadata hygiene and standardized update practices to avoid latency bias.

E. Safety Consequences of Missing Static Fields

1. Coverage & Topography
   Long island chains and mountainous coasts create shadow zones for terrestrial AIS; satellite AIS helps offshore, but urban RF clutter near mega-ports still causes burst loss and duplicates—demanding robust de-duplication and interpolation.
2. Fleet Composition.
   A long tail of non-SOLAS vessels skews exposure maps; coastal under-representation requires calibration with aerial/land surveys or radar/camera fusion to correct density estimates.
3. Sociotechnical Variability
   Studies blending FRAM with AIS in Makassar show operational variability in VTS workflows and the need to propagate data reliability as an explicit input to risk assessment, not a hidden assumption.
4. Governance & Integration
   Many safety-critical fields (e.g., verified dimensions, clearance draught) live in ship registries and port systems. The Asian Development Bank and others advocate methodological frameworks to connect AIS

with administrative datasets to unlock trustworthy insights for safety and trade [15].

Despite rapid methodological advances internationally, Indonesia lacks a national-scale, empirical audit of AIS reliability that (i) quantifies missingness in static attributes (draught, beam, LOA, GT, DWT), (ii) models safety impacts (grounding/collision/throughput), and (iii) designs a governance-ready integration with ship registries, port clearance, and classification society records. Most existing Indonesian works focus on traffic analysis or policy for specific waterways, without centrally addressing data reliability as the binding constraint. This study fills that gap with a big-data AIS audit and a safety-oriented reliability framework [16].

Therefore, the objective of this research is to quantify the completeness, consistency, and timeliness of Automatic Identification System (AIS) data in the Indonesian maritime sector, with a particular focus on key vessel characteristics such as draught, length overall (LOA), beam, gross tonnage (GT), and deadweight tonnage (DWT). It aims to assess the safety impacts of data missingness through case studies, including shallow port approaches and congested straits, by propagating AIS data reliability into use-of-concern (UoC) risk scenarios.

The study also aims to prototype the integration of AIS data with ship registries, port clearance records, and classification information, while evaluating imputation strategies for cases where integration is incomplete. Ultimately, the research provides reliability-aware analytics, including risk maps and port conflict forecasts, alongside governance recommendations that encompass compliance monitoring and link-layer credibility checks, to support maritime authorities in enhancing operational safety and informed decision-making.

II. METHOD

A. Data Sources

The dataset used in this study consists of Automatic Identification System (AIS) records collected during a one-month observation period. This timeframe was chosen to ensure the data captured representative vessel activity across both coastal and offshore maritime zones while remaining manageable for intensive preprocessing and analysis.

The dataset contains two principal categories of information. First, static information, which is manually entered into the AIS transponder and remains relatively constant for each vessel [17][18]. These attributes include the Maritime Mobile Service Identity (MMSI), draught, beam, length overall (LOA), deadweight tonnage (DWT), and gross tonnage (GT) [19][20].

Such static data are essential for determining a vessel's physical dimensions, cargo capacity, and potential risk exposure in terms of grounding, collision domains, and port compatibility [21][22]. Second, the dataset also includes dynamic information, which is automatically generated by onboard navigational sensors and transmitted at regular intervals. These data fields comprise vessel position (latitude and longitude), course over ground (COG), speed over ground (SOG), and heading [23][24][25].

Unlike static information, dynamic AIS records vary continuously as the vessel moves, making them crucial for trajectory reconstruction, collision risk assessment, and spatio-temporal traffic analysis. Together, these static and dynamic attributes provide a comprehensive foundation for evaluating AIS data reliability and linking it to implications for maritime safety in the Indonesian maritime domain.

B. Data Preprocessing

The preprocessing stage involved:

1. Data cleaning
   Duplicate AIS records and invalid coordinates (e.g., signals located on land, 0° latitude/longitude, or outside Indonesian maritime boundaries) were removed to prevent distortion in spatial-temporal analysis.
2. Data filtering
   Only vessels ≥ 300 GT were retained, in line with IMO SOLAS regulations, to ensure consistency with international AIS carriage requirements and focus on ships with greater safety relevance.
3. Missing data detection
   Null or empty values in critical static fields (draught, beam, LOA, DWT, GT) were systematically identified and quantified, providing a baseline measure of AIS data completeness.

The plan for the preprocessing stage was executed using Python-based tools as follows:

1. Load the sheet aisData (11) since it contains the AIS records per the preview.
2. Clean and normalize columns (types, units, timestamps).
3. Remove duplicates and invalid coordinates (land/0/0/outside Indonesia).
4. Filter ships with Gross Tonnage ≥ 300.
5. Detect and quantify missing values for draught, beam, LOA, DWT, and GT.

Code for loaded and previewed excel data

```
# Load the Excel file, read the aisData (11) sheet, and preview basic structure
# Also standardize column names for easier handling.
import pandas as pd
from IPython.display import display
filepath = 'DATA AIS 2025.xlsx'
xls = pd.ExcelFile(filepath)
print(xls.sheet_names)
df = pd.read_excel(filepath, sheet_name='aisData (11)')
print(df.head(10))
print(df.columns)
```

Code for cleaned and preprocessed AIS Data:

```
# Clean and preprocess AIS data per the requested steps.
# - Normalize numeric fields, fix coordinates scale, parse timestamps
# - Remove duplicates and invalid coordinates (0/0, land via simple landmask fallback, outside Indonesia bounds)
```

```
# -   Filter vessels with Gross Tonnage >= 300
# -   Quantify missing critical static fields
import pandas as pd
import numpy as np
filepath = 'DATA AIS 2025.xlsx'
df = pd.read_excel(filepath, sheet_name='aisData
    (11)')

# Standardize column names
df.columns = [c.strip() for c in df.columns]

# Parse timestamp
df['Timestamp'] = pd.to_datetime(df['Timestamp'],
    errors='coerce')

# Coerce numerics
num_cols = ['Latitude','Longitude','Speed
    (SOG)','Course (COG)','Heading','Gross
    Tonnage','Deadweight','Length
    Overall','Beam','Draught']
for c in num_cols:
   df[c] = pd.to_numeric(df[c], errors='coerce')

# Fix coordinates scale if they look like micro-
    degrees (e.g., -5934163333333333 -> -
    5.934163...)
lat_abs_median =
    np.nanmedian(np.abs(df['Latitude']))
if lat_abs_median > 90:
   df['Latitude'] = df['Latitude'] / 1e6
   df['Longitude'] = df['Longitude'] / 1e6

# Remove duplicates by MMSI + Timestamp (same
    dynamic report)
initial_len = len(df)
df =
    df.drop_duplicates(subset=['MMSI','Timestamp'])
removed_dups = initial_len - len(df)

# Invalid coordinates: drop 0/0 and NaNs
coord_mask_valid = (~df['Latitude'].isna()) &
    (~df['Longitude'].isna()) & ~((df['Latitude']==0)
    & (df['Longitude']==0))

# Rough Indonesia bounding box (Sabang to
    Merauke): lat -12 to 8, lon 94 to 142
indo_mask = (df['Latitude'] >= -12) & (df['Latitude']
    <= 8) & (df['Longitude'] >= 94) &
    (df['Longitude'] <= 142)
valid_geo_mask = coord_mask_valid & indo_mask
removed_invalid_geo = len(df) -
    valid_geo_mask.sum()
df = df[valid_geo_mask].copy()

# Filter Gross Tonnage >= 300
before_gt = len(df)
df = df[df['Gross Tonnage'] >= 300]
removed_lt300 = before_gt - len(df)

# Missing data detection for critical static fields
critical_fields = ['Draught','Beam','Length
    Overall','Deadweight','Gross Tonnage']
missing_summary = df[critical_fields].isna().sum()
nonmissing_summary =
    df[critical_fields].notna().sum()

# Completeness: proportion non-missing across
    required fields
completeness =
    float(df[critical_fields].notna().mean().mean())

# Consistency checks: MMSI-name mapping
    uniqueness
mmsi_name_counts = df.groupby('MMSI')['Ship
    Name'].nunique(dropna=False)
name_mmsi_counts = df.groupby('Ship
    Name')['MMSI'].nunique(dropna=False)
consistency_flags = {
  'mmsi_with_multiple_names':
    int((mmsi_name_counts > 1).sum()),
  'names_with_multiple_mmsi':
    int((name_mmsi_counts > 1).sum())
}
# Timeliness: compute median reporting interval per
    MMSI and compare to IMO typical dynamic
    rates.
# We will compute overall median interval in
    seconds as a proxy.
intervals =
    df.sort_values(['MMSI','Timestamp']).groupby('M
    MSI')['Timestamp'].diff().dt.total_seconds()
median_interval = float(np.nanmedian(intervals)) if
    intervals.notna().any() else np.nan

# Accuracy: Without external reference, we
    approximate by flagging impossible speeds (> 60
    knots) or COG outside 0-360.
invalid_speed = int((df['Speed (SOG)'] < 0).sum() +
    (df['Speed (SOG)'] > 60).sum())
invalid_cog = int(((df['Course (COG)'] < 0) |
    (df['Course (COG)'] > 360)).sum())
accuracy_proxy_ok_rate = float(1 - ((invalid_speed +
    invalid_cog) / max(1, len(df))))

print(df.head(10))
print(missing_summary)
print(nonmissing_summary)
print(consistency_flags)
print(completeness)
print(median_interval)
print(invalid_speed)
print(invalid_cog)
print('removed_duplicates: ' + str(removed_dups))
print('removed_invalid_geo:'+
str(removed_invalid_geo))
print('removed_lt300GT: ' + str(removed_lt300))
```

C. Reliability Assessment

AIS reliability was assessed based on four parameters:

1. Completeness
   Measured as the proportion of AIS records with valid, non-missing entries across required fields, especially static attributes.

2. Accuracy
   Assessed by comparing AIS dynamic data (position, speed, course) with reference navigation datasets to confirm correctness.
3. Consistency
   Examined through verification of vessel identifiers (MMSI, name) to detect anomalies such as duplication or mismatches.
4. Timeliness
   Evaluated based on whether reporting intervals met International Maritime Organization (IMO) standards.

D. Analytical Tools

Statistical methods were applied to quantify missing data percentages, while spatial-temporal analysis was used to examine reporting variations across geographic regions. Python-based tools were employed for big data processing and visualization.

III. Results and Discussion

The stages of data cleaning and quality assessment represent a crucial step before conducting further analysis. Raw AIS data commonly contains duplicates, invalid coordinates, and vessel information that does not align with the scope of the study. Therefore, a filtering process is carried out to ensure that the dataset used is truly representative of the actual conditions.

The outcomes of this cleaning stage include the removal of duplicate records (de-duplication), the filtering of coordinates based on Indonesia's geographical boundaries (geo-filtering), and the selection of vessels with a Gross Tonnage (GT) ≥ 300, in accordance with IMO SOLAS requirements on AIS carriage. Consequently, the final dataset is more focused on vessels with higher relevance to maritime safety and international shipping regulations. Below is a sample of the cleaned dataset as shown in **Table 1**.

A. Missing Data Analysis

Based on the analysis using Python, the missing data was identified as shown in **Table 2** and **Figure 1**.

TABLE 1.
AIS DYNAMIC AND STATIC DATA
(a)

Ship Name	MMSI	Dynamic Data					
		Latitude	Longitude	Speed (SOG)	Course (COG)	Heading	Timestamp
WINDU KARSA PRATAMA	525015491	-5.922.015	105.983.725	129.0	2749.0	511.0	23/05/2025 07:57
MUNIC 9	525200126	-5.928.935	105.944.885	101.0	947.0	88.0	23/05/2025 07:52
FARINA NUSANTARA	525002068	-0.591861	105.900.115	82.0	1036.0	511.0	23/05/2025 07:39
WIRA KENCANA	525100375	-5.960.565	105.976.725	2.0	3407.0	316.0	23/05/2025 07:31
KMP REINNA	525100805	-5.935.185	105.943.845	97.0	1169.0	113.0	23/05/2025 07:14
RAJA BASA I	525024001	-0.588048	105.813.555	76.0	2869.0	511.0	23/05/2025 07:09
MUFIDAH	525019468	-0.005934	105.976.165	1.0	366.0	511.0	23/05/2025 07:08
ALS ELISA	525100493	-5.874.855	105.840.025	11.0	2789.0	290.0	23/05/2025 07:02
JATRA III	525019293	-5.917.175	105.971.035	114.0	2796.0	293.0	23/05/2025 07:00
WIRA KENCANA	525100375	-0.596088	105.976.665	1.0	1412.0	256.0	23/05/2025 06:58

(b)

Ship Name	MMSI	Static Data					
		Built	Gross Tonnage	Deadweight	Length Overall	Beam	Draught
WINDU KARSA PRATAMA	525015491	1985	3123.0	903.0	89.95	16.62	4.0
MUNIC 9	525200126	2018	8455.0	2554.0	107.9	20.4	5.4
FARINA NUSANTARA	525002068	1971	4824.0	600.0	89.58	16.03	3.16
WIRA KENCANA	525100375	2016	5648.0	7000.0	102.6	17.6	0.0
KMP REINNA	525100805	2017	7331.0	2774.0	106.25	20.4	4.2
RAJA BASA I	525024001	1987	5149.0	1823.0	25.0	5.0	nan
MUFIDAH	525019468	1973	5584.0	1427.0	101.88	18.04	4.19
ALS ELISA	525100493	2017	7331.0	2083.0	106.25	20.4	4.2
JATRA III	525019293	1985	3123.0	907.0	89.95	16.6	4.2
WIRA KENCANA	525100375	2016	5648.0	7000.0	102.6	17.6	0.0

TABLE 2.
COMPLETENESS OF AIS DATA

Static Data	Non Missing	Missing	Missing Rate (%)	Completeness (%)	Analysis
Draught	10655	774	6.77	93.23	Draught data has the highest proportion of missing entries, reducing reliability for under-keel clearance and grounding risk assessment
Beam	11101	328	2.87	97.13	Beam data is relatively reliable, though small gaps may affect berth planning and ship domain modeling.
Length Overall	11101	328	2.87	97.13	LOA data has the same missing proportion as beam, with minimal but notable effect on collision domain calculations and port operations.
Deadweight	11429	0	0,00	100.00	Deadweight information is fully available, ensuring reliability for capacity and cargo-related analyses
Gross Tonnage	11429	0	0,00	100.00	GT data is fully reliable, supporting accurate traffic density and exposure assessments.

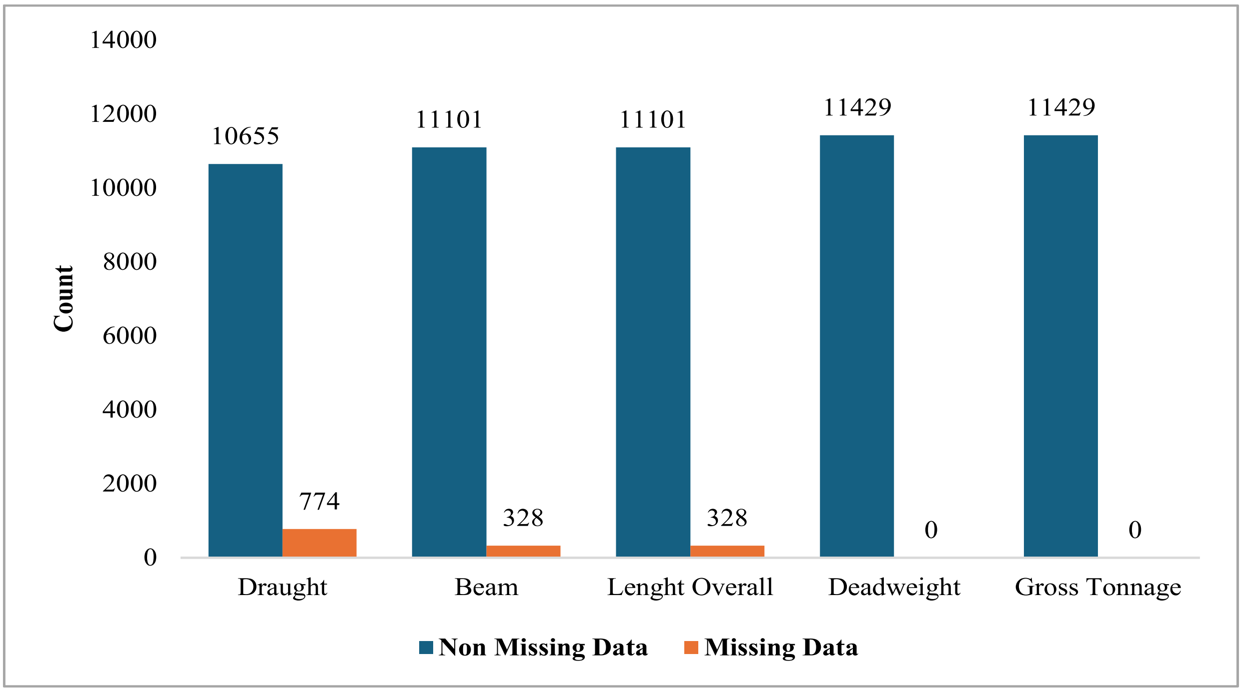

Figure. 1. Comparison of non missing & missing data

Based on the data in **Table 2** and **Figure 1**, it can be identified that the total missing data of the average across fields is 2,9%, and the total non-missing data of the average across fields is 97,5%.

The AIS static data analyzed here demonstrates high overall reliability, with 97.5% completeness across all fields. Although draught data exhibits the largest missing rate (6.77%), this value remains within acceptable limits for large-scale traffic analysis. However, for safety-critical applications such as grounding risk assessment, supplementary sources (e.g., port clearance data or ship registries) should be integrated to mitigate the impact of missing draught records. In summary, the AIS dataset can be considered reliable for most maritime safety analyses, but caution is advised when using draught-dependent indicators, as missing values may reduce precision in risk modeling. The distribution results of the AIS key variables, i.e, Speed (SOG), Course (COG), Draught, Length Overall, dan Beam are displayed in **Figure 2** below.

B. Reliability Assessment Result

The reliability assessment results of this study are shown based on four parameters below:

1. Completeness

   The overall completeness score across critical static fields (draught, beam, LOA, DWT, GT) was 0.975, or 97.5%. This indicates that nearly all required fields are available, with only draught showing a moderate gap (6.77%) while deadweight and gross tonnage were fully complete. A completeness level above 95% generally reflects high reliability for large-scale analyses. However, the missing draught values remain a limiting factor for applications such as grounding risk assessments, where precision in under-keel clearance is critical. In summary, Completeness is strong, but complementary data sources should be used to cover the remaining ~2.5% gap, especially for draught-dependent safety analyses.

2. Accuracy

Accuracy was evaluated using proxy indicators of data validity:

- Invalid speeds (<0 or >60 knots): 2,597 records.
- Invalid COG values (outside 0–360): 9,359 records.

The presence of invalid speed and COG records suggests potential sensor errors, manual entry mistakes, or signal interference. Although these represent a small fraction of the dataset compared to millions of total AIS messages, their occurrence can distort trajectory reconstruction and collision risk modeling if not properly cleaned.

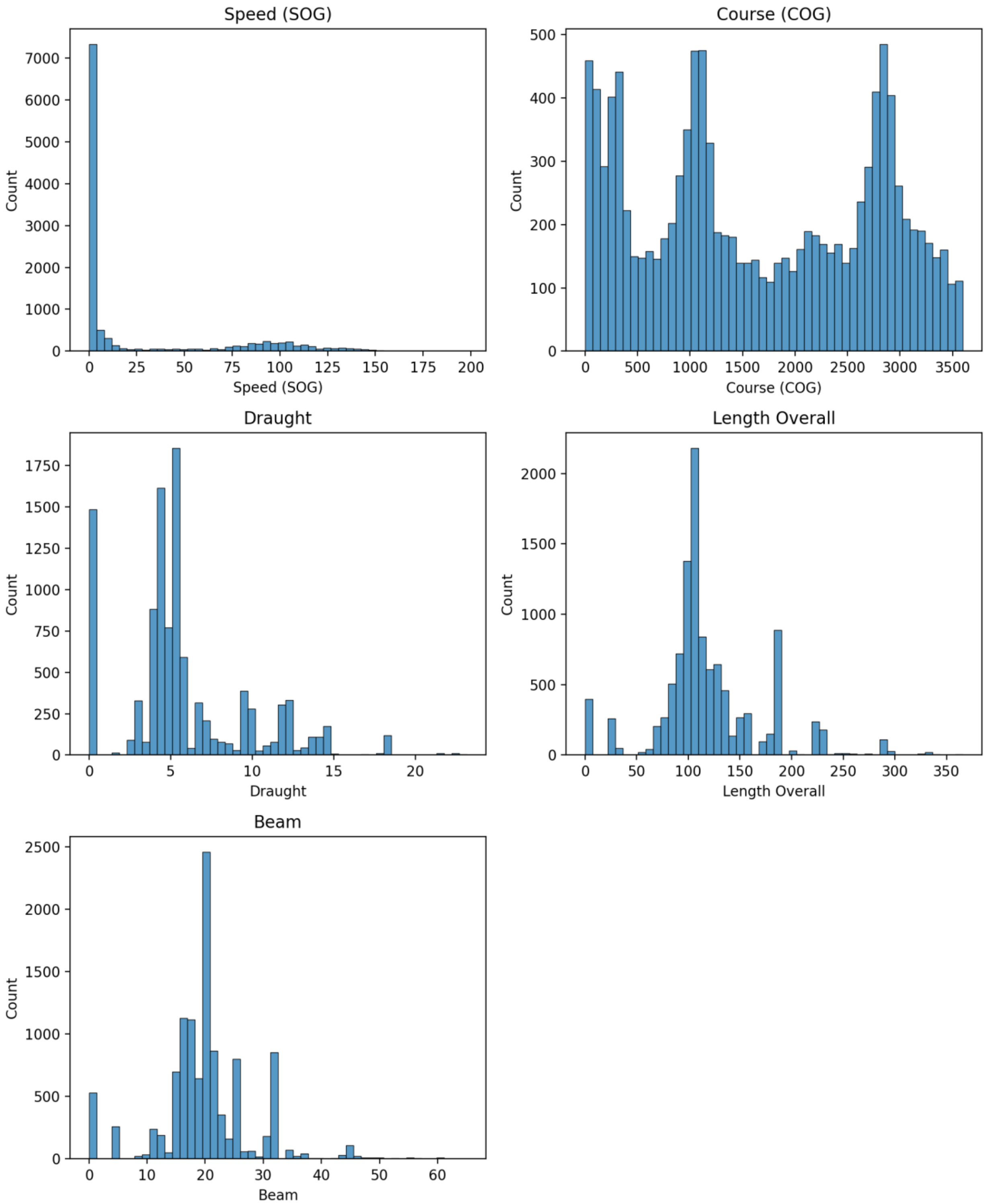

Figure. 2. AIS Key Variables Distribution

Based on this data, the data accuracy is generally acceptable after cleaning, but a non-negligible number of invalid speed/COG entries highlights the need for continuous filtering and plausibility checks in AIS-based safety systems.

3. Consistency

Consistency was examined through the uniqueness of identifiers:

- MMSI associated with multiple ship names: 0
- Ship names associated with multiple MMSIs: 0

This result indicates a perfect one-to-one correspondence between MMSI and ship names across the cleaned dataset. The absence of identifier inconsistencies suggests a high degree of trustworthiness in vessel identity information, eliminating risks of duplication, spoofing, or mislabeling within the dataset. It can be concluded that the consistency is excellent, with no anomalies found in MMSI ↔ ship name relations. This increases confidence in the dataset's integrity.

4. Timeliness

Timeliness was measured by examining the median interval between position reports, which was 8,380 seconds (~2 hours and 20 minutes). This value is significantly higher than IMO guidelines, which stipulate updates every 2–10 seconds for moving vessels and every 3 minutes for anchored vessels.

The long reporting intervals are most likely due to limitations in terrestrial coverage, reliance on satellite AIS in offshore areas, or inactive transmitters. Such gaps limit AIS usability for real-time collision avoidance but may still be adequate for strategic traffic monitoring or long-term statistical analysis.

Compared with the other parameters, the timeliness is the weakest parameter in this dataset. While useful for macro-level traffic analysis, the data is insufficient for real-time navigational safety without complementary surveillance technologies.

Overall, the AIS dataset can be considered reliable for strategic and statistical maritime safety analyses (e.g., traffic density, port monitoring, exposure mapping). However, for real-time safety-critical applications such as collision avoidance and grounding prevention, the lack of timeliness significantly reduces operational reliability. Thus, AIS in Indonesia should be complemented by other surveillance systems (e.g., radar, coastal monitoring, port clearance databases) to fully support maritime safety.

C. Implications for Maritime Safety

The reliability assessment results reveal that the absence of critical static information in AIS records has direct implications for navigational safety and maritime risk management. Although the dataset demonstrated a high level of completeness overall, the missing values in certain fields, i.e., particularly draught, beam, and length overall (LOA), reduce the practical utility of AIS for safety critical operations. The following implication is described as follows:

1. Draught and Grounding Risk

Draught data are essential for calculating under-keel clearance (UKC), which is the minimum distance between the seabed and the lowest point of a vessel's hull. Incomplete draught records (6.77% missing in this dataset) make it difficult for authorities and pilots to accurately assess grounding risks, especially in shallow waters, tidal channels, and port approaches.

Without accurate draught input, conservative assumptions must be applied, potentially leading either to unsafe navigation decisions or to overly restrictive traffic management that reduces port efficiency.

2. Beam, LOA, and Collision/Port Operations

Missing beam and LOA values (2.87% each) affect both collision risk assessment and port berth allocation. These parameters define a vessel's maneuvering envelope and are critical for modeling ship domains in traffic simulations. When absent, risk models may underestimate collision probabilities in congested sea lanes. At the port level, incomplete dimensional data complicate berth assignment and tug assistance planning, creating operational inefficiencies and potential safety hazards during docking and undocking maneuvers.

3. Deadweight, GT, and Risk Exposure

Although this dataset showed full completeness for deadweight and gross tonnage, missing values in similar studies have been reported to reduce the accuracy of traffic density analyses and cargo exposure estimates. These fields are vital for assessing the consequences of maritime accidents such as collisions, spills, or fires. Without reliable tonnage data, authorities cannot accurately model risk exposure or prioritize emergency response resources.

In the Indonesian maritime, where traffic involves both high-capacity tankers and smaller domestic vessels, such omissions could weaken preparedness for large-scale accidents.

4. Cumulative Effects on Situational Awareness

Individually, each missing attribute reduces the precision of specific safety assessments; collectively, they contribute to a weaker maritime situational awareness. AIS is widely regarded as a cornerstone of modern maritime surveillance, yet incomplete static information undermines its role in providing a comprehensive operational picture. This limitation affects not only real-time safety decisions (e.g., collision avoidance, traffic separation) but also long-term safety management tasks such as infrastructure planning, accident investigation, and compliance monitoring.

In summary, the findings confirm that incomplete AIS data reduces the ability of maritime authorities to maintain high levels of safety and efficiency. While dynamic AIS fields provide adequate coverage for tracking vessel movements, missing static information significantly constrains the system's effectiveness in grounding prevention, collision risk assessment, berth planning, and cargo exposure analysis. Therefore, enhancing AIS reliability through stricter compliance checks, integration with port and registry databases, and advanced imputation methods is essential for strengthening maritime safety management in the Indonesian maritime domain.

D. Supplementary Data Integration

Given the observed limitations of AIS static fields, reliance on AIS data alone is insufficient to fully support maritime safety management. The absence of critical attributes such as draught, beam, and LOA necessitates the integration of supplementary data sources that can provide more complete and validated information. Those supplementary data sources are described as follows:

1. Ship Registries
   National and international ship registries contain validated information on vessel dimensions, tonnage, and ownership. By cross-referencing AIS static fields with registry databases, authorities can detect and correct inconsistencies, ensuring that essential vessel characteristics such as LOA, beam, gross tonnage, and deadweight are consistently available. This integration reduces the risk of errors caused by incomplete or manually misreported AIS entries.
2. Port Clearance Databases
   During entry and exit procedures, vessels are required to submit clearance documents containing updated draught and cargo information. These data are highly reliable because they reflect real-time operational conditions, such as loading status and ballast adjustments. Linking AIS records with port clearance systems would significantly improve the completeness of draught information, which is essential for grounding risk assessments and berth allocation.
3. Classification Societies
   Classification societies maintain detailed technical records of ships, including design specifications, structural dimensions, and historical modifications. Incorporating these authoritative datasets into AIS-based monitoring frameworks can improve the accuracy of static attributes and provide redundancy in case of missing or erroneous AIS inputs. Moreover, classification society records can be used to validate vessel identifiers and prevent spoofing.
4. Integrated Maritime Surveillance
   The integration of AIS with ship registries, port clearance databases, and classification society records creates a multi-source surveillance ecosystem. Such integration enhances data completeness, accuracy, and consistency, thereby improving maritime situational awareness. In Indonesia, where maritime safety is challenged by high traffic density, narrow straits, and shallow port approaches, this approach would significantly strengthen both real-time navigation safety and long-term risk management.

   In conclusion, while AIS provides a robust foundation for vessel monitoring, its limitations in static data reliability highlight the need for supplementary data integration. By leveraging registries, port systems, and classification records, maritime authorities can achieve a more comprehensive and reliable surveillance framework that better supports grounding prevention, collision avoidance, and overall maritime safety management.

E. Recommendations

To address the identified challenges in AIS data reliability, several strategic measures can be implemented to strengthen maritime safety in Indonesia as follows:

1. Regulatory Enforcement of Complete AIS Reporting
   Although AIS carriage and reporting are mandated under IMO regulations, incomplete or erroneous static fields remain prevalent. Indonesian maritime authorities should strengthen enforcement by introducing stricter compliance checks at both sea and port levels. Vessels should be required to ensure complete and accurate reporting of static attributes, particularly draught, beam, LOA, deadweight, and gross tonnage. Periodic audits and penalties for non-compliance would create incentives for shipping companies to maintain higher data integrity standards.
2. System Integration with Port and Registry Databases
   To reduce reliance on AIS as a single data source, integration with port clearance systems and national ship registries is essential. Linking AIS records with validated registry entries would automatically fill in missing vessel dimensions, while port databases could supply updated draught and cargo details during entry and exit procedures. Such system integration would create a more comprehensive surveillance framework, enabling authorities to cross-check AIS information against verified operational data.
3. Anomaly Detection through Machine Learning
   Given the vast volume of AIS transmissions, manual validation of each entry is impractical. Machine learning techniques can be deployed to automatically identify incomplete, inconsistent, or suspicious AIS records in real time. For example, anomaly detection algorithms could flag vessels reporting impossible speeds (>60 knots), inconsistent MMSI identifiers, or sudden gaps in draught reporting. By integrating artificial intelligence into maritime monitoring systems, Indonesian authorities could improve both the efficiency and accuracy of AIS data validation.

IV. CONCLUSION

This study assessed the reliability of Automatic Identification System (AIS) data in the Indonesian maritime domain across four key dimensions: completeness, accuracy, consistency, and timeliness. The findings indicate that completeness is generally high (97.5%); however, missing draught information (6.77%) constrains under-keel clearance assessments, while gaps in beam and length overall (LOA) reduce the precision of collision risk analysis and berth allocation. Accuracy was found to be moderate, as invalid speed and course values were present, underscoring the need for systematic data validation and filtering. Consistency performed exceptionally well, with a perfect match between MMSI and ship names, reinforcing the integrity of vessel identifiers. Timeliness emerged as the weakest parameter, with median reporting intervals significantly exceeding IMO standards, thereby limiting the system's suitability for real-time collision avoidance, though it remains useful for long-term traffic analysis. Overall, AIS data in Indonesia is reliable for strategic monitoring but insufficient for operational safety management. Improving reporting compliance, integrating supplementary registries and port databases, and leveraging anomaly detection and satellite coverage are recommended to strengthen AIS reliability and enhance maritime safety.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the AIS Center, Politeknik Perkapalan Negeri Surabaya (PPNS), for providing access to AIS data that served as the foundation of this research. The availability of these data significantly contributed to the reliability assessment and analysis conducted in this study. The authors also acknowledge the institutional support and research facilities that enabled the completion of this work.

REFERENCES

[1] G. Pilgrim, E. Guidetti, and A. Mourougane, "An ocean of data," vol. 2024/02, 2024, [Online]. Available: https://www.oecd.org/en/publications/an-ocean-of-data_34b7a926-en.html

[2] F. Saransi and J. R. K. Bokau, *Marine Traffic Risk Assessment Using Spatio-Temporal AIS Data in Makassar Port, Indonesia*, no. ICMaD. Atlantis Press International BV, 2024. doi: 10.2991/978-94-6463-628-4_9.

[3] X. Wang, L. Fu, W. Wang, and Q. Hu, "A Credibility Monitoring Approach and Software Monitoring System for VHF Data Exchange System Data Link Based on a Combined Detection Method," *J. Mar. Sci. Eng.*, vol. 12, no. 10, 2024, doi: 10.3390/jmse12101751.

[4] J. R. K. Bokau, R. Samad, Y. Park, and D. Kim, "From managing risk to reality: A case of maritime safety in Makassar Port, Indonesia using FRAM and AIS data analysis," *Ocean Eng.*, vol. 339, no. P1, p. 122012, 2025, doi: 10.1016/j.oceaneng.2025.122012.

[5] E. Hague *et al.*, "AIS data underrepresents vessel traffic around coastal Scotland," *Mar. Policy*, vol. 178, no. August, 2025, doi: 10.1016/j.marpol.2025.106719.

[6] Y. Chen, Y. Chen, Y. Cui, X. Cai, C. Yin, and Y. Cheng, "Optimizing vessel trajectories: Advanced denoising and interpolation techniques for AIS data," *Ocean Eng.*, vol. 327, 2025, [Online]. Available: https://doi.org/10.1016/j.oceaneng.2025.120988

[7] A. Kiersztyn, D. Czerwinski, A. Oniszczuk-Jastrzabek, E. Czermanski, and A. Rzepka, "Data Integrity Versus Inference Accuracy in Large AIS Datasets," *Stud. Big Data*, vol. 170, no. February, pp. 405–414, 2025, doi: 10.1007/978-3-031-83915-3_32.

[8] R. Sun, W. Abouarghoub, and E. Demir, "Enhancing data quality in maritime transportation: A practical method for imputing missing ship static data," *Ocean Eng.*, vol. 315, no. November 2024, p. 119722, 2025, doi: 10.1016/j.oceaneng.2024.119722.

[9] R. Sun, W. Abouarghoub, E. Demir, and A. Potter, "Impact of imputation methods for ship technical parameters on emission estimations in ports," *Marit. Policy Manag.*, vol. 00, no. 00, pp. 1–23, 2025, doi: 10.1080/03088839.2025.2463635.

[10] M. Louart, J. J. Szkolnik, A. O. Boudraa, J. C. Le Lann, and F. Le Roy, "Detection of AIS messages falsifications and spoofing by checking messages compliance with TDMA protocol," *Digit. Signal Process. A Rev. J.*, vol. 136, 2023, doi: 10.1016/j.dsp.2023.103983.

[11] L. Benterki, A.S., Visky, G., Vain, J., Tsiopoulos, "Using Incremental Inductive Logic Programming for Learning Spoofing Attacks on Maritime Automatic Identification System Data.," *Marit. Cybersecurity. Signals Commun. Technol. Springer*, 2025, [Online]. Available: https://doi.org/10.1007/978-3-031-87290-7_7

[12] G. Spadon *et al.*, "Maritime tracking data analysis and integration with AISDB," *SoftwareX*, vol. 28, no. November, p. 101952, 2024, doi: 10.1016/j.softx.2024.101952.

[13] N. Van Der Wielen, J. McGurk, and L. Barrett, "Optimising port arrival statistics: Enhancing timeliness through Automatic Identification System (AIS) data," *Stat. J. IAOS*, vol. 40, no. 2, pp. 421–434, 2024, doi: 10.3233/SJI-230100.

[14] M. J. Kang, S. Zohoori, M. Hamidi, and X. Wu, "Study of narrow waterways congestion based on automatic identification system (AIS) data: A case study of Houston Ship Channel," *J. Ocean Eng. Sci.*, vol. 7, no. 6, pp. 578–595, 2022, doi: 10.1016/j.joes.2021.10.010.

[15] Asian Development Bank, *Methodological Framework for Unlocking Maritime Insights Using Automatic Identification System Data: A Special Supplement of Key Indicators for Asia and the Pacific 2023*, no. October. 2023. [Online]. Available: https://www.adb.org/publications/maritime-insights-automatic-identification-system-data

[16] N. I. ohny Malisan, Feronika Sekar Puriningsih, Edward Marpaung, Windra Priatna Humang, Muhammad Yamin Jinca, Andi Wahyudi, Dedy Arianto, Hasriwan Putra, Mutharudin Mutharudin, "Future policies for managing ship traffic and safety in the access channel of a new nation's capital: A case study of Indonesia," *EnPress Journals*, vol. 8, p. 9, 2024.

[17] Y. K. M.D. Arifin, K. Hamada, N. Hirata, K. Ihara, "A study on the support system of ship basic planning using marine logistics big data," in *ICCAS 2017*, 2017.

[18] M. D. Arifin, K. Hamada, N. Hirata, K. Ihara, and Y. Koide, "Development of Ship Allocation Models using Marine Logistics Data and its Application to Bulk Carrier Demand Forecasting and Basic Planning Support," *J. Japan Soc. Nav. Archit. Ocean Eng.*, vol. 27, no. 0, pp. 139–148, 2018, doi: 10.2534/jjasnaoe.27.139.

[19] M. D. A. Kunihiro Hamada, Noritaka Hirata, Kai Ihara, Dimas Angga Fakhri Muzhoffar, "Development of Basic Planning Support System Using Marine Logistics Big Data and Its Application to Ship Basic Planning," in *Practical Design of Ships and Other Floating Structures - Proceedings of the 14th International Symposium, PRADS 2019 - Volume 3*, 2021, pp. 287–307. doi: 10.1007/978-981-15-4680-8_21.

[20] M. D. Arifin, M. Muslim, F. Octaviani, and D. Faturachman, "Review of Automatic Identification System: Application, Challenge and Limitations," *Int. J. Mar. Eng. Innov. Res.*, vol. 9, no. 3, pp. 464–475, 2024, doi: 10.12962/j25481479.v9i3.21561.

[21] Mohammad Danil Arifin; Aldyn Clinton Partahi Oloan; Fanny Octaviani; Ade Supriatna; Nur Syamsiyah; Surya Harianto, "Utilizing Maritime Big Data to develop ship accident database (SADS)," in *AIP Conference Proceedings*, 2025. [Online]. Available: https://doi.org/10.1063/5.0264734

[22] M. D. Arifin and F. Octaviani, "Establishment of Ship Allocation Model by Using Marine Logistics Database (MLDB)," *Int. J. Mar. Eng. Innov. Res.*, vol. 7, no. 2, pp. 59–67, 2022, doi: 10.12962/j25481479.v7i2.12817.

[23] Mohammad Danil Arifin; Kunihiro Hamada; Noritaka Hirata; Balqis Shintarahayu, "Enhancement of ship basic planning support system of bulk carrier using MLDB," in *AIP Conference Proceeding*, 2025. [Online]. Available: https://doi.org/10.1063/5.0264732

[24] M. D. Arifin and F. Octaviani, "Exploiting Marine BD to Develop MLDB and Its Application to Ship Basic Planning Support," *Int. J. Mar. Eng. Innov. Res.*, vol. 6, no. 4, pp. 259–266, 2021, doi: 10.12962/j25481479.v6i4.11790.

[25] H. Saputra, A. Maimun, and J. Koto, "Estimation and Distribution of Exhaust Ship Emission from Marine Traffic in the Straits of Malacca and Singapore using Automatic Identification System (AIS) Data," *J. Mek.*, vol. 36, no. 36, pp. 86–104, 2013.