5th ECMWF-ESA ML Workshop Bologna 13–17/04/2026

Hyperparameter-Tuned ML Models for Predicting Sea-Level Anomalies in the Venice Lagoon

MedEWSa

Introduction: Accurate forecasting of sea-level height is crucial for improving Early Warning Systems in vulnerable coastal areas such as the Venice Lagoon. Traditional approaches often struggle to represent the complex and nonlinear interactions driving sea-level variability. Within the MedEWSa project, advanced Artificial Intelligence techniques are explored to enhance predictive capabilities. In this study, multiple Machine Learning models are trained using meteorological predictors from the ERA5 reanalysis. The models aim to improve the prediction of sea-level anomalies over extended time ranges. This approach supports the development of more reliable and operational Early Warning Systems.

Authors:
Mehri Hashemi Devin [1]
Antonello Squintu[1]
Enrico Scoccimarro [1]
(1) CMCC Foundation

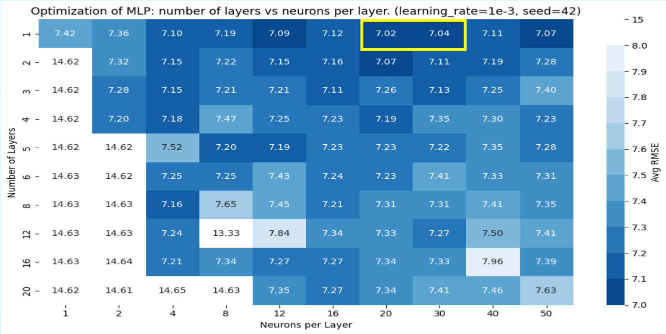
1. Data & Methodology

Data: Meteorological predictors were obtained from the ERA5 reanalysis (e.g. MSLP, precipitation and wind components over the Mediterranean region). The daily sea-level anomalies of Venice (tide-gauge observations) is used as the target.
Models: Eight models were evaluated that include of Neural Networks (MLP, Sequential, PyTorch-based architectures), tree-based ensembles (Random Forest, AdaBoost, Gradient Boosting, XGBoost) and Linear Regression.

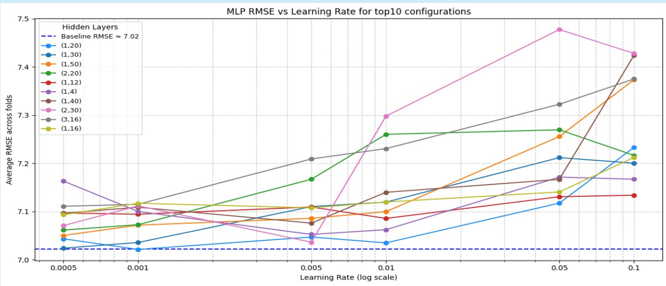
Hyperparameter optimization: Hyperparameter tuning is performed in a 5-fold cross-validation (2004–2019) framework using standardized predictors, with tailored multi-step procedures.

2. Neural Networks hyperparametrization

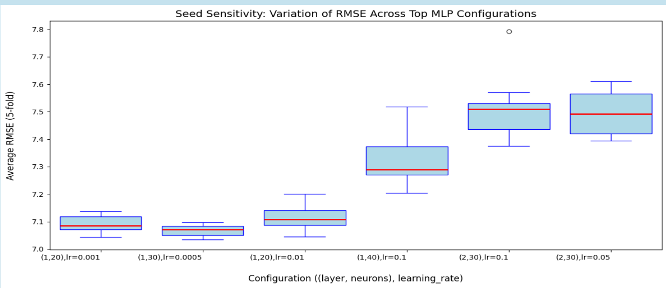
- Number of layers vs number of neurons per layer



- Learning rate tuning for top configurations (RMSE-based)



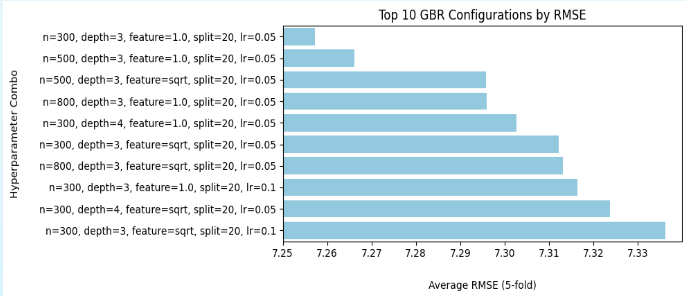
- Robustness analysis via multiple random seeds on best and worst configurations



- Final selection using statistical testing (p-value for top 2 config: 0.06)

3. Tree-Based hyperparametrization

Tree-based ensembles were tuned using a grid search over key hyperparameters number of trees, maximum depth, max_features, min_samples_split, min_samples_leaf. Similar to NN models then the top-performing configurations were further refined including learning rate and evaluated through seed sensitivity analysis to assess robustness.

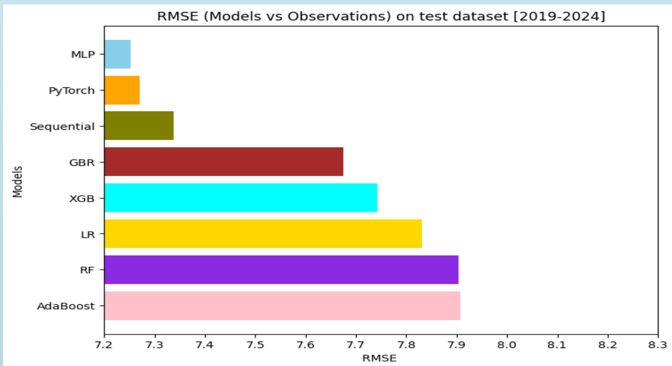


4. Final Model Selection

Final hyperparameter configurations after tuning and selection, reflecting optimal performance, robustness, and statistical significance.

Model	Selected parameters
LR	NA
Sequential	1 layer, 40 neurons per layer, learning rate 0.0005
MLP	1 layer, 30 neurons per layer, learning rate 0.0005
pyTorch	1 layer, 40 neurons per layer, learning rate 0.001
RF	700 trees, maximum depth 14
AdaBoost	600 trees, maximum depth 6, learning rate 1
GBR	300 trees, maximum depth 3, learning rate 0.05, features share considered in each iteration (feature): 1
XGB	300 trees, maximum depth 5, features share considered in each iteration (colsample): 0.7

The optimized models were trained again on the whole training dataset and tested on the period from 2019 to 2024.



5.Conclusion

- Neural Networks achieve the highest predictive skill, outperforming linear regression and tree-based models
- Hyperparameter tuning and robustness analysis play a key role in improving model performance and stability
- The optimized framework is well-suited for integration into operational Early Warning Systems, improving forecasting capabilities

cmcc Centro Euro-Mediterraneo sui Cambiamenti Climatici

CITTÀ DI VENEZIA