FLP-XR: Future Location Prediction on Extreme Scale Maritime Data in Real-time

George Theodoropoulos

University of Piraeus

co-authors: Andreas Patakis, Andreas Tritsarolis, Yannis Theodoridis

The Challenge of Maritime Mobility Prediction

Core Problem: Predicting where ships will be in the future.

Why It Matters:

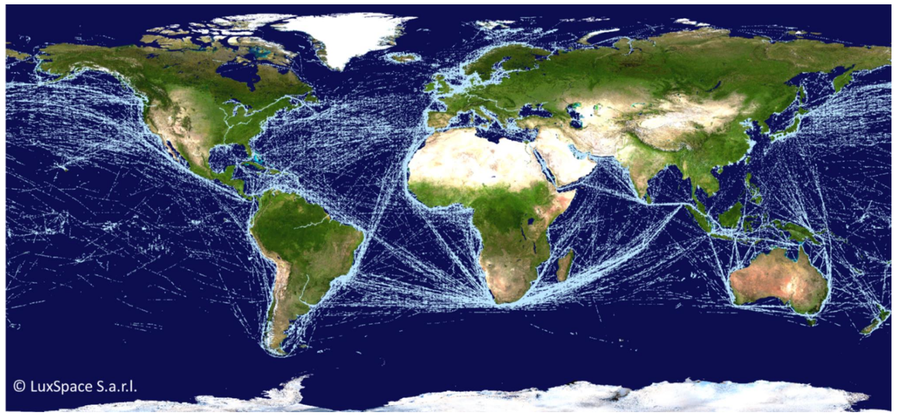
- Collision Avoidance (Safety)
- Route Optimization & Efficiency (Operations)
- Port Management & Logistics (Planning)
- Environmental Monitoring
- Search and Rescue Operations

The Scale is Immense:

- AIS data streams continuously from thousands of vessels globally.
- This creates massive datasets requiring real-time processing capabilities.

Existing Challenges: High volume, high velocity, limited resources (especially at Edge/Fog).

Need: Fast prediction methods suitable for low-power edge devices.



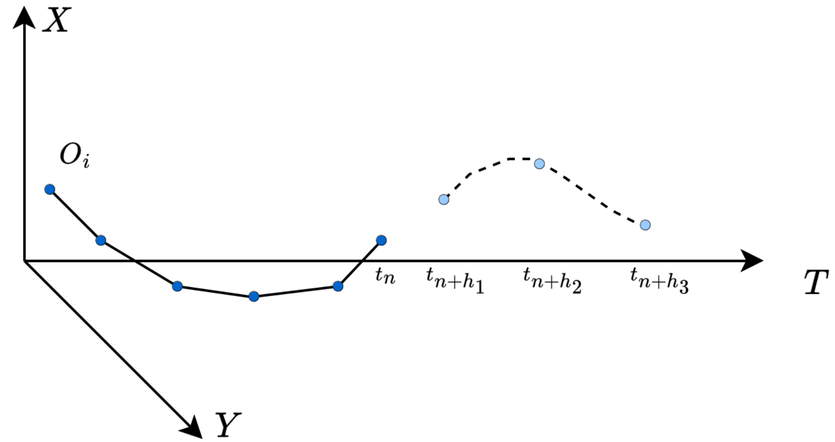
Satellite-AIS-based map of global ship traffic (esa.int)

FLP Problem Formulation

- Given (i) a stream of AIS records <**VID, lon(t), lat(t), t**> generated by a set of vessels, where **VID** is the vessel's unique identifier (e.g., MMSI) and <**lon(t), lat(t)**> is the pair of vessel coordinates recorded at timestamp **t**, and (ii) a prediction time horizon **Δt**,
- … the problem at hand is to train a Machine Learning (ML) model able to forecast a vessel's anticipated location <**lon(t + Δt), lat(t + Δt)**>

$$V_i(t + \Delta t) = [lon(t + \Delta t), lat(t + \Delta t)]$$

Horizon Values (**Δt**) play a big role in the quality of the final result. The values used in our study are 10, 20, …, 60 min., in an effort to address short-to-medium term predictions.



Predicting the future location of vessel for three distinct horizon values.

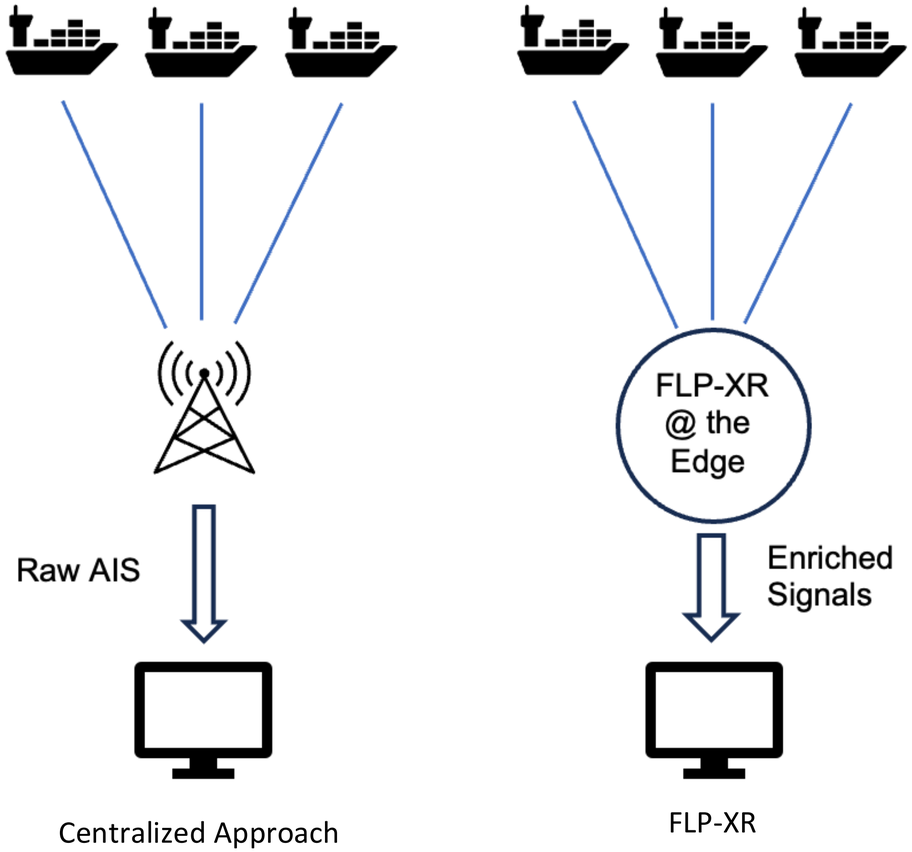
Introducing FLP-XR: A Fast and Efficient Solution

Future Location Prediction - Extreme Gradient Boosting approach (XGBoost)

- **FLP-XR =** Future Location Prediction + eXtreme scale & XR for Edge/Fog.
- **Core Objective:** Provide accurate predictions quickly, enabling deployment on resource-limited hardware.

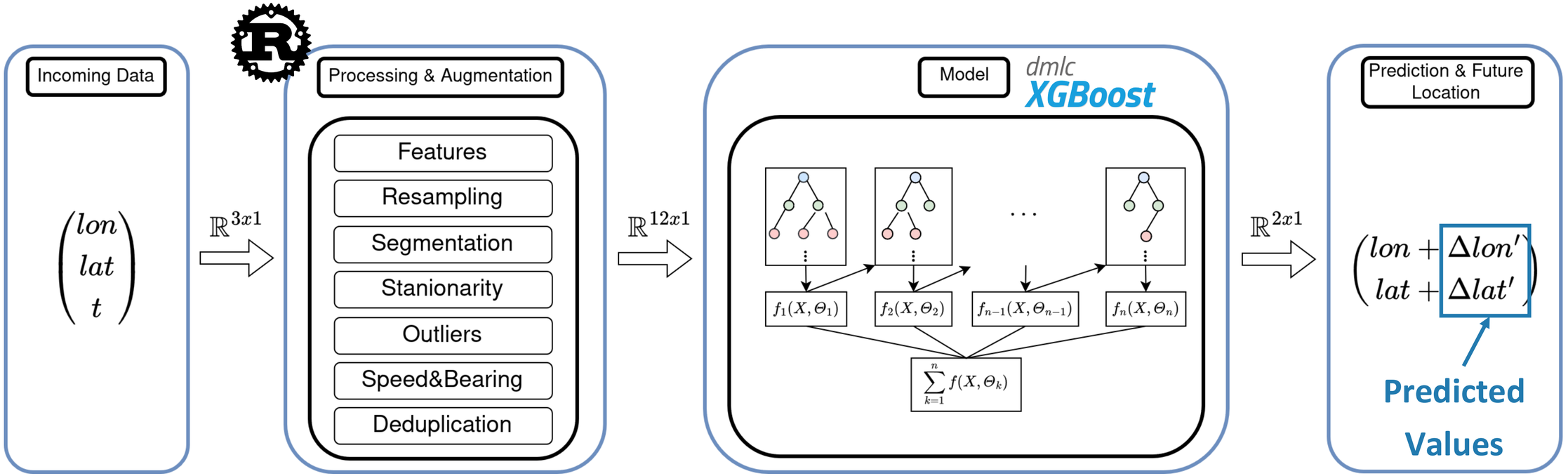
Key Claims:

- Achieves high accuracy in many scenarios (comparable to SotA).
- Trained significantly faster than traditional RNNs – up to 300x speedup wrt SotA!
- Inference extremely fast – up to 3000x faster than SotA!



Difference between centralized and in-situ data analytics

The FLP-XR architecture



- **Input:** Continuous stream of AIS records <Vessel ID, lon(t), lat(t), Timestamp>.
- **Output:** Predicted location (lon, lat) at specific future time horizons (e.g., 10 min, 20 min) based on the current position.
- **How Horizons are Handled:** Preprocess historical data to create training examples for each fixed horizon separately. This simplifies the model and makes it faster.

Data Preparation & Key Features

A crucial step: Preprocessing & Feature Engineering (in **Rust**):

- **Extrapolated Differences:** Calculates the difference between the vessel's actual position and where it would be if continuing its last linear trend (Δt). Deviations signal potential turning points or speed changes.
- **Historical Averages:** Computes moving averages of speed over recent intervals (e.g., average speed over 30 minutes) to capture typical behavior trends.
- **Speed & Heading Differences:** Derives features like velocity differences from predicted path, heading change rate (heading_diff), and standard deviation of historical headings. These quantify the vessel's dynamic state and maneuvering tendencies.
- **Vessel Context Features (v_type, origin_port):** Incorporates specific vessel metadata (e.g., container ship vs. tugboat) and context information like its last known port or destination area. This provides important domain knowledge influencing future behavior

FEATURE NAME	DESCRIPTION
v_type	vessel type
coords (lon, lat)	vessel location coordinates
sp & br	vessel speed & bearing, calculated at a given timestamp
extrap_diff (lon, lat)	difference between the latest and future coordinates, assuming linear extrapolation
last_diff (lon, lat)	difference between the latest and previous coordinates
origin & orig_dist	POI closest to starting point of vessel's trip & its distance from the latest location
Δt	prediction time horizon

Experimental Setup & Datasets

We evaluated FLP-XR using three real-world AIS datasets:

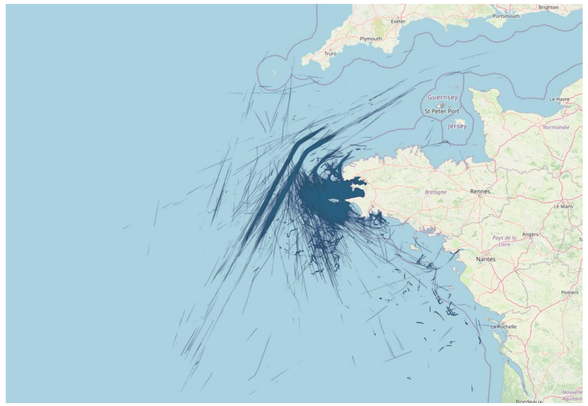
- **Brest Dataset**, open dataset, ~6 months data
- **Piraeus Dataset**, open dataset, ~ 1 year data
- **Aegean Dataset**, provided by MarineTraffic/Kpler (MobiSpaces project)

Geographic Scope: datasets spanning different ports and seas.

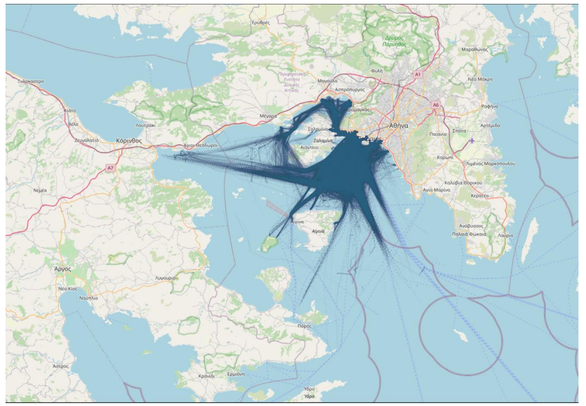
Hardware tested, for high- to low-resource devices:

- High-end GPU Server
- NUC-i7 (Standard x86 CPU)
- Raspberry Pi 4 (ARM-based, low-power device)

Comparison Platform: Nautilus SotA model* was used for baseline performance.



Brest Dataset



Piraeus Dataset

Aegean Dataset

* Tritsarolis, A., et al. (2024). On Vessel Location Forecasting and the Effect of Federated Learning. In Proceedings of the 25th IEEE MDM Conference.

Performance Comparison: Accuracy vs. SotA

FLP-XR generally provides competitive accuracy, sometimes outperforming Nautilus.

- **Example (Brest)**: At 10 min horizon, error is lower for FLP-XR than Nautilus.
- **Stability Note**: FLP-XR errors increase linearly with prediction horizon across different vessel types and locations. This predictability can be advantageous.

Key Observation: While often competitive, there are specific cases (e.g., Piraeus vessels at longer horizons) where the SotA model might have a slight edge in error. However, FLP-XR's consistent performance is very usable for Edge scenarios.

Prediction error ($\mu \pm \sigma$) in meters of FLP-XR and Nautilus per dataset.

Dataset	Model	10min.	20min.	30min.	40min.	50min.	60min.
Brest	FLP	**429 ± 544**	1066 ± 1178	**1749 ± 1796**	2364 ± 2470	3119 ± 3215	3935 ± 3986
	SotA	509 ± 806	**988 ± 1379**	2112 ± 2489	**2197 ± 2349**	**2798 ± 2705**	**3912 ± 3770**
Piraeus	FLP	339 ± 369	838 ± 811	1374 ± 1281	1874 ± 1759	2598 ± 2378	**3334 ± 2852**
	SotA.	**234 ± 501**	**492 ± 1078**	**442 ± 1104**	**831 ± 2020**	**1475 ± 3283**	3818 ± 5668
Aegean	FLP	**176 ± 615**	**460 ± 1040**	**809 ± 1508**	**1123 ± 1914**	**1543 ± 2488**	2003 ± 3027
	SotA.	264 ± 392	615 ± 607	1205 ± 1720	1563 ± 1795	3246 ± 4022	**1978 ± 1566**

Efficiency Under Constraints: Speed on Edge Hardware

Training (Crucial):

- On NUC-i7: Average training time reduced from 3.6 hours to just under 2 minutes! (300x speedup).
- This allows models to be trained much more frequently or adapted on edge devices.

Inference (Real-time):

- On Raspberry Pi 4:
  - Training was manageable (~minutes/hours).
  - Inference time per prediction is extremely fast, enabling real-time processing of the data stream at low power consumption. Speedup magnitude here is even more dramatic.

Dataset	Model	Hardware	Total Training (s)	Inference per record (μs)
Brest	FLP-XR	Xeon+A100	69	0.56
	SotA		35,909	1,345
	FLP-XR	NUC-i7	186	12
	FLP-XR	RasPi 4	1,119	118
Piraeus	FLP-XR	Xeon+A100	60	0.61
	SotA		13,161	1,327
	FLP-XR	NUC-i7	123	14
	FLP-XR	RasPi 4	811	121
Aegean	FLP-XR	Xeon+A100	86	0.28
	SotA		14,637	1,410
	FLP-XR	NUC-i7	643	10
	FLP-XR	RasPi 4	4,020	105

Comparison of training and inference(per AIS message) times between FLP-XR and Nautilus.

Deployment Potential in Edge/Fog Computing

The Big Advantage: Extreme speed gains!

- Training takes minutes/hours on a laptop.
- Inference runs instantly or near-instantly, even on low-power devices like Raspberry Pi.

Efficiency Gains Explained:

- Dedicated XGBoost training hardware (like the NUC) optimized for speed helps significantly.
- Efficient Rust preprocessing minimizes overhead before model inference.

Practical Impact: This allows complex prediction tasks to be handled locally ("in situ"), reducing latency and bandwidth usage. It paves the way for deploying large-scale analytics services directly on Edge hardware, independent of cloud connectivity.

Example Use-Case	Hardware Platform	FLP-XR capabiltiies
Real-time predictions using Edge/Fog deployments in busy ports with high vessel density, using limited hardware.	Raspberry Pi 4	**~10K records/second**
Fog/Cloud deployment in large sea areas, requiring predictions using hardware with low power consumption	Intel NUC (i7)	**Up to 100K+ records/second**
Global-scale monitoring systems or highly accurate analysis where latency is less critical	Cloud/GPU Cluster	**Very large datasets/batch sizes**

Conclusion & Future Directions

FLP-XR addresses the challenge of **fast and accurate** maritime vessel location prediction.

By leveraging XGBoost and a tailored data processing pipeline (including Rust preprocessing), it achieves remarkable speed improvements against SotA (~300x training, ~3000x inference) while often maintaining SotA accuracy levels or being very competitive in many scenarios.

Key Contribution: Demonstrates the viability of gradient boosting for complex maritime prediction tasks previously dominated by computationally expensive deep learning models (like RNNs). Highlights that efficiency isn't just about model structure but also preprocessing and hardware adaptation.

DATA STORIES
Data Science Lab. @ University of Piraeus

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

MobiSpaces
new data spaces for green mobility