

Dynamic Inference of Regional Traffic States in a Dense Waterway Using AIS Window Sequences

Enze Wu*

School of Information Engineering, Jiangsu Maritime Institute, China

* Corresponding author Email: wuenzenj@qq.com

Abstract: Regional vessel monitoring requires both a description of current traffic and an estimate of its near-term evolution. This study evaluates a lightweight approach to dynamic traffic-state inference using sequences of Automatic Identification System (AIS) observations. A lower Yangtze waterway case area is divided into spatial cells, and vessel-level observations are aggregated into 10-minute windows. The inference target is the proportion of observed vessels with mean speed below 2 kn in a future window. A gradient-boosted tree model combines current traffic summaries with two preceding windows. Experiments use previously extracted observations representing 1,806,306 AIS messages and 6,014 vessel identities over approximately 4.7 days, with chronological training, validation and test periods. On 7,094 test samples, the 10-minute forecast achieves a mean absolute error of 0.0719, compared with 0.0810 for persistence and 0.0762 for a current-window model. These correspond to relative reductions of 11.2% and 5.7%. Accuracy for three operational traffic states reaches 81.15%. The advantage in continuous prediction remains at 20- and 30-minute horizons, although abrupt changes remain difficult. The results support adding short histories to vessel-window monitoring in this case area. The inferred states describe observed speed composition and should not be interpreted as independently validated congestion or collision-risk levels.

Keywords: Automatic identification system; Dense waterway; Dynamic inference; Gradient boosting; Regional traffic state; Short-term prediction.

1. Introduction

Automatic Identification System (AIS) observations provide positions, identities and motion attributes that can support maritime traffic monitoring. Their applications include route estimation, anomaly detection, collision prediction and navigation planning [1]. For a traffic operator monitoring a busy waterway, an aggregate description can complement individual vessel tracks: it identifies locations where many observed vessels are moving slowly and shows whether this condition is likely to persist into the next observation window. The usefulness of such a description depends on a clear definition of the monitored quantity and a credible evaluation of its future estimates.

Current regional activity can be obtained with straightforward engineering. A stream processor can maintain records by Maritime Mobile Service Identity (MMSI), update a sliding or tumbling window, and associate the records with a spatial grid. A key-value store can then expose vessel counts and speed summaries. This capability provides the baseline for the present study. The research question is whether a short sequence of these summaries improves future regional-state estimates beyond simply carrying the current summary forward. Storage, deduplication and grid assignment are therefore treated as supporting operations, with no claim of methodological novelty.

AIS quality must be considered when constructing regional indicators. Emmens et al. discuss both the analytical opportunities and the limitations of AIS observations, including noise associated with equipment, external conditions and dense traffic [2]. Zhao et al. examine spatial and temporal errors and their preprocessing [3]. Consequently, a large number of messages cannot be interpreted directly as the same number of vessels. This study first averages speed within each vessel-cell-window and then gives each observed

vessel equal weight in the cell summary. The resulting statistic concerns the observed vessel population; it does not estimate vessels that failed to report.

Existing maritime studies address several related but distinct targets. Liu et al. represent traffic as a particle system and assess complexity through spatial and dynamic relationships [4]. Yu et al. review collision-risk assessment, hotspot detection and path planning in restricted waters [5]. Li et al. compare machine-learning and deep-learning methods for individual vessel trajectory prediction in different traffic environments [6]. At a larger spatial scale, March et al. use AIS observations to quantify changes in marine traffic occupancy [7]. These studies motivate regional analysis, but their targets do not provide a direct performance reference for predicting the speed composition of a grid cell. Comparisons in the present work therefore use methods that predict exactly the same observable quantity.

This paper presents an applied case study of dynamic regional inference from AIS window sequences. The contribution is a reproducible formulation and evaluation of a limited operational task: predicting the future low-speed vessel proportion using current and recent regional observations. We compare a training-period grid mean, persistence, a current-window learning model and a model with short history. Evaluation uses later observations that are excluded from model fitting and selection, and includes continuous errors, derived state agreement and failure-oriented subgroup analyses. The study does not introduce a new learning algorithm or claim that low speed alone measures congestion.

2. Data and regional representation

2.1. Case area and data provenance

The study rectangle spans 118.85–119.15° E and 32.10–

32.30° N in the lower Yangtze region. It was chosen during an earlier exploratory examination of traffic activity. It is a convenience case area rather than an official port or navigable-channel polygon. Results include all retained observations within the rectangle and may therefore include anchoring, berthing and maneuvering activity as well as through traffic. The term dense waterway refers to the concentration of observed vessel activity in this setting, without implying an independently measured capacity threshold.

The source collection covers approximately 21–26 June 2023. The present experiments use a locally retained intermediate dataset containing 507,480 vessel-cell records at five-minute resolution. These records represent 1,806,306 accepted dynamic AIS messages and 6,014 distinct MMSIs. The earlier extraction admitted message types 1, 2, 3, 18, 19 and 27, nine-digit MMSIs, positions inside the rectangle, and reported speeds between 0 and 40 kn. It removed five consecutive exact duplicate records and 5,330 records whose displacement from the previous accepted observation implied a speed above 50 kn over a positive interval of at most one hour. These are preprocessing heuristics rather than externally validated quality labels.

The intermediate extraction also required a course value in the inclusive interval 0–360°. A value of 360° can denote unavailable course; no course values or course-derived features are used in the present analysis. Because the original source was not locally readable during this reanalysis, preprocessing was not repeated against the full message stream. The scope of verification is therefore the retained summaries and their documented extraction, rather than a new end-to-end audit of every raw message. The timestamps used are recorded reception times in UTC.

2.2. Vessel weighting and traffic states

The rectangle is divided into 0.01° by 0.01° cells, approximately 0.94 km east–west and 1.11 km north–south at the study latitude. Non-overlapping 10-minute windows are used. When combining the original five-minute summaries, mean speeds for the same MMSI, cell and new window are weighted by the original message counts. This reconstructs the accepted-message mean speed for that vessel in the 10-minute window. Each vessel then contributes once to the regional statistics. A vessel visiting two cells contributes once to each visited cell; cell counts should therefore not be summed to estimate unique vessels in the entire region.

Let $V_{g,t}$ be the set of observed vessels in cell g during window t , and let $n_{g,t} = |V_{g,t}|$. For vessel i , let $v_{i,g,t}$ be its accepted-message mean speed. The regional low-speed proportion is

$$q_{g,t} = \frac{1}{n_{g,t}} \sum_{i \in V_{g,t}} \mathbf{1}(v_{i,g,t} < 2 \text{ kn}). \quad (1)$$

The mean and sample standard deviation of the vessel mean speeds are denoted by $\mu_{g,t}$ and $s_{g,t}$. Each eligible cell-window is described by (q, n, μ, s) . We require at least five observed vessels so that a single vessel cannot determine the full regional summary. This threshold is an analysis choice, not a completeness guarantee. The first and last 10-minute bins of the retained record are conservatively excluded to avoid incomplete collection boundaries.

There are 31,777 eligible cell-windows in 69 cells. The observed vessel count per eligible window has a median of 9, a mean of 10.30 and a maximum of 34. Figure 1 shows their spatial distribution and chronological coverage. Blank map cells have no eligible windows and are not asserted to be empty water or land.

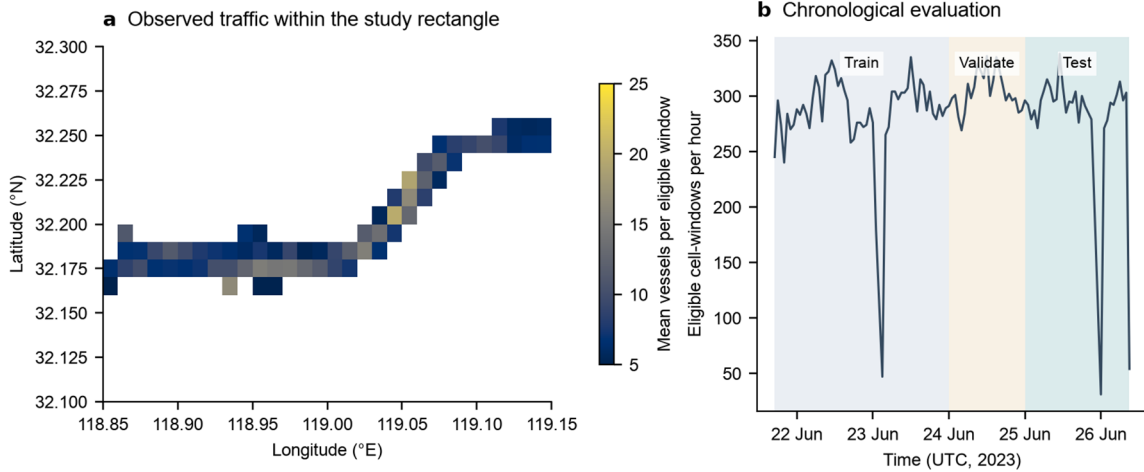


Figure 1. Study area and data coverage. (a) Mean observed vessel count in eligible 10-minute windows for each 0.01° cell. Only cells with at least one window containing five or more vessels are shown; no shoreline or official channel boundary is implied. (b) Number of eligible cell-windows per UTC hour. Shading identifies training, validation and test periods. Coverage is descriptive and does not measure receiver detection probability. Source data contain the grid and hourly aggregates.

For an interpretable display, we define three operational states: mostly moving when $q < 1/3$, mixed speeds when $1/3 \leq q < 2/3$, and mostly low speed when $q \geq 2/3$. A mostly low-speed state can reflect normal anchoring or berth activity. State names therefore refer only to speed composition. Their boundaries are fixed before fitting the prediction models and are not expert congestion annotations.

3. Dynamic inference and evaluation

3.1. Prediction from current and recent windows

At the end of window t , the prediction target is $q_{g,t+h}$ for $h = 1, 2, 3$, corresponding to 10, 20 and 30 minutes between window ends. Each target is a future window aggregate, not

an instantaneous vessel configuration. Figure 2 summarizes the information flow. The current-window feature vector contains q, n, μ, s , the two grid indices, and sine and cosine encodings of UTC hour. The dynamic feature vector additionally contains q, n, μ from windows $t - 1$ and $t - 2$, together with their changes from $t - 1$ to t . This yields eight current-window features and seventeen dynamic features.

Samples require eligible current and two preceding windows, as well as an observed eligible target window. Lagged observations are joined by exact timestamps and cell indices. No interpolation connects missing windows. Both learned models are evaluated on this same sample set. For the longer horizons, an unobserved intervening future window is not used as an input and does not itself invalidate an otherwise observed target. Eligibility of the target is checked retrospectively for evaluation; the model does not predict

whether sufficient vessels will report in that future window.

Gradient-boosted regression trees provide a nonlinear mapping from these features to the future proportion. Gradient boosting is an established additive learning approach [8]. We use its histogram-based implementation and do not modify the underlying algorithm. To allow the learning model to retain a persistence component when validation supports it, the final estimate is

$$\hat{q}_{g,t+h} = \text{clip}_{[0,1]}[\alpha f_h(\mathbf{z}_{g,t}) + (1 - \alpha)q_{g,t}], \quad (2)$$

where f_h is fitted separately for each horizon and α is chosen using validation data. When $\alpha = 0$, the estimate is persistence. Both the current-window and dynamic models use the same parameter search and blend candidates. This comparison evaluates the additional history under matched model-family and selection conditions.

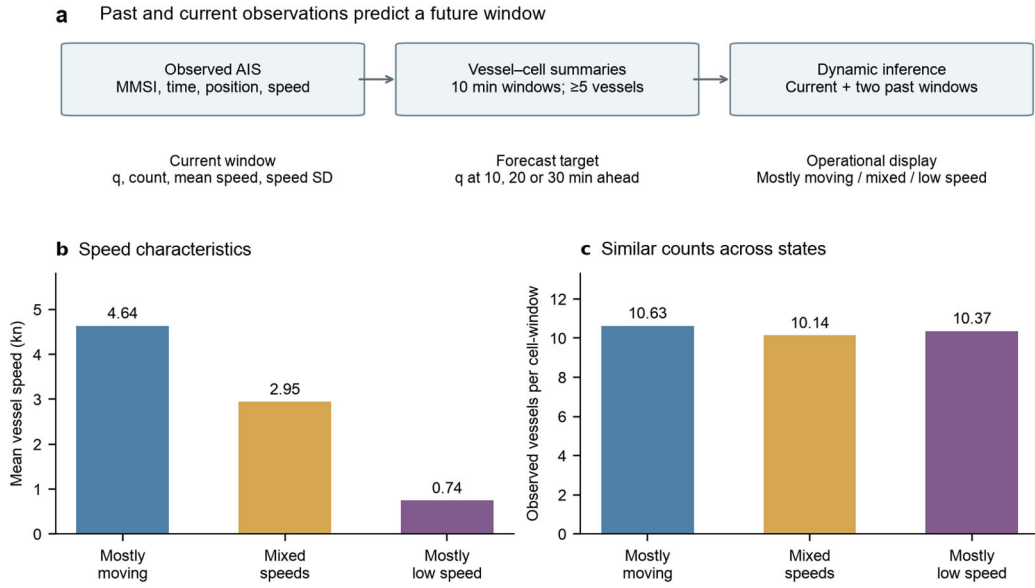


Figure 2. Dynamic inference and interpretation of regional summaries. (a) Observations available through the current window form features for a future low-speed proportion; the predicted proportion can be converted into three operational states. (b,c) Mean speed and mean observed vessel count across the 15,360 eligible training-period cell-windows, grouped by the fixed state definitions. There are 8,342 mostly moving, 4,062 mixed-speed and 2,956 mostly low-speed windows. Bars are descriptive means with no uncertainty intervals. Speed differences partly follow from the state definition and are not an independent validation of state labels.

3.2. Chronological protocol and baselines

The evaluation follows the chronological deployment question rather than randomly mixing windows. Time-series evaluation requires attention to dependence and model assumptions; standard cross-validation can be valid under particular autoregressive conditions [9]. Here, a later holdout provides a direct test of operation after the fitting period. Training targets end no later than 24 June 2023 at 00:00 UTC. Validation targets start on 24 June and end no later than 25 June at 00:00 UTC. Test targets start on 25 June and continue to the end of the eligible record. Observed histories may cross

a split boundary because they would be available at prediction time, while fitting targets remain within their designated period.

Table 1 lists the resulting sample counts. Model parameters are fitted only on training data, selection uses validation mean absolute error (MAE), and the selected models are frozen for testing. No refitting on validation targets is performed. Later test observations may enter later feature windows as they arrive; this is observation-updated forecasting with fixed model parameters. The experiment concerns later times in the same study area and is not a test of transfer to an unseen waterway.

Table 1. Chronological prediction samples by horizon. Each sample is a cell with a complete observed history and an eligible target window.

Horizon	Training	Validation	Test	Test cells
10 min	11,862	5,855	7,094	61
20 min	11,697	5,845	6,990	58
30 min	11,566	5,847	6,937	59

Four methods are compared. Grid mean assigns the cell’s mean training target, with the overall training mean as a fallback for an unseen cell. Persistence assigns the current

proportion to the future window. Current window uses the eight instantaneous summary and context features. Dynamic history uses all seventeen features. Persistence is particularly

important because substantial temporal stability can make a learning model appear useful against an unnecessarily weak comparator.

For both learned models, the search combines maximum leaf counts of 7 or 15, 100 or 200 boosting iterations, squared-error or absolute-error training loss, and blend weights of 0, 0.25, 0.50, 0.75 or 1. The learning rate is 0.05, minimum leaf size is 30, L2 regularization is 1, and automatic early stopping is disabled. The implementation is scikit-learn 1.9.1 with seed 20260917. The selected 10-minute models both use 15 leaves, 200 iterations, absolute-error loss and a blend weight of 1. Thus, their main comparison is not explained by different selected blend weights.

3.3. Metrics and uncertainty

MAE measures the average absolute difference between the predicted and observed proportions. Root mean square error (RMSE) gives greater weight to large deviations. All targets share the same bounded scale, allowing direct comparison across cells. We avoid percentage error because the observed proportion can be zero; the choice of forecast-error measure requires attention to such properties [10]. State accuracy and macro-averaged F1 are obtained by applying the fixed three-state thresholds to both predictions and observed

targets. They measure agreement with operational categories, not accuracy against external traffic-safety labels.

For the primary 10-minute comparison, uncertainty is summarized with a paired block bootstrap. All test cells within the same non-overlapping two-hour UTC block are resampled together, preserving their within-block dependence. We draw 1,000 replicates from 17 available time blocks and report percentile 95% intervals for the difference in MAE. Positive differences indicate lower error for dynamic history. These intervals describe variability within a short test period; they do not account fully for longer-term dependence or establish generalization across seasons.

Additional analyses divide test samples by current observed vessel count and by the subsequently observed magnitude of change. A change of at least 0.20 in q defines the changing subgroup. This is a retrospective diagnostic, not a regime selector available to the model. Binary threshold checks at $q = 1/3$, $1/2$ and $2/3$ assess whether categorical agreement depends on the display threshold.

4. Experimental results

4.1. Added history improves continuous prediction

Table 2. Performance on the 7,094 common test samples at the 10-minute horizon. Lower MAE and RMSE and higher state agreement indicate better performance for the defined task.

Method	MAE	RMSE	State accuracy	Macro F1
Grid mean	0.1011	0.1428	75.32%	0.7475
Persistence	0.0810	0.1204	80.10%	0.7949
Current window	0.0762	0.1117	80.03%	0.7914
Dynamic history	0.0719	0.1048	81.15%	0.8047

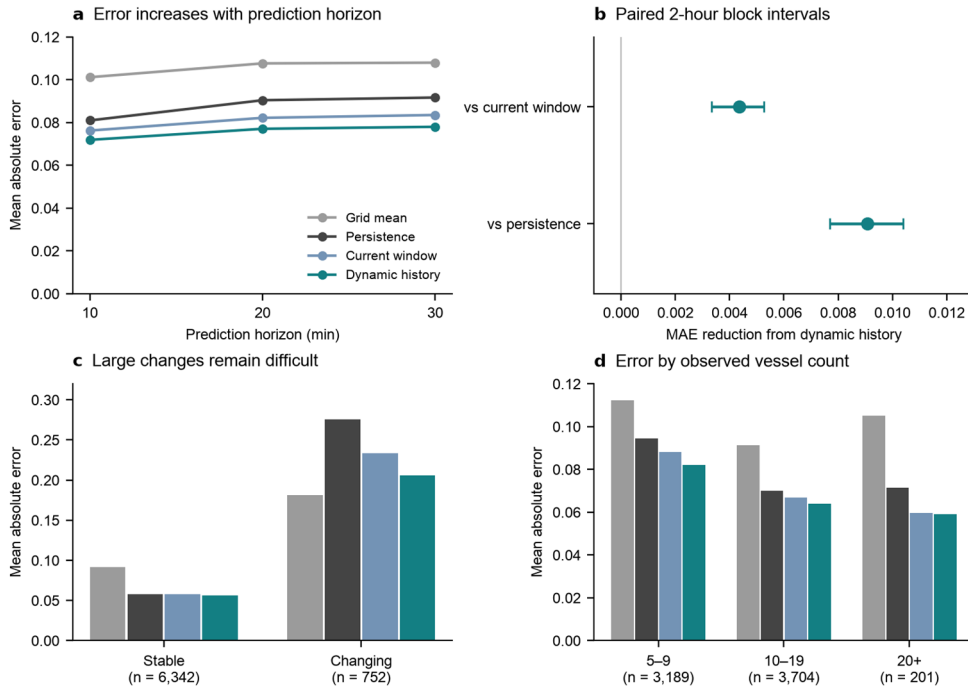


Figure 3. Held-out forecast performance and error diagnostics. (a) MAE at 10, 20 and 30 minutes for all four methods, with 7,094, 6,990 and 6,937 respective test samples. (b) Paired MAE reductions for dynamic history at 10 minutes; markers show the observed difference and horizontal bars show percentile 95% intervals from 1,000 resamples of 17 two-hour time blocks. (c) MAE for samples with observed absolute change below 0.20 and at least 0.20. (d) MAE by current observed vessel count. Colors in (c,d) follow the method colors in (a). The subgroup bars are descriptive point estimates, and subgroup selection in (c) uses future observations only for evaluation.

Dynamic history achieved the lowest overall MAE and RMSE among the four tested methods at the 10-minute horizon (Table 2). Its MAE was 0.0719, compared with

0.0810 for persistence and 0.0762 for the current-window model. The corresponding reductions were 11.2% and 5.7%. In the original proportion scale, the improvement over

persistence was 0.00908, or approximately 0.91 percentage points of the low-speed vessel proportion. This modest absolute difference is more informative than the relative percentage alone.

The paired block interval for the MAE reduction relative to persistence was 0.00769–0.01039. Relative to the current-window model, the reduction was 0.00435 with an interval of 0.00334–0.00526. Both intervals were above zero within this evaluation period (Figure 3b). The current-window model reduced continuous error relative to persistence but had slightly lower state accuracy and macro F1. Consequently, better proportion estimates did not automatically improve every threshold-based measure.

The dynamic model also had the lowest overall MAE at the two longer horizons. At 20 minutes, its MAE was 0.0770, compared with 0.0904 for persistence and 0.0822 for the current-window model. At 30 minutes, the corresponding values were 0.0780, 0.0916 and 0.0835. Figure 3a shows this pattern. Sample availability differed slightly across horizons, so the curve should not be interpreted as a perfectly matched cohort experiment. All within-horizon method comparisons used identical samples.

4.2. Stable periods and abrupt changes behave differently

The 10-minute test set contained 6,342 samples with an

absolute change below 0.20 and 752 samples with a change of at least 0.20. For the stable group, MAEs were 0.0579 for persistence, 0.0576 for current window and 0.0560 for dynamic history. On changing samples, the errors increased to 0.2757, 0.2333 and 0.2056, respectively. Additional history therefore reduced error relative to the two near-term comparators, but it did not resolve abrupt regional changes.

The historical grid mean had a changing-subgroup MAE of 0.1813, below the dynamic model's 0.2056, despite its worse overall performance. This is an important counterexample to a claim of universal superiority. Because this subgroup is defined by a large departure from the current observation, a forecast that shrinks toward a long-term regional mean can perform favorably within it. The result does not demonstrate an advance rule for identifying those windows. Stable-subgroup state accuracy also favored persistence, at 86.75%, over dynamic history, at 85.79%.

Dynamic-history MAE was 0.0820 for 5–9 vessels, 0.0639 for 10–19 vessels and 0.0590 for at least 20 vessels (Figure 3d). These groups contained 3,189, 3,704 and 201 test samples. Larger counts can make the empirical proportion less sensitive to one vessel, but differences in location and traffic composition are also possible explanations. The small highest-count subgroup does not support a general claim about performance under extreme density.

4.3. State agreement and a regional sequence

Table 3. Binary threshold sensitivity on the common 10-minute test set. A positive target has an observed low-speed proportion at or above the indicated threshold.

Threshold	Positive samples	Persistence accuracy	Dynamic accuracy	Persistence F1	Dynamic F1
1/3	4,065	88.85%	88.85%	0.9027	0.8990
1/2	2,867	87.54%	88.74%	0.8458	0.8535
2/3	1,803	91.09%	92.20%	0.8248	0.8342

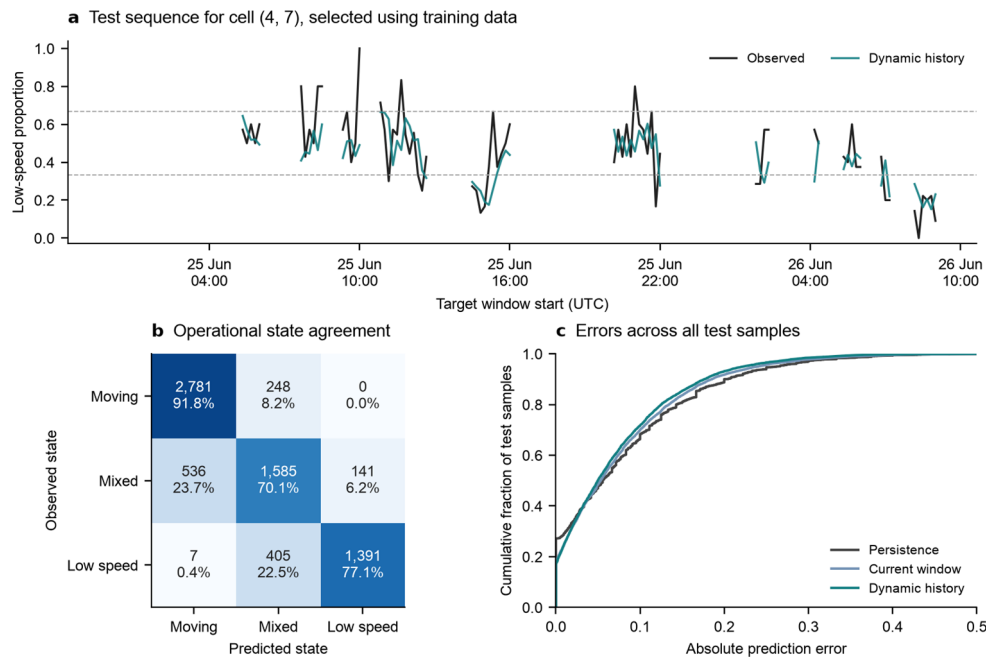


Figure 4. Temporal behavior and state agreement at the 10-minute horizon. (a) Observed and predicted proportions for a cell selected by training variability, plotted against the target-window start in UTC. Dashed horizontal lines mark the operational state boundaries; missing test windows remain gaps. (b) Confusion matrix for all 7,094 test samples. Each cell contains its count and percentage within the observed-state row. (c) Empirical cumulative distributions of absolute error for persistence, current window and dynamic history. The example illustrates temporal behavior; it does not determine the reported aggregate performance.

The operational three-state accuracy of dynamic history was 81.15%, with macro F1 of 0.8047. Figure 4b shows the

confusion matrix, including counts and row-normalized percentages. To illustrate a temporal sequence without

selecting a visually favorable prediction, the example cell in Figure 4a was chosen solely by the largest training-period standard deviation of q among cells with at least 100 eligible training windows. The entire available test sequence for that cell is displayed, and gaps are left unconnected. Aggregate conclusions rely on all test cells rather than this illustration.

Binary agreement depended on the threshold. At a threshold of 0.50, dynamic history achieved 88.74% accuracy, compared with 87.54% for persistence. At 2/3, the respective accuracies were 92.20% and 91.09%. At 1/3, both methods achieved 88.85%, while persistence had the higher positive-class F1, 0.9027 versus 0.8990 (Table 3). These results support reporting the continuous target alongside any categorical display instead of selecting a favorable threshold after evaluation.

5. Discussion

The main finding is that two preceding AIS windows added useful predictive information to current regional summaries in this case area. The matched current-window model already had access to grid location, time of day and the current low-speed proportion. The remaining improvement therefore supports using recent regional history within the tested learning family. It does not identify a physical mechanism or establish that a more complex architecture would perform better. The tree learner, grid representation and temporal features are established techniques; the paper's value is the explicit target, fair baselines and measured application results.

The approach can complement an existing streaming monitor. A Kafka-based ingestion process and a Redis record keyed by vessel identity can maintain the observations needed for each cell-window. After a window closes, three recent regional summaries are sufficient to form the dynamic inputs. The output would contain the current observed proportion, the predicted future proportion and an operational state label. Incomplete histories should be identified explicitly; a persistence fallback could be offered, but its performance on missing-history cases was not tested here. The present experiments are offline and do not establish end-to-end stream throughput, fault tolerance or production latency.

Several boundaries matter for interpretation. First, only approximately 4.7 days and one previously explored region are available. The test covers about 33 hours, and 17 bootstrap blocks provide limited information about variability. Second, complete-history selection retains 7,094 of 9,126 eligible target windows in the test period, approximately 77.7%. Reported errors therefore describe a comparatively observable subset. A system evaluated across reception gaps may perform differently. Third, the earlier intermediate extraction imposes quality and course-availability filters that cannot be independently rechecked here against the cloud-resident source.

Fourth, the low-speed proportion is affected by anchoring, berth operations, maneuvering and vessel composition. Without berth polygons, navigational status validation, channel capacity or operator annotations, it cannot establish congestion, vessel delay or collision risk. Changes in the proportion may arise from vessels entering and leaving a cell as well as from individual speed changes. Fifth, the 0.01° grid, minimum count of five and speed threshold of 2 kn are fixed design choices. Sensitivity to grid resolution, reporting loss and alternative speed thresholds has not been tested. The categorical threshold checks address display boundaries only. Independent waterways, longer observation periods and

operational annotations would be needed to assess broader monitoring value.

6. Conclusion

This study formulated dynamic regional traffic-state inference as prediction of the future low-speed vessel proportion from AIS window sequences. In a lower Yangtze case area, adding two preceding windows reduced 10-minute MAE by 11.2% relative to persistence and by 5.7% relative to a matched current-window model. Improvements in overall continuous prediction remained at 20 and 30 minutes. Simple baselines remained competitive in stable periods and selected changing cases, which limits claims of general superiority. The results support a modest addition to existing vessel-window monitoring: retaining recent regional history to improve short-term estimates while keeping the output tied to an observable, explicitly defined traffic quantity.

Acknowledgment

This work was supported by the Research Project of Jiangsu Maritime Institute (Grant No. 2025ZKyb02).

References

- [1] Tu, E., Zhang, G., Rachmawati, L., Rajabally, E., & Huang, G. (2018). Exploiting AIS data for intelligent maritime navigation: A comprehensive survey from data to methodology. *IEEE Transactions on Intelligent Transportation Systems*, 19(5), 1559-1582. <https://doi.org/10.1109/TITS.2017.2724551>
- [2] Emmens, T., Amrit, C., Abdi, A., & Ghosh, M. (2021). The promises and perils of Automatic Identification System data. *Expert Systems with Applications*, 178, Article 114975. <https://doi.org/10.1016/j.eswa.2021.114975>
- [3] Zhao, L., Shi, G., & Yang, J. (2018). Ship trajectories pre-processing based on AIS data. *Journal of Navigation*, 71(5), 1210-1230. <https://doi.org/10.1017/S0373463318000188>
- [4] Liu, Z., Wu, Z., Zheng, Z., & Yu, X. (2022). A molecular dynamics approach to identify the marine traffic complexity in a waterway. *Journal of Marine Science and Engineering*, 10(11), Article 1678. <https://doi.org/10.3390/jmse10111678>
- [5] Yu, H., Meng, Q., Fang, Z., Liu, J., & Xu, L. (2022). A review of ship collision risk assessment, hotspot detection and path planning for maritime traffic control in restricted waters. *Journal of Navigation*, 75(6), 1337-1363. <https://doi.org/10.1017/S0373463322000650>
- [6] Li, H., Jiao, H., & Yang, Z. (2023). AIS data-driven ship trajectory prediction modelling and analysis based on machine learning and deep learning methods. *Transportation Research Part E: Logistics and Transportation Review*, 175, Article 103152. <https://doi.org/10.1016/j.tre.2023.103152>
- [7] March, D., Metcalfe, K., Tintoré, J., & Godley, B. J. (2021). Tracking the global reduction of marine traffic during the COVID-19 pandemic. *Nature Communications*, 12(1), Article 2415. <https://doi.org/10.1038/s41467-021-22423-6>
- [8] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189-1232. <https://doi.org/10.1214/aos/1013203451>
- [9] Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120, 70-83. <https://doi.org/10.1016/j.csda.2017.11.003>
- [10] Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of*

Forecasting, 22(4), 679-688.
<https://doi.org/10.1016/j.ijforecast.2006.03.001>