ATLANTES: A SYSTEM OF GPS TRANSFORMERS FOR GLOBAL-SCALE REAL-TIME MARITIME INTELLIGENCE

Henry Herzog, Joshua Hansen, Yawen Zhang, & Patrick Beukema
Allen Institute for AI
Seattle, WA, USA
{henryh,joshuah,yawenz,patrickb}@allenai.org

ABSTRACT

Billions of humans depend on healthy oceans for prosperity and sustenance. Unsustainable exploitation of the oceans exacerbated by climate change are threatening coastal communities worldwide. Accurate and timely monitoring of maritime activity is an essential step to effective governance and to inform future policy. In support of this complex global-scale effort, we built *Atlantes* [1], a machine learning based system that provides the first ever real-time view of vessel behavior at global scale. Atlantes leverages a series of bespoke transformers to distill a high volume continuous stream of GPS messages (120M/day) emitted by hundreds of thousands of vessels into easily observable behaviors. The combination of low latency and high performance enables operationally relevant decision-making and successful interventions on the high seas where illegal and exploitative activity is too common. Atlantes is already in use by hundreds of organizations worldwide. Here we provide an overview of the machine learning model and infrastructure that enables this system to function efficiently and cost-effectively at global-scale and in real-time.

1 INTRODUCTION

Every day, over 600,000 ships transmit a steady stream of their geolocations using an automatic shipboard broadcasting system known as the automatic identification system (AIS) (Maritime, 2025). At its core, an AIS message is simply a unique identifier, a timestamp, and a location (latitude and longitude), typically alongside metadata such as the current speed and heading. Every day, over 100 million AIS messages are broadcasted and recorded across a constellation of satellites, compiled into streams, and delivered to a diverse user base including coastal state governments, environmental organizations, and commodities traders (fig: 1A). While AIS was originally designed to facilitate collision avoidance, today it supports a far broader set of use cases including supply chain tracking, route optimization, insurance and compliance, commodity tracking, environmental monitoring, and beyond. Alongside satellite imagery, AIS is one of the core datasets for global-scale maritime intelligence. However, unlike satellite imagery, which provides just a fleeting glimpse of a sliver of global traffic at comparatively slow latency, AIS offers a near comprehensive streaming source of real-time information.

We stress that this work was significantly different from a conventional machine learning research effort in that it was entirely application-driven (Rolnick et al., 2024). Our primary goal was to develop a GPS model capable of meeting user's needs (minimal latency and expert performance). Additionally, the model had to be highly computationally efficient to be cost-effective to run in production and to enable rapid iteration. To achieve this, we designed several custom transformer models and trained them on a large dataset of GPS trajectories annotated at message-level granularity by a group of leading maritime analysts. The entire code base is open-sourced (GitHub), allowing users to deploy this model or fine-tune for other uses.

[1] Our base model architecture is called ATLAS (AIS transformers learning for active subpaths). *Atlantes* refers to a collection of atlases.

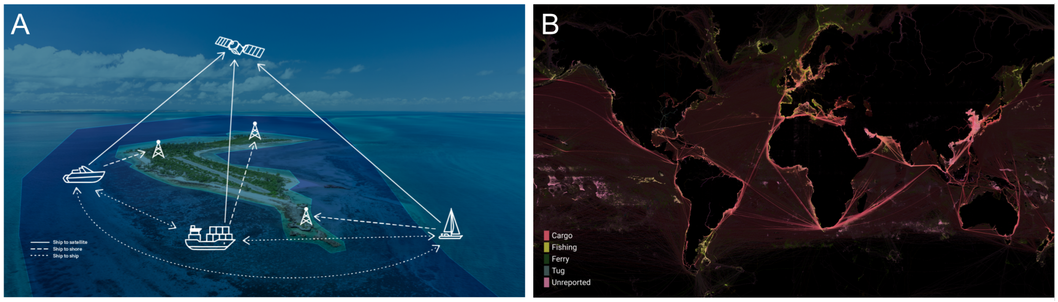


Figure 1: (A) Depiction of the Automatic Identification System. (B) All broadcasted messages from 2023 (20B) color coded by vessel category.

1.1 Previous Work

AIS data have been modeled with many different machine-learning based approaches for a variety of tasks including vessel type classification (e.g. cargo vs. tanker), activity classification (e.g. transiting vs. navigating), anomaly detection, and route optimization (see Yang et al. (2024) and Wolsing et al. (2022) for a review). However, prior research has focused on specific regions and/or short time-frames, emphasizing historical analysis over real-time classification. Data availability and limited GPS specific ML tooling (including for annotation), have hampered the development of scalable robust models.

2 Methods

2.1 Tasks and Datasets

Atlantes tackles two classification tasks with AIS sequences: (1) entity classification to determine whether a sequence is from a vessel or a buoy, and (2) activity classification to determine the vessel's activity given the most recently broadcasted AIS message. Entity classification is a requirement because without context, it is not possible to distinguish activity patterns on the basis of a single message. For example, a vessel may transport a buoy resulting in vessel-like trajectories for a period of time, or a vessel may drift without power resembling a buoy-like trajectory.

Entity classification dataset: We applied domain-expert heuristics and leveraged AIS metadata to sample buoys with high precision but low recall across the entire population of AIS from 2024 ($\sim$20B messages). The result was a dataset of 1.8M entities (175,857 buoy labels and 1,643,737 vessel labels). All buoy data was included, while vessel data was stratified by vessel type.

Activity classification dataset: We could not rely on metadata or heuristics to machine annotate vessel trajectories by activity class because no existing metadata reliably separates these behaviors (even if one sacrifices significant recall).

To build an effective global model we compiled a training dataset that covered the entire planet and all vessel types, continuously between January 2022 and June 2024. We hired 20 expert maritime analysts capable of distinguishing other behaviors from fishing. The analysts were trained to ensure consistent labeling against domain definition, and label every GPS message into five classes: transiting, anchored, fishing, moored, or other. This work was conducted using a custom-built platform designed specifically for trajectory annotation (Fig. 5). During the initial sampling phase, 85% of monthly stratified samples came from expert-identified regions, with the remaining 15% from other areas. These data were used to train an initial model, which was later improved via active learning. Experts iteratively labeled new training samples informed by the model predictions. In total, experts labeled over 7,500 track-months ($\sim 10k$ messages in a track-month) of AIS data, encompassing more than 15 million messages. These labels include 558,347 instances of fishing, 502,490 of transiting, 123,696 of moored, and 105,299 of anchored activities.

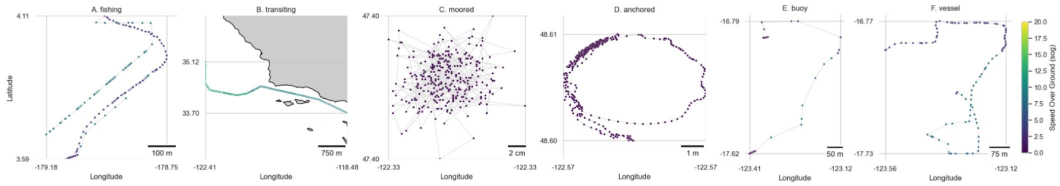


Figure 2: Examples of activity from each class (A-D). Example buoy (E) and vessel (F).

2.2 Model architecture and training

We found that modern transformers are well suited to GPS modeling, but several critical differences to natural language motivated a bespoke approach. Unlike written language, AIS/GPS sequences exhibit highly irregular spatial and temporal frequencies, and can contain high amounts of noise. These features can be handled by GPS specific embeddings and a well designed cost function that mimics how humans interpret the irregularity.

The model is transformer-based with three key components: continuous point embedding (CPE) layers, 1D CNN layers, and transformer encoder inspired by Liang et al. (2022). The CPE layers embed the raw GPS sequences, which are irregular in space and time, into a representation suitable for the encoder layers. This is achieved by computing the spatiotemporal differences between successive anchor points within a configurable window (n=9 messages). This step eliminates the need for message interpolation or extensive feature engineering. Subsequent CNN layers extract local patterns, and the transformer encoder learns global representations across the entire sequence via multi-head self-attention. The ATLAS architecture includes 6 CPE layers, 3 CNN layers, and (Activity=9/Entity=4) transformer layers. In addition to the model, activity classifications are passed through several post-processing layers, including confidence thresholding, geo-fencing around marine infrastructure ((Bastani et al., 2023)), and speed filters. Two different models were trained using this architecture, one for entity classification and a second for activity classification.

Models were trained using class-weighted cross-entropy loss, with weights selected based on the minimum validation loss over 4 epochs, using a batch size of 256, with a learning rate of 1e-4. A single training run takes just 6 hours on four H100 GPUs. The lightweight design of the model, consisting of 4.7M parameters, makes it computationally efficient and suitable for deployment on modest hardware.

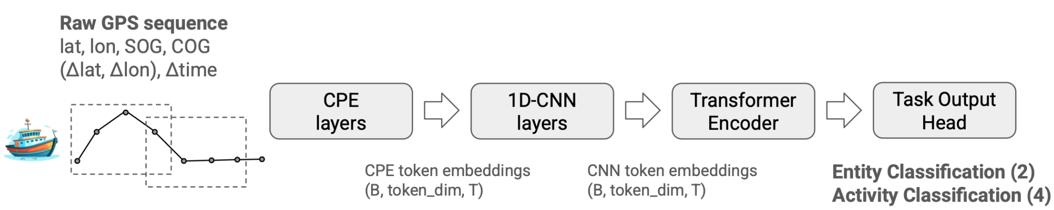


Figure 3: *ATLAS* model architecture.

2.3 How good is this model really?

While we conducted extensive offline evaluations, our primary evaluation method involved experts reviewing the model's outputs in an online staging environment (prior to exposing results to all users). After addressing different sources of false positives identified through this feedback process, we conducted third-party evaluation on globally stratified random samples. For entity classification, our model achieved an accuracy of 97.5%. For activity classification, our model achieved an overall accuracy of 71% (90% for fishing vessels and 51% for unknown vessels). Because there is no other model that offers real-time global predictions, we are not able to produce direct comparison of this model to prior work. We share these metrics because they best reflect existing user experience.

The best context we can provide about how "good" our performance is is that we are approaching human performance but not quite there. We measured inter-annotator agreement on this task at 85%. It is not possible to know with certainty what a vessel is doing in real-time from a single GPS message. There may be multiple plausible activities that will only converge after a more prolonged period of engaging in that activity. Humans also typically use many different data sources, in addition to the GPS sequence, to classify a trajectory (such as google images of the vessel that the model does not have access to). Naturally, performance improves with additional context. At the extreme, classification of GPS messages conditioned on future data, such as analysis of historical data, is a different problem, and should approach 100% accuracy.

2.4 Deployment

Upon receiving new messages, a speed-based change-point detection algorithm identifies any putative change in behavior, triggering a three-stage pipeline involving pre-processing, the model inference, and post-processing. Entity classification is performed offline using historical AIS data stored in a database and is only calculated if metadata for the activity classification tasks lacks that class. This information is typically missing for new trajectories or vessels not previously on the water broadcasting. If insufficient context exists (n=500 messages), the entity classifier defaults to unknown. For activity classification, entity information determines whether specific activities are possible (e.g., fishing events are only associated with vessels).

The system processes 28 activity classifications per second using 20 CPUs and 5 T4 GPUs, with a configurable message context of up to 2048 messages spanning 30 days (approximately 5B GPS messages are inferenced including the context each day).

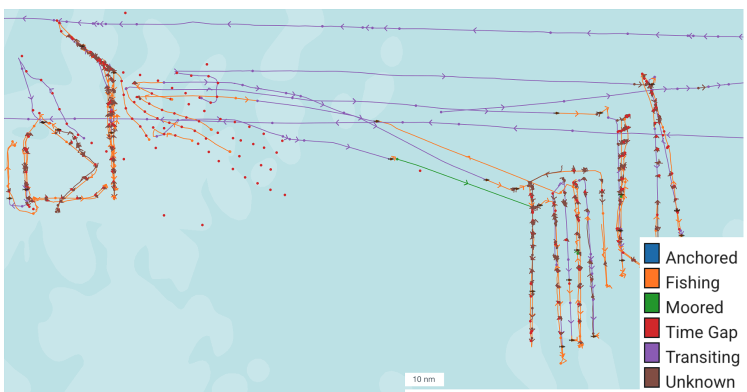


Figure 4: Example classifications of vessel movement patterns. Indian ocean, February 2025.

3 Conclusion/Impact Statement

Our primary contribution is to enable real-time analysis and therefore actionable intelligence. Real-time insights require classifying behavior with more limited information, and care must be taken to ensure that the intelligence we provide is both accurate and worth taking action on. This model is merely one component in a much more complex and global effort to monitor and protect the planet's oceans. We hope that by open sourcing the models, we can enable other researchers and activists to better understand the strengths and limitations of this approach, and also foster more widespread adoption of maritime transparency.

ACKNOWLEDGMENTS

We wish to thank the annotation team for curating and refining the dataset that made all of these models possible. We also thank the engineering team for their work in developing the annotation tool which enabled scalable annotations. Additionally, we thank Spire Maritime for data. Thanks to Karen Goodfellow for help with UI and UX design.

REFERENCES

Favyen Bastani, Piper Wolters, Ritwik Gupta, Joe Ferdinando, and Aniruddha Kembhavi. Satlaspretrain: A large-scale dataset for remote sensing image understanding, 2023. URL `https://arxiv.org/abs/2211.15660`.

Yuxuan Liang, Kun Ouyang, Yiwei Wang, Xu Liu, Hongyang Chen, Junbo Zhang, Yu Zheng, and Roger Zimmermann. Trajformer: Efficient trajectory classification with transformers. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, CIKM '22, pp. 1229–1237, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392365. doi: 10.1145/3511808.3557481. URL `https://doi.org/10.1145/3511808.3557481`.

Spire Maritime. Real-time ais vessel tracking data. Technical report, Spire Maritime, 2025.

David Rolnick, Alan Aspuru-Guzik, Sara Beery, Bistra Dilkina, Priya L. Donti, Marzyeh Ghassemi, Hannah Kerner, Claire Monteleoni, Esther Rolf, Milind Tambe, and Adam White. Application-driven innovation in machine learning, 2024. URL `https://arxiv.org/abs/2403.17381`.

Konrad Wolsing, Linus Roepert, Jan Bauer, and Klaus Wehrle. Anomaly detection in maritime ais tracks: A review of recent approaches. *Journal of Marine Science and Engineering*, 10(1), 2022. ISSN 2077-1312. doi: 10.3390/jmse10010112. URL `https://www.mdpi.com/2077-1312/10/1/112`.

Ying Yang, Yang Liu, Guorong Li, Zekun Zhang, and Yanbin Liu. Harnessing the power of machine learning for ais data-driven maritime research: A comprehensive review. *Transportation Research Part E: Logistics and Transportation Review*, 183:103426, 2024. ISSN 1366-5545. doi: https://doi.org/10.1016/j.tre.2024.103426. URL `https://www.sciencedirect.com/science/article/pii/S1366554524000164`.

A APPENDIX

A.1 CHANGE POINT DETECTOR LOGIC SUMMARY

The Change Point Detector identifies significant shifts in observational data based on time and speed over ground (SOG) metrics. It operates under the assumption that the last element in the input list represents the most recent observation. The pseudocode below takes a list of speed values and returns True or False based on whether a change was detected.

By combining time-based and SOG-based analyses, the detector determines if a changepoint has occurred and clearly communicates the reasoning behind each detection outcome.

A.2 ANNOTATION TOOL

While developing Atlantes, we developed a custom machine learning annotation tool designed to streamline and enhance the classification of vessel behavior using Automatic Identification System (AIS) data. Traditional data annotation methods relied on multiple annotators using ad-hoc tools, leading to inefficiencies in updating and refining machine learning models. Our in-house tool addresses these challenges by providing an intuitive, efficient platform for subject matter experts to visualize, evaluate, and label AIS data.

This platform is optimized for processing vast data streams—approximately 150 million AIS messages daily—allowing for rapid and accurate annotation. Its design prioritizes speed and simplicity, enabling annotators to efficiently generate high-quality datasets essential for developing robust machine learning models. Within a few months of deployment, the tool facilitated the annotation of over 2 million AIS positions with only a small team of annotators.

Algorithm 1 Example Pseudocode for SOG-based Changepoint Detection

```
Require: sogs: A list of vessel speeds over ground (floats)
Require: baseSampleSize, changepointProbabilityThreshold, MIN_MESSAGES,
    MAX_NUM_MESSAGES, MAX_DURATION
procedure isSogBasedChangepoint(sogs)
    currentNumObservations ← length(sogs)
    if currentNumObservations < baseSampleSize then
        return False
    end if
    currentObservations ← convertToNumpyArray(sogs)
    baseSample ← getBaseSample(currentObservations)
                                                ▷ Initialize the base speed distribution
    (currentMean, currentStd) ← initializeBaseSpeedDistribution(baseSample)
                                ▷ Compute a standardized sum (details depend on your method)
    standardizedSum ← calculateStandardizedSum(currentObservations)
    prob ← calculateProbability(standardizedSum)
    if prob < changepointProbabilityThreshold then
        return True
    else
        return False
    end if
end procedure
                        ▷ Main logic that calls the above procedure and applies additional checks
numMessages ← length(sogs)
timeBasedCPD ← TimeGapDetector()
isLessThanMinMessages ← (numMessages < MIN_MESSAGES)
if timeBasedCPD.detectsChangepoint() then
    return True
end if
if isLessThanMinMessages then
    return False
end if
if numMessages > MAX_NUM_MESSAGES then
    return True
end if
if (lastTime − firstTime) > MAX_DURATION then
    return True
end if
if isSogBasedChangepoint(sogs) then
    return True
end if
                        ▷ Final check repeated, for example, if needed in a different context
currentNumObservations ← length(sogs)
if currentNumObservations < baseSampleSize then
    return False
end if
```

Anchored	2.26/sec
Fishing	0.47/sec
Moored	12.4/sec
Transiting	15.5/sec
Unknown	8.5/sec
Total	39.1/sec

Table 1: Inference rates by activity type.

The tool not only accelerates dataset creation but also integrates expert domain knowledge directly into the annotation process, improving the quality and adaptability of AI models. This innovation represents a significant advancement in maritime data annotation, supporting the development of AI systems capable of detecting complex vessel behaviors, such as fishing activities and covert transshipments. Ultimately, the tool contributes to enhancing global maritime monitoring and ocean conservation efforts.

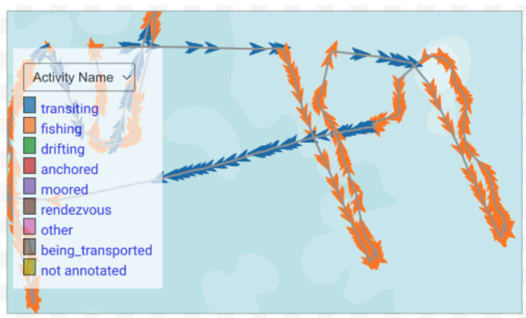


Figure 5: A. Depiction of annotation process. GPS sequences are annotated at message level granularity by assigning a class to each individual message.

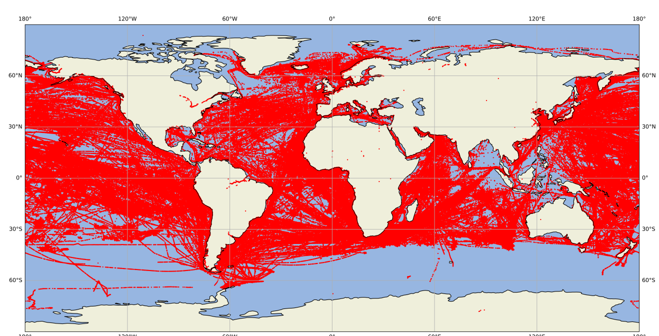
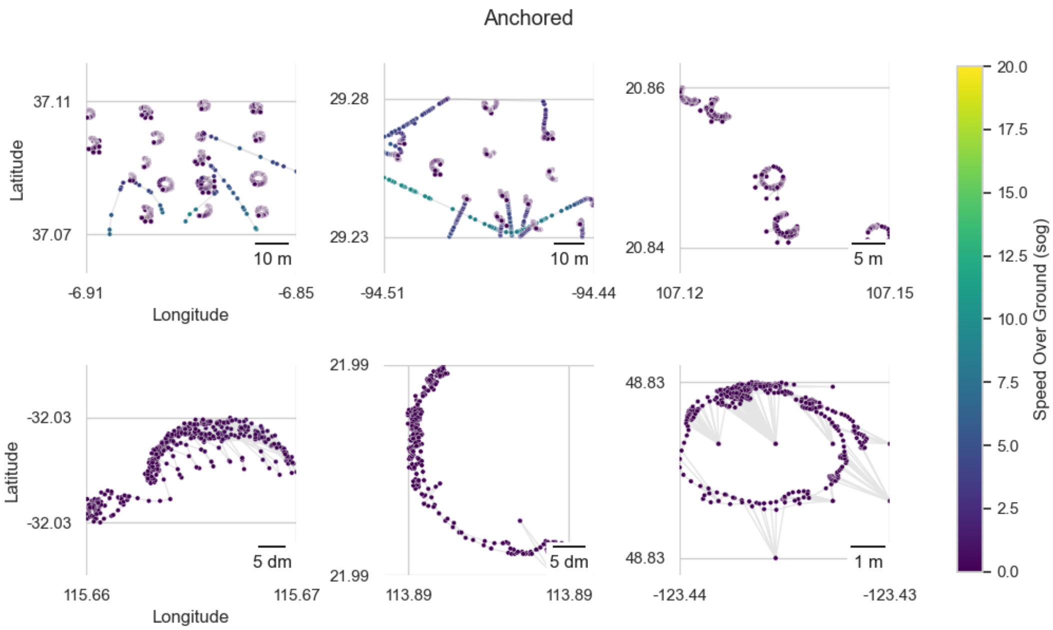
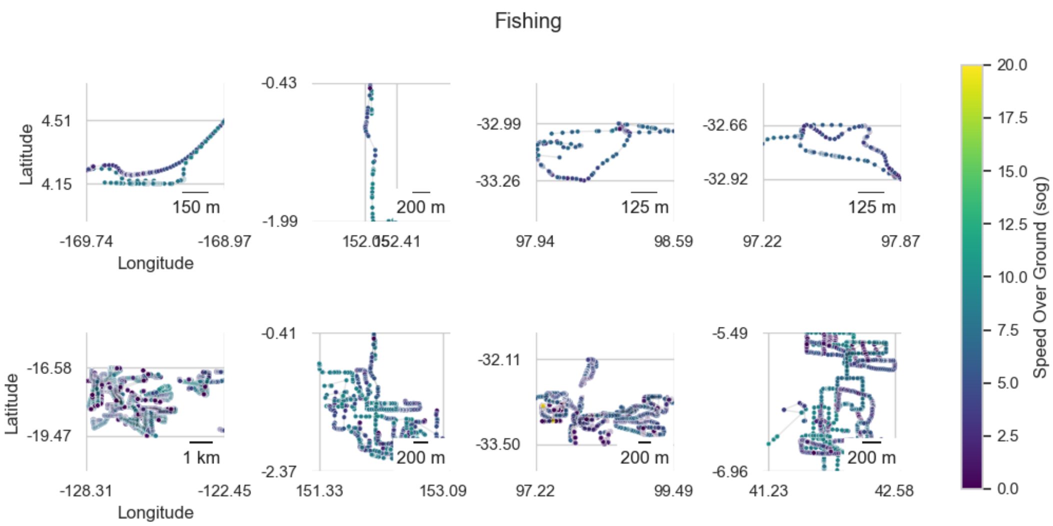
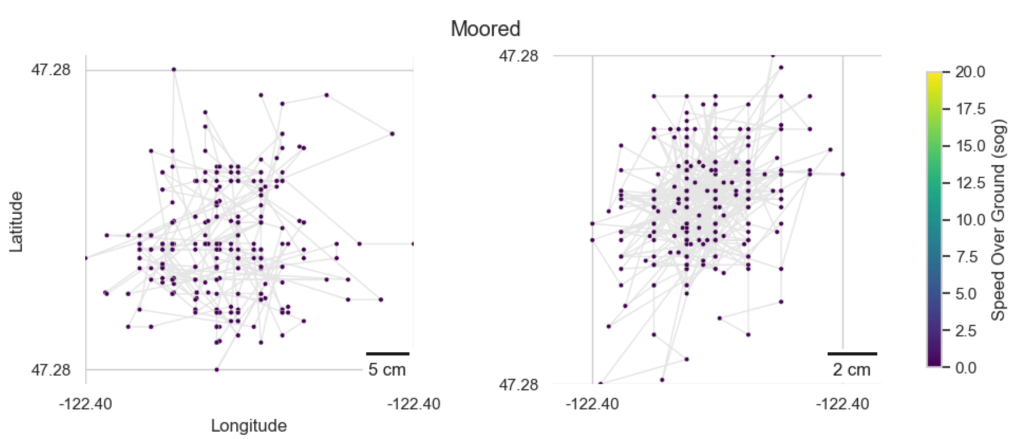


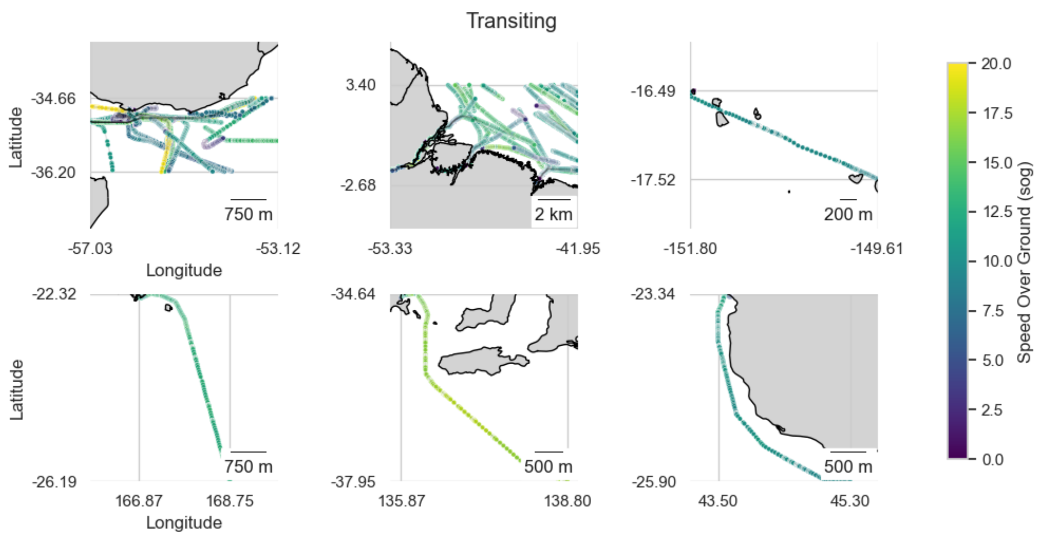
Figure 6: Geographic Distribution of Labeled Messages Used for Activity Classification

A.3 Additional Examples of Activity Predictions


Anchored
Latitude
Longitude
37.11
37.07
-6.91
-6.85
10 m
29.28
29.23
-94.51
-94.44
10 m
20.86
20.84
107.12
107.15
5 m
-32.03
-32.03
115.66
115.67
5 dm
21.99
21.99
113.89
113.89
5 dm
48.83
48.83
-123.44
-123.43
1 m
Speed Over Ground (sog)
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0


Fishing
Latitude
Longitude
4.51
4.15
-169.74
-168.97
150 m
-0.43
-1.99
152.05
152.41
200 m
-32.99
-33.26
97.94
98.59
125 m
-32.66
-32.92
97.22
97.87
125 m
-16.58
-19.47
-128.31
-122.45
1 km
-0.41
-2.37
151.33
153.09
200 m
-32.11
-33.50
97.22
99.49
200 m
-5.49
-6.96
41.23
42.58
200 m
Speed Over Ground (sog)
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0


Moored
Latitude
Longitude
47.28
47.28
-122.40
-122.40
5 cm
47.28
47.28
-122.40
-122.40
2 cm
Speed Over Ground (sog)
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0



Transiting
Latitude
Longitude
-34.66
-36.20
-57.03
-53.12
750 m
3.40
-2.68
-53.33
-41.95
2 km
-16.49
-17.52
-151.80
-149.61
200 m
-22.32
-26.19
166.87
168.75
-34.64
-37.95
135.87
138.80
500 m
-23.34
-25.90
43.50
45.30
Speed Over Ground (sog)
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0