Università di Genova | DICCA DIPARTIMENTO DI INGEGNERIA CIVILE, CHIMICA E AMBIENTALE

MeteOcean
www.dicca.unige.it/meteocean meteocean@dicca.unige.it

A machine learning approach to achieve accurate time series forecast of sea-wave conditions

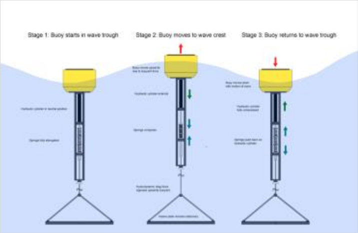
Giulia Cremonini, Daniele Lagomarsino, Agnese Seminara, Giovanni Besio

giulia.cremonini@edu.unige.it

Motivation

Università di Genova | DICCA DIPARTIMENTO DI INGEGNERIA CIVILE, CHIMICA E AMBIENTALE

There are myriad reasons why predicting wave conditions is important



Prediction of wave conditions is very important in several engineering applications

THE MAIN AIM

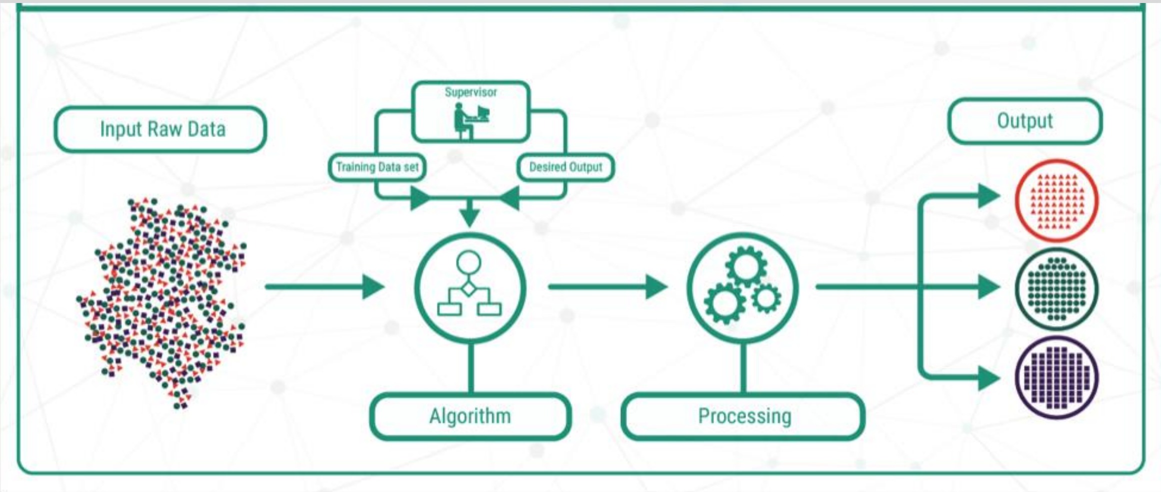
Development of a machine learning framework for the estimation and forecasting of sea conditions

Università di Genova | DICCA DIPARTIMENTO DI INGEGNERIA CIVILE, CHIMICA E AMBIENTALE

Supervised Machine Learning

Because wave models can be computationally expensive, a new approach with machine learning is developed here

S. L. is the machine learning task of learning a function that maps an input to an output based on example input-output pairs.



A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples.

WHY?

Supervised Machine Learning

Università di Genova | DICCA DIPARTIMENTO DI INGEGNERIA CIVILE, CHIMICA E AMBIENTALE

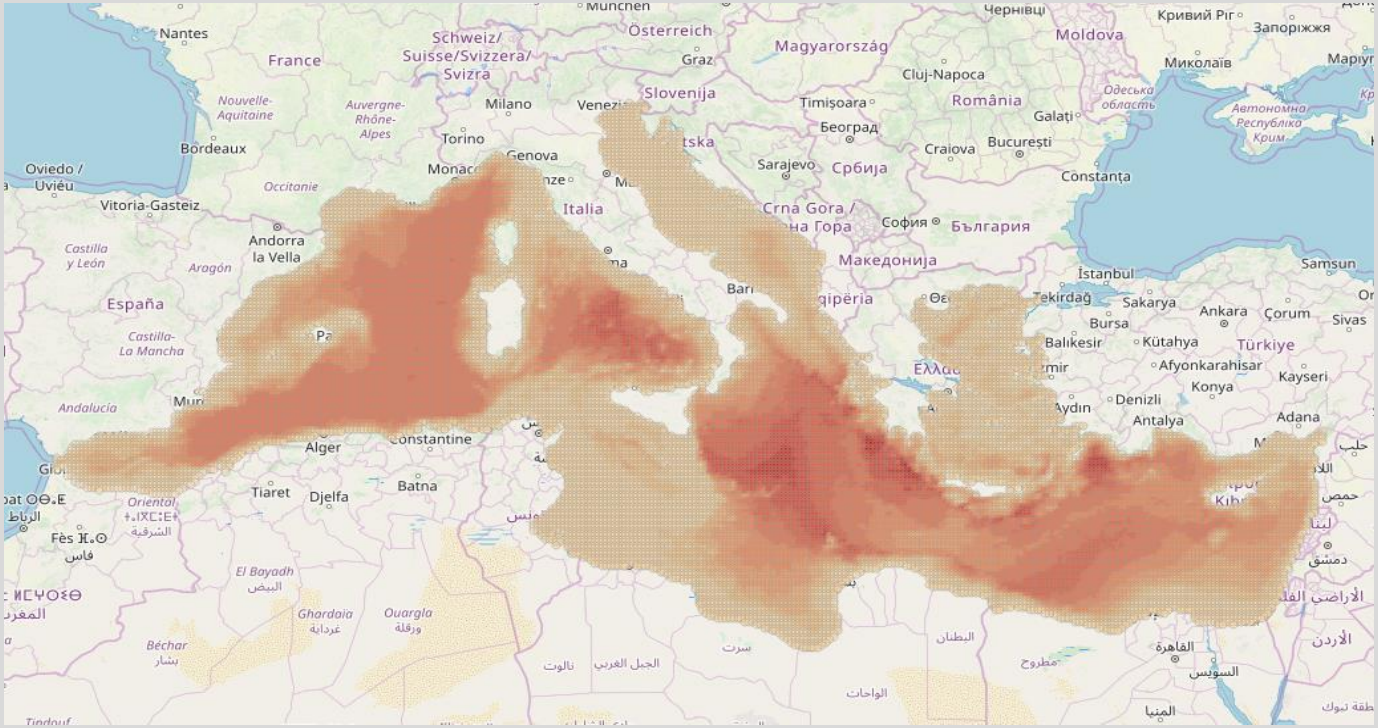
- Database with high spatial and temporal resolution (> 300 000 data) ⟶ **LARGE AMOUNT OF DATA**
- Accurate forecast of wave conditions ⟶ **DECREASE OF COMPUTATION COSTS**
- Improvement of the evaluation for various lead times and for different met-ocean variables

Università di Genova
DICCA DIPARTIMENTO DI INGEGNERIA CIVILE, CHIMICA E AMBIENTALE

Data & Methods

- Hindcast Database: with re-analysis of atmospheric and wave conditions over the whole Mediterranean Basin
- 40 years (1979-2018) time series of wave and wind parameters, hourly defined, with a 0,1°x0,1° spatial resolution

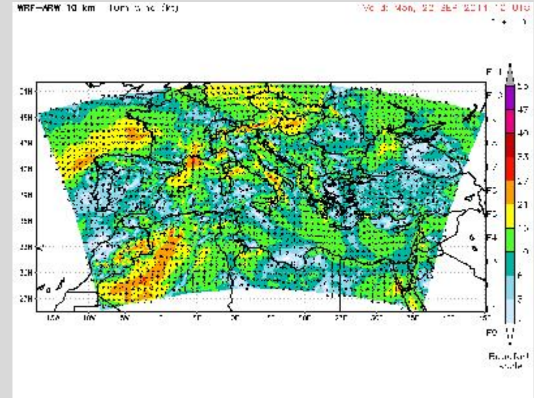
In our analysis we retain only the wave height time series



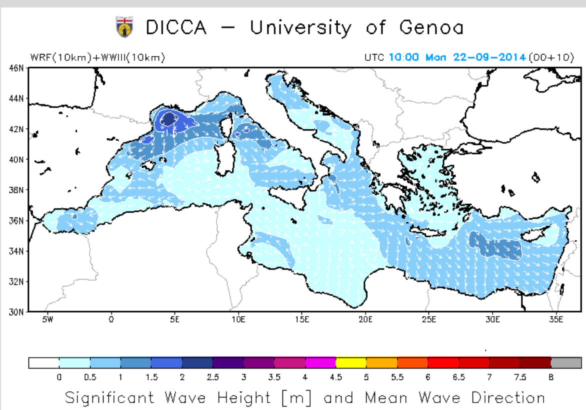
http://www3.dicca.unige.it/meteocean/hindcast.html

The numerical model chain

Università di Genova
DICCA DIPARTIMENTO DI INGEGNERIA CIVILE, CHIMICA E AMBIENTALE



1. WRF-ARW: the non-hydrostatic mesoscale model to provide the 10m wind fields of Mediterranean Sea



2. WaveWatchIII used for re-analysis of wave conditions, with a spatial resolution of 10 km at the latitude of 45° N

Forecast Service: 120 hours wave forecasts for the Mediterranean Basin, published daily.

Università di Genova | DICCA DIPARTIMENTO DI INGEGNERIA CIVILE, CHIMICA E AMBIENTALE

Data & Methods

Regularized Linear Regression

Regularized Least Squares is a family of methods for solving the least-squares problem while using regularization to further constrain the resulting solution

$$\min_{w \in \mathbb{R}^D} \frac{1}{n} \sum_{i=1}^{n} (y_i - w^\top x_i))^2 + \lambda w^\top w, \quad \lambda \geq 0.$$

where λ is a regularizer and helps preventing overfitting by controlling the stability of the solution

My problem is: $Y = wX$

and the *goal* is to find the coefficients (w) in order to solve the equation for a specific value of λ.

Università di Genova | DICCA DIPARTIMENTO DI INGEGNERIA CIVILE, CHIMICA E AMBIENTALE

Data & Methods

Our Approach

1. Definition of training- validation dataset (30 years) and test set (10 years)
2. K-fold Cross-Validation:

- Partition of the T-V set into K subsets
- For each subset a RLS analysis is performed and the RMSE error is calculated
- A single subsample is retained as the validation data for testing the model, and the remaining k − 1 subsamples are used as training data

→ ***the mean λ value (over the K iterations), referred to the smallest error, defines the model to use***

From a matrix point of view, in each analysis

- Input X: is built considering a short time frame (Memory, ΔT) of H, shifted of an hour
- Output Y: is the $\Delta T + y^*$ element of the series (y^* represent the lead time of prediction)

$$Hs = [x_1, \dots, x_N]$$

$$X = \begin{bmatrix} x_1 & \cdots & x_{\Delta T} \\ \vdots & \ddots & \vdots \\ x_{N-\Delta T+1} & \cdots & x_{N-y^*} \end{bmatrix}$$

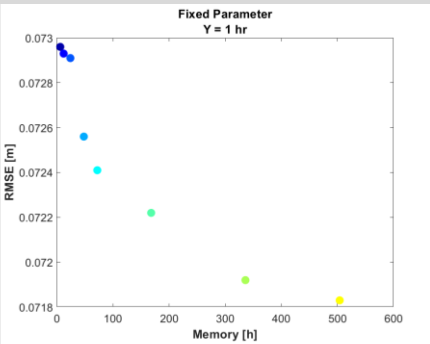
$$Y = \begin{bmatrix} y_{\Delta T+y^*} \\ \vdots \\ y_N \end{bmatrix}$$

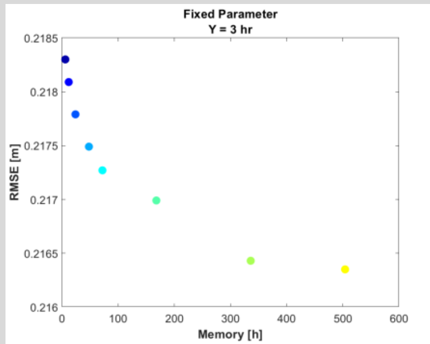
Analysis of each combination $\Delta T - y^*$: $\Delta T = 6, 12, 24, 48, 72, 168, 336, 504\ hr$; $y^* = 1, 3, 6, 12, 24, 48, 72, 96, 120, 144, 168\ hr$

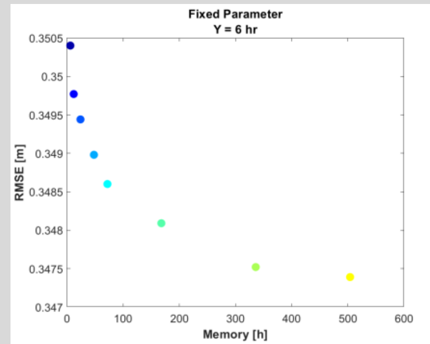
Results

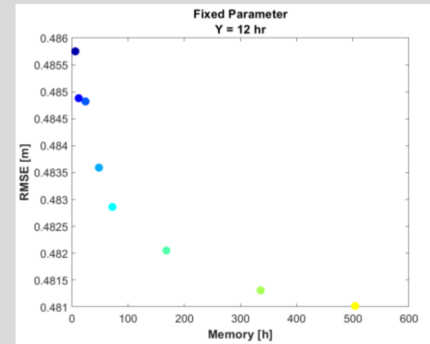
Performances

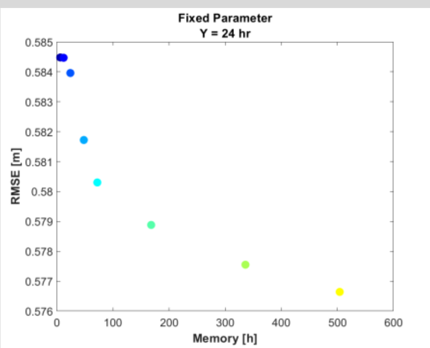
The RMSE-ΔT plots for each lead times

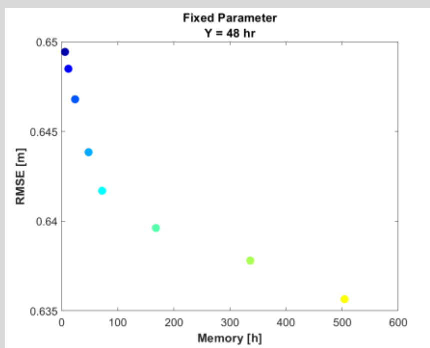


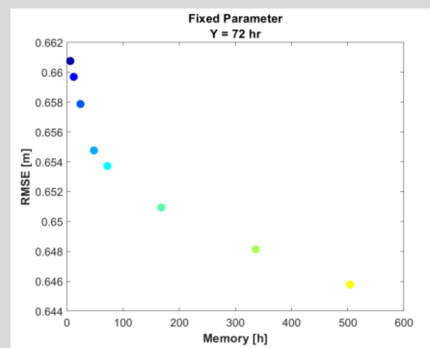


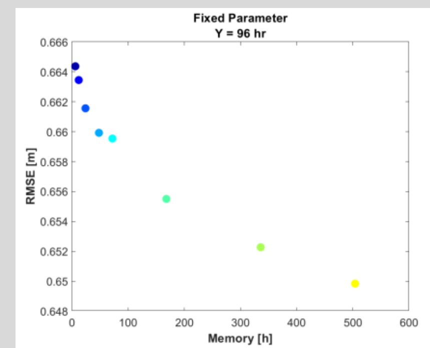


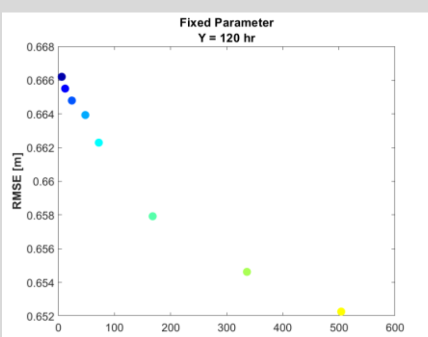


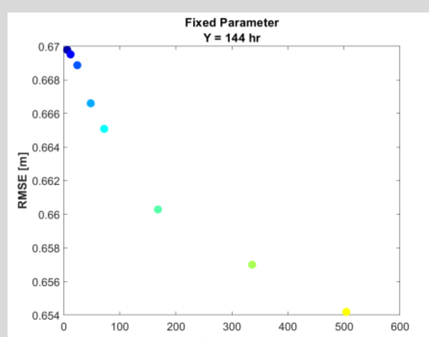


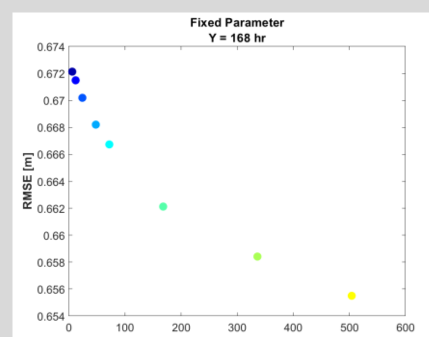












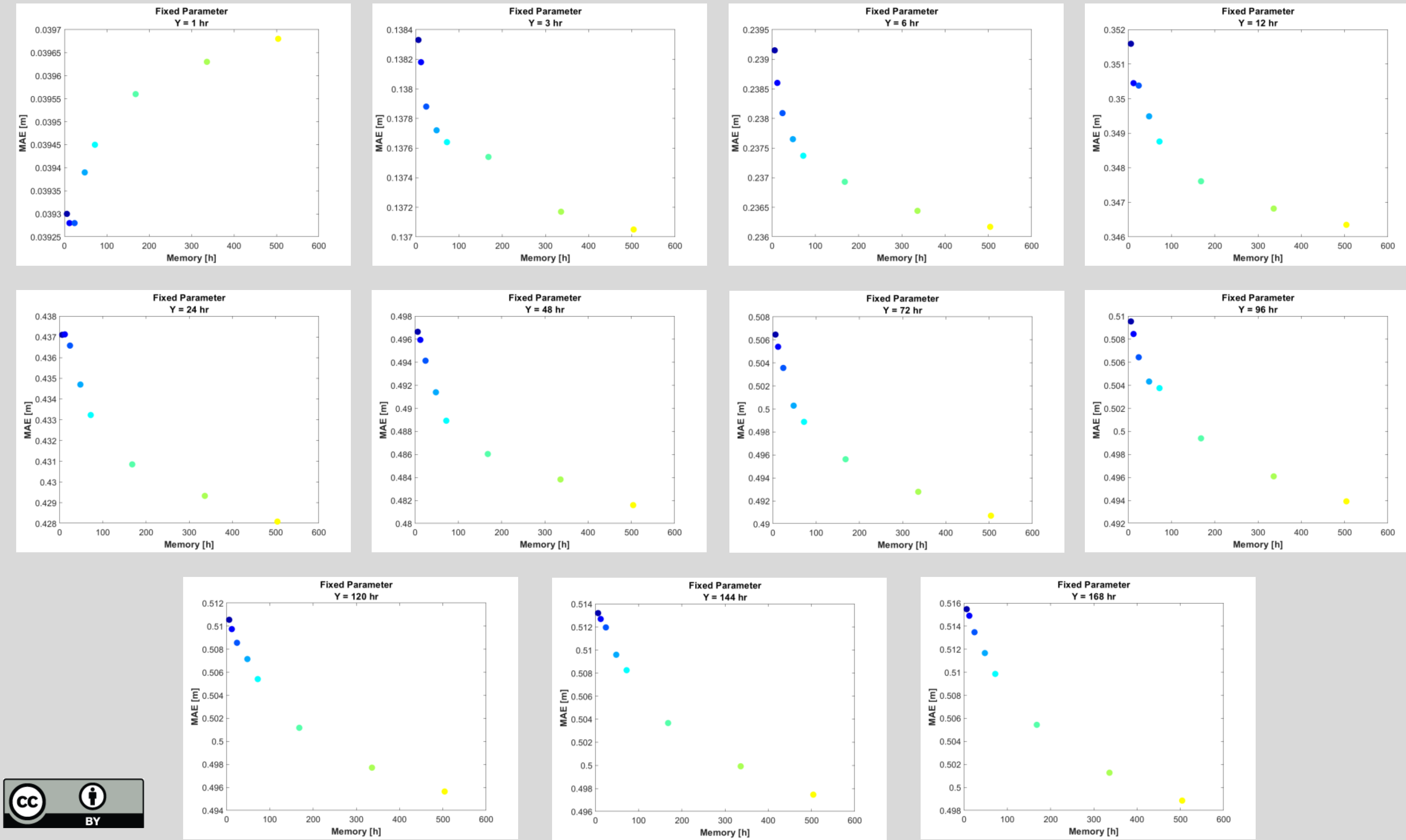
Università di Genova
DICCA DIPARTIMENTO DI INGEGNERIA CIVILE, CHIMICA E AMBIENTALE

Università di Genova
DICCA DIPARTIMENTO DI INGEGNERIA CIVILE, CHIMICA E AMBIENTALE

Results

Performances

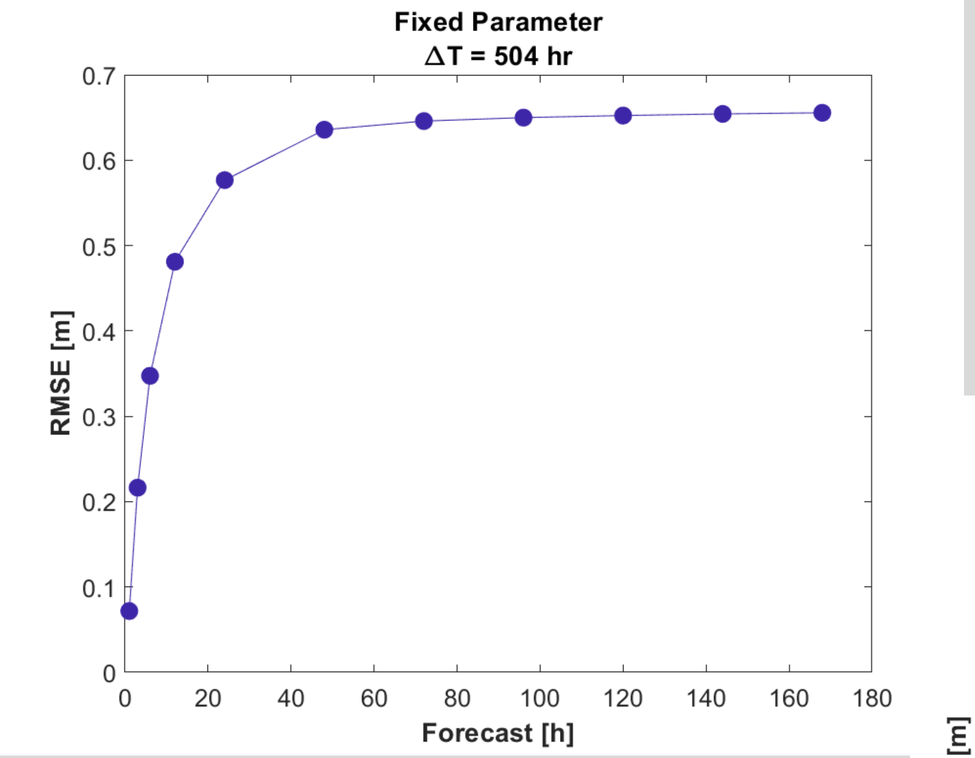
The MAE-ΔT plots for each lead times

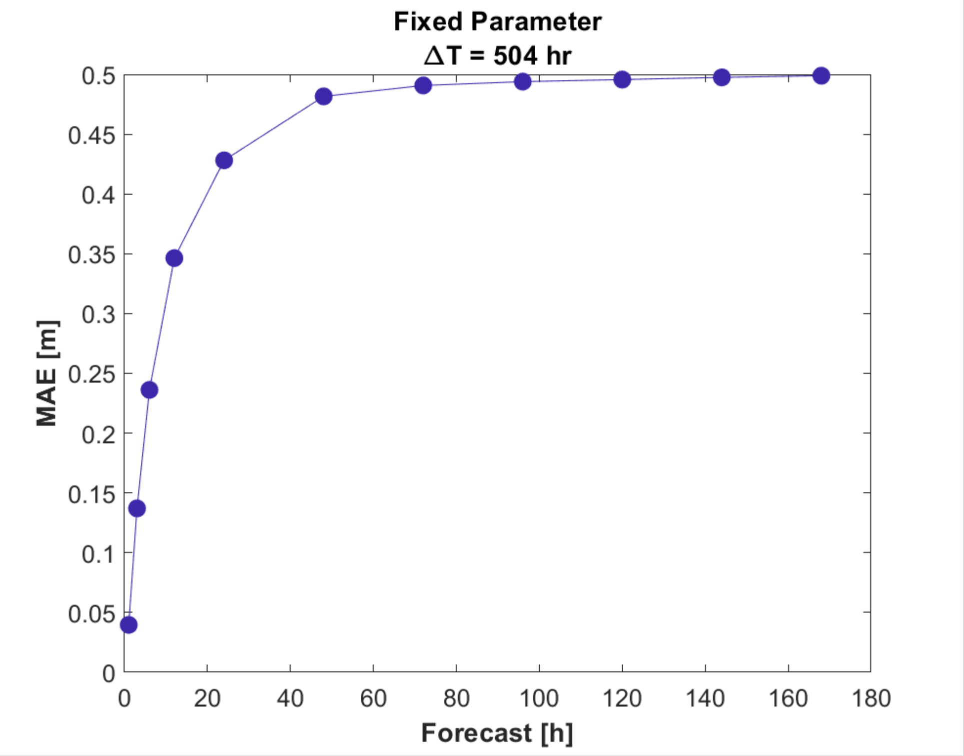


Università di Genova
DICCA DIPARTIMENTO DI INGEGNERIA CIVILE, CHIMICA E AMBIENTALE

Results

Best Performance: $\Delta T = 504hr$





Università di Genova | DICCA DIPARTIMENTO DI INGEGNERIA CIVILE, CHIMICA E AMBIENTALE

Conclusions

- We report the RMSE and MAE trend for each value of forecast horizon analysed
- A way to see the best results obtained is selecting the minimum value of the error as the window increases
- The best result obtained is for: $\Delta T = 504\ \mathrm{hr}$

… And Then…

- Multivariate linear regression: including other features of the wave field
- Explore Non-Linear Models