Thesis Topic: Machine Learning for Gap Filling in Hydromast Time-Series Data

What's this about?

The Hydromast sensor collects velocity measurements in rivers and other water bodies, producing valuable time-series data for hydrological studies. However, real-world deployments are prone to data gaps and malfunctions caused by sensor dropouts and harsh environmental conditions. These problems reduce the reliability of downstream analyses. This project investigates machine learning methods to reconstruct missing Hydromast velocity data and detect faults. Both classical approaches (e.g., k-Nearest Neighbors, regression-based interpolation) and advanced time-series models (e.g., LSTMs, temporal autoencoders) will be explored. The work will involve creating synthetic gaps in continuous datasets, testing model performance, and applying the best approaches to real cases of missing Hydromast data.

What will you learn?

- Data analysis from the sensors: cleaning, preparing input features
- Implementing machine learning models for forecasting
- Design metrics (RMSE, MAE, correlation) to compare reconstructions on both synthetic and real missing segments.

Why is this important?

Environmental monitoring depends on continuous and reliable data, but sensors in extreme conditions frequently fail or produce incomplete datasets. By applying machine learning, this project aims to provide more accurate reconstructions, ensuring that Hydromast data remains usable for hydrological analysis, validation studies, and real-time monitoring. The results will demonstrate how ML can complement innovative environmental sensing technologies, bridging the gap between raw sensor outputs and reliable scientific insights.

Requirements: Good command of python and statistics.

Supervisor: Bauyrzhan Zhakanov, Laura Piho