Data Assimilation Project for MOHID

A proposal

by Mariangel (Angie) Garcia

TÉCNICO LISBOA

MARETEC
MARINE, ENVIRONMENT & TECHNOLOGY CENTER

Something in common

- Understand behavior of atmosphere/ocean
- Work with community of scientists to solve problems that matter
- "… and to foster the transfer of knowledge and technology **for the betterment of life on earth**"

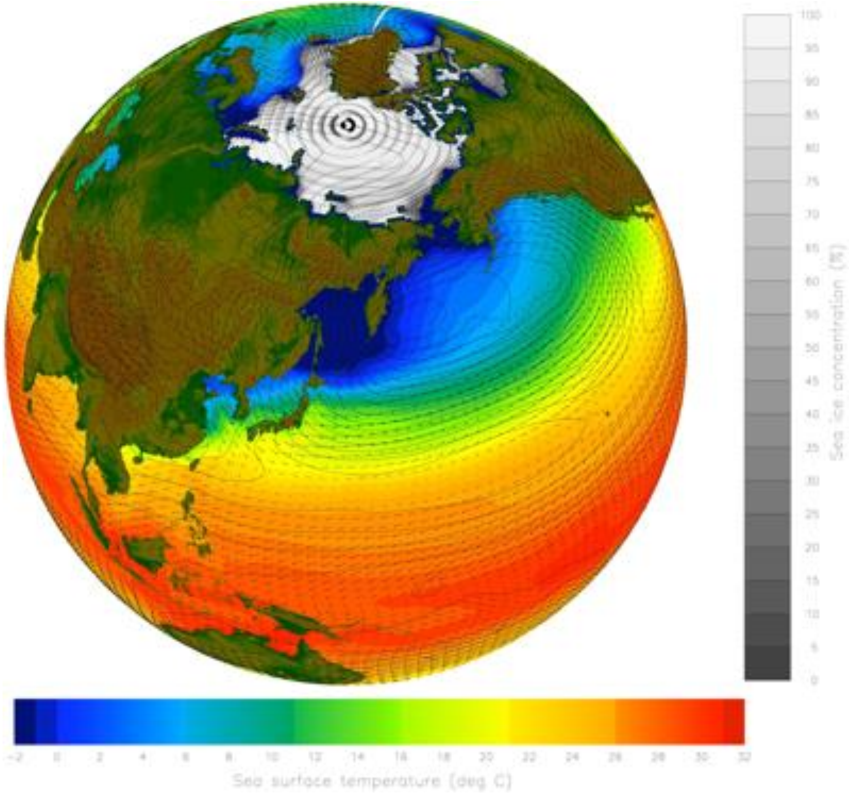
NCAR's Mission

Tools to understand climate

Interagency Ocean Observing Committee (IOOC)

Observations

Tools to understand climate

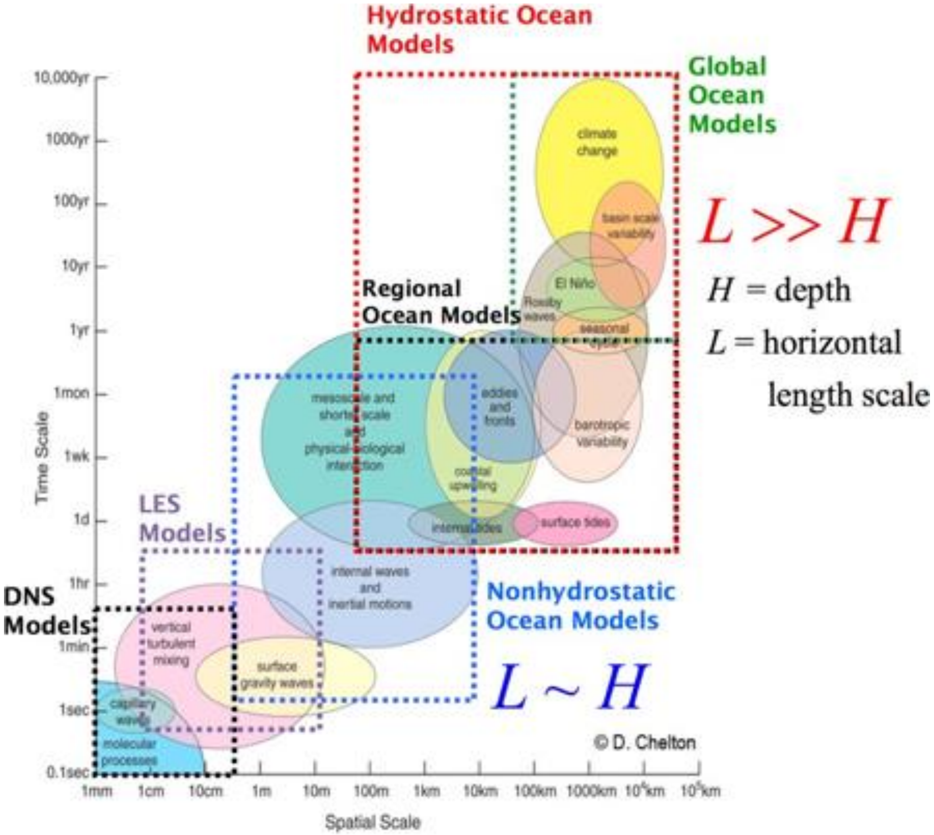


Numerical Models

Practical Applications

Modeling Challenge

photo: Raincoast GeoResearch



Ocean Modeling & Data Assimilation

Previews work

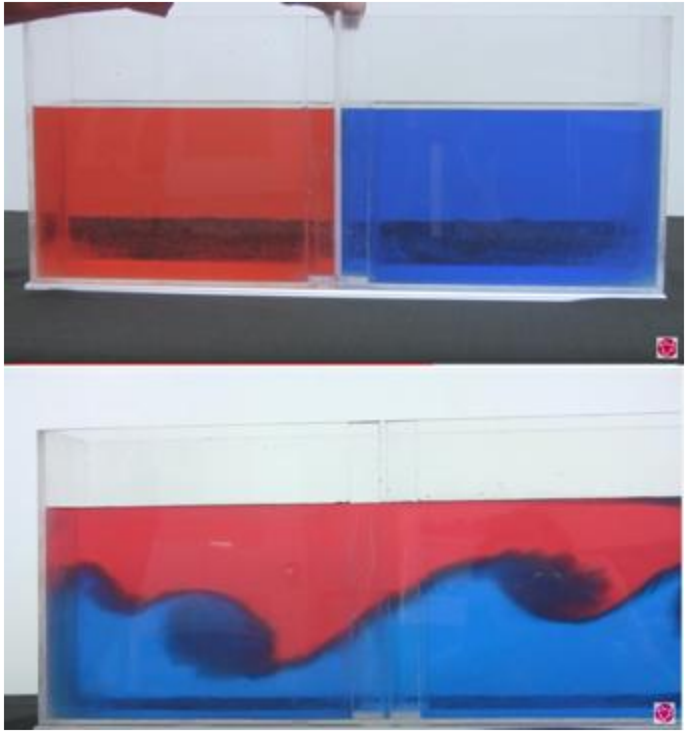
Former Research Project

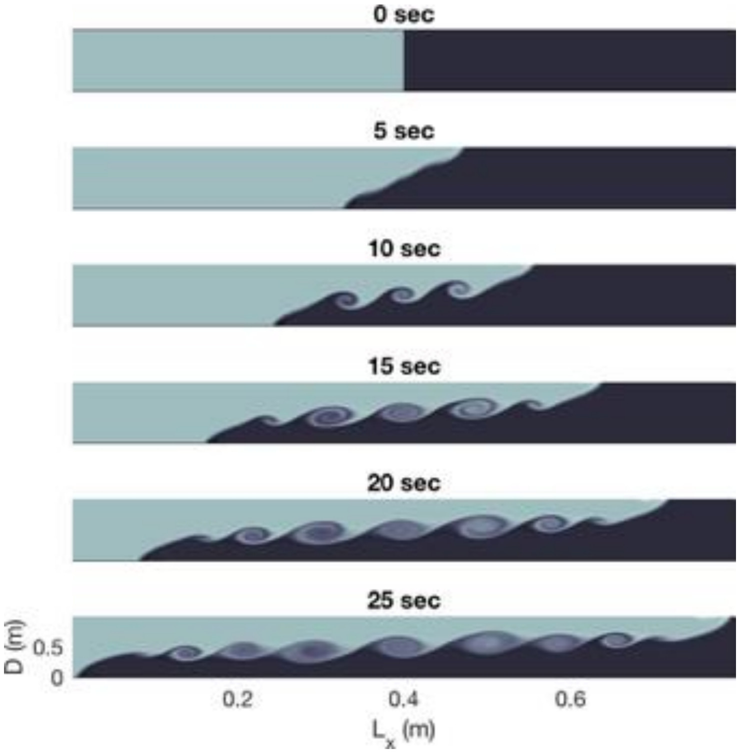
GCEM framework

General Curvilinear Environmental Modeling (GCEM)

1. *Unified Curvilinear Ocean Atmosphere Model (UCOAM)*
   - **General Curvilinear Coastal Ocean Model (GCCOM)**
   - *General Curvilinear Atmosphere Model (GCAM)*
2. *Distributed Coupling Tools (DCT)*
3. Computational Environment (CE)
   - *Cyber-infrastructure Web Application Framework (CyberWeb)*
4. **Data Assimilation Unit (DAU)**

Benchmark Test Cases: 3D Lock Release





Benchmark Test Cases: 3D Internal Wave Beam

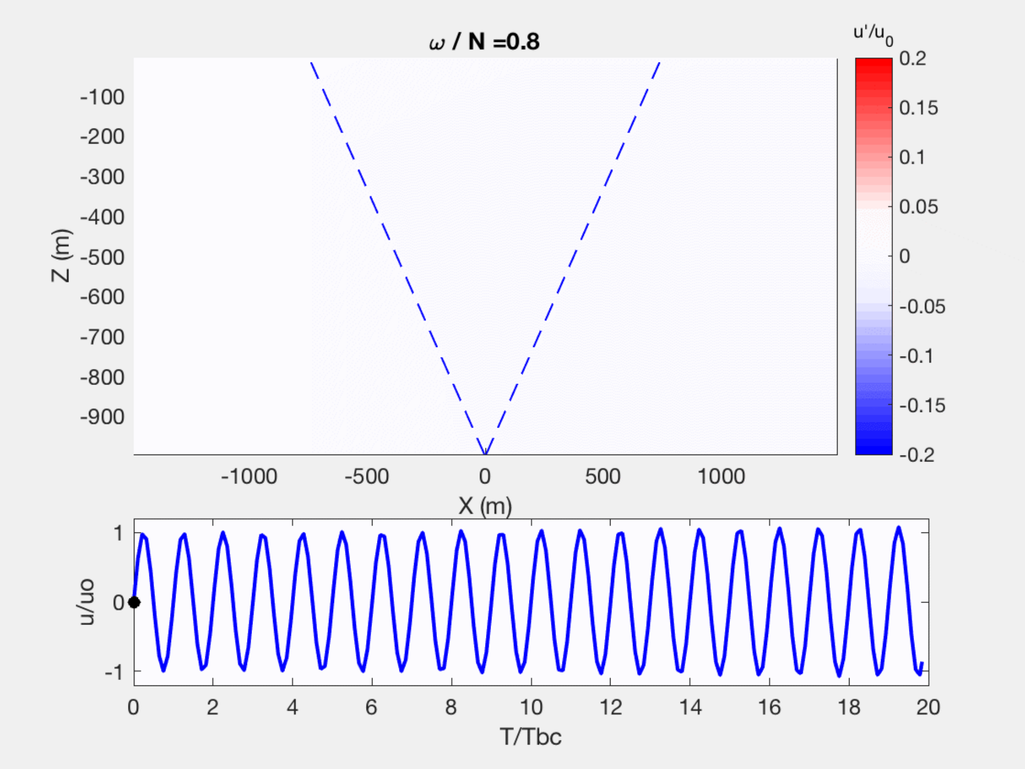


Fig 12. Instantaneous stream-wise velocity at t = 19.5T_bc.The theoretical beam angle are shown between lines at each side, dashed hydrostatic, solid line non-hydrostatic.

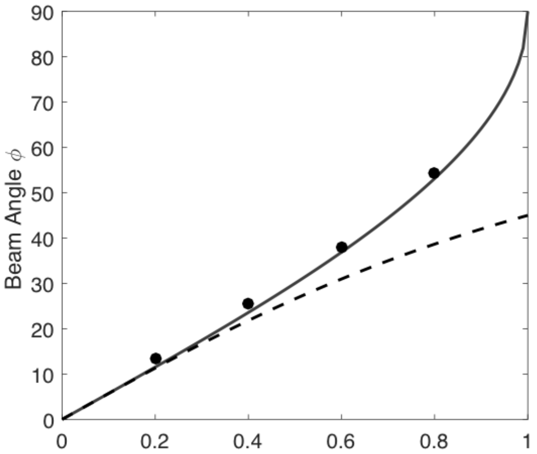
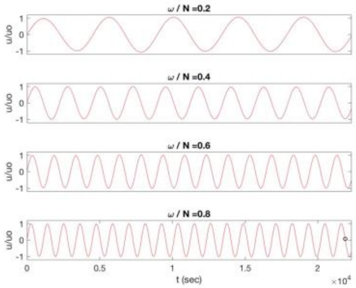


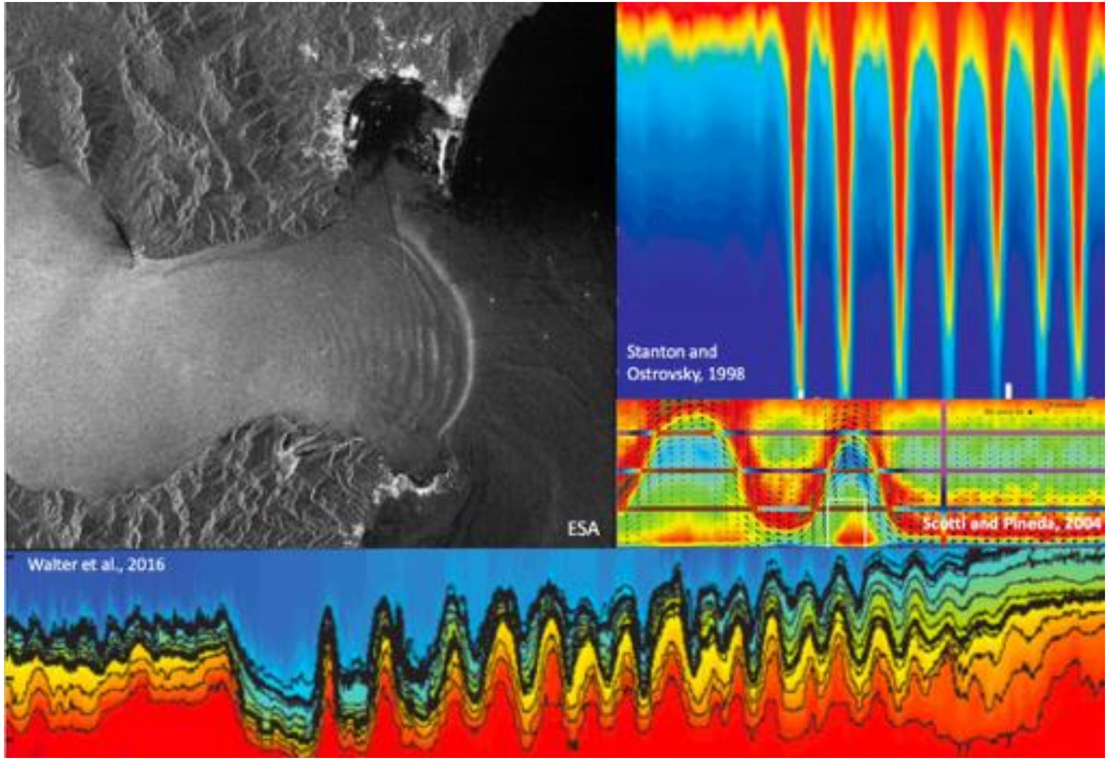
Fig 11. The theoretical curves for the internal wave beam angle, dashed line hydrostatic, solid line , non-hydrostatic and black dot numerical value



- Mariangel Garcia et al. (2018). “Validation of the Non-hydrostatic General Curvilinear Coastal Ocean Mode (GCCOM) for Stratified Flows ”. In: *Under preparation*
- P. F. Choboter et al. (2016). “Nesting nonhydrostatic GCCOM within hydrostatic ROMS for multiscale Coastal Ocean Modeling”. In: *OCEANS 2016 MTS/IEEE Monterey*, pp. 1–4
- Mohammad Abouali and Jose E Castillo (2013). “Unified Curvilinear Ocean Atmosphere Model (UCOAM): A vertical velocity case study”. In: *Math. Comput. Model.* 57.9-10, pp. 2158–2168
- Sean Vitousek and Oliver B. Fringer (2014). “A nonhydrostatic, isopycnal-coordinate ocean model for internal waves”. In: *Ocean Modelling* 83, pp. 118 –144. ISSN: 1463-5003
- O B Fringer, M Gerritsen and R L Street (2006). “An unstructured-grid, finite-volume, nonhydrostatic, parallel coastal ocean simulator”. In: *Ocean Model.* 14.3-4, pp. 139–173. ISSN: 14635003
- Zhigang Lai et al. (2010). “A nonhydrostatic version of FVCOM: 1. Validation experiments”. In: *Journal of Geophysical Research: Oceans* 115.11, pp. 1–23
- Masoud Jalali, Narsimha R. Rapaka and Sutanu Sarkar (2014). “Tidal flow over topography: effect of excursion number on wave energetics and turbulence”. In: *Journal of Fluid Mechanics* 750, 259283
- M. Garcia et al. (2016). “Interfacing an ensemble Data Assimilation system with a 3D nonhydrostatic Coastal Ocean Model, an OSSE experiment”. In: *OCEANS 2016 MTS/IEEE Monterey*, pp. 1–11

Reference

Solving sub-mesoscale processes



Density stratification within the ocean interior supports the propagation of internal gravity waves.

FACT: Model errors are currently inevitable.

Uncertainty Quantification(UQ)

UQ is the process by uncertainty is estimated in a system.

$$Y - \hat{Y} = e$$

Where *e* is an unknown error

Uncertainty Reduction(UR)

UR which has the purpose of reducing the uncertainty in a system.

FACT: Model errors are currently inevitable.

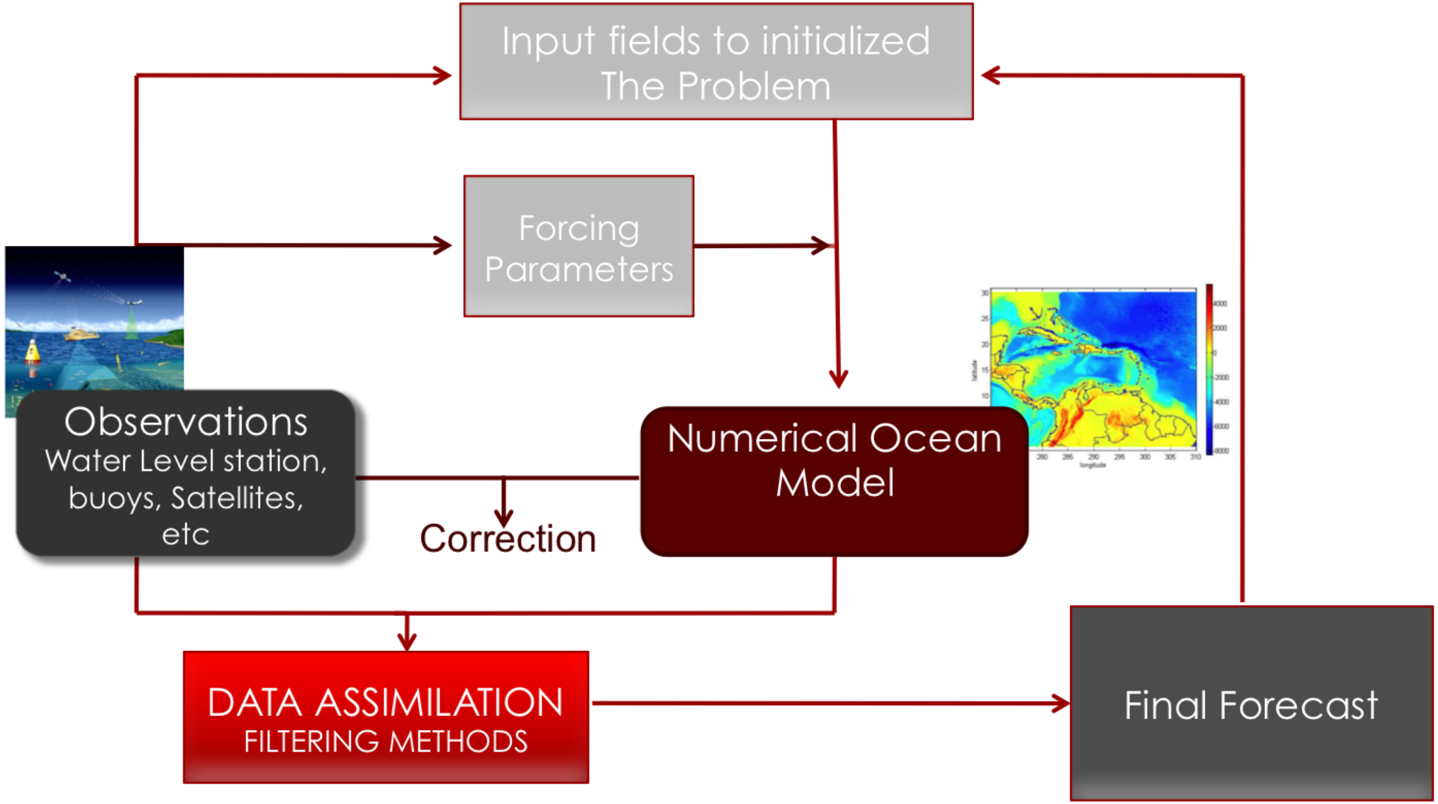
DA attempt to do UQ, it consists of three components:

1. set of observations
2. a dynamical model
3. data assimilation scheme

The main goal

Reduce the uncertainty in the entire system

Data Assimilation Philosophy



Question to be addressed

Question to be addressed?

- **What models do we use?**
- What assimilation algorithms do we use?
- What type of observations do we assimilate?
- What are the observation errors?
- What are the model and analysis errors?

MOHID
Water Modelling System

Question to be addressed

Question to be addressed?

- What models do we use?
- **What assimilation algorithms do we use?**
- What type of observations do we assimilate?
- What are the observation errors?
- What are the model and analysis errors?

Assimilation approaches

Variational approach

- Optimal Interpolation
- 3DVar
- 4DVar

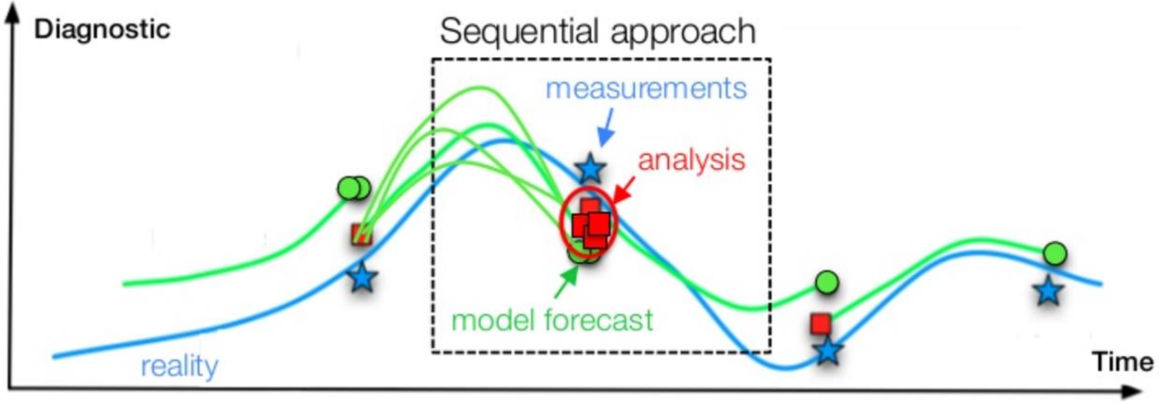
Sequential approach

- Kalman Filter *Kalman,1960*
- EnKF *Evensen,1994*
- ETKF *Bishop& Hunt,2001*
- EAKF*Anderson,2001*
- Particle Filter *Non Gaussian*
- ESRKF*Tippett,2003*
- Hybrid: OI EnsKF,SSEnsKF

[8]Kalnay 2003.
[9]Evensen 2006.

Ensemble Kalman Filter

→ Key idea: "optimal combination of observations and forward model"



Ensemble Kalman filter (EnKF)

- Forecast step → uncertainty propagation
  - Explicit propagation of the error statistics
  - Nonlinear extension of the Kalman filter
- Analysis step → Kalman filter update equation

Stochastic characterization
Estimation of error covariance matrices

Kalman gain matrix

KF Vs EnKF

variable	definition
$\boldsymbol{x}$	model state
$\boldsymbol{\eta}$	process noise
$\boldsymbol{\epsilon}$	measurement noise
$\boldsymbol{y}^o, \boldsymbol{y}^f$	observations
	(^{o} measured and ^{f} predicted)
N	ensemble size

variable	definition
$\boldsymbol{P}$	covariance of model state
$\boldsymbol{Q}$	covariance of process noise
$\boldsymbol{R}$	covariance of measurement noise
$\boldsymbol{M}$	model time propagation operator
$\boldsymbol{H}$	observation operator
$\boldsymbol{K}$	Kalman gain

Forecast (^{f})

KF	EnKF
$\boldsymbol{x}^f_{i+1} = \boldsymbol{M}_i \boldsymbol{x}^a_i$ $\boldsymbol{P}^f_{i+1} = \boldsymbol{M}_i \boldsymbol{P}^a_i \boldsymbol{M}^T_i + \boldsymbol{Q}_i$	$\boldsymbol{x}^{ef}_{j,i+1} = \boldsymbol{M}_i(\boldsymbol{x}^{ea}_{j,i}) + \boldsymbol{\eta}^e_{j,i}$ $\boldsymbol{x}^f_{i+1} = \frac{1}{N}\sum_{j=1}^{N} \boldsymbol{x}^{ef}_{j,i+1}$ $\boldsymbol{P}^f_{i+1} = \frac{1}{N-1}\sum_{j=1}^{N}(\boldsymbol{x}^{ef}_{i+1} - \boldsymbol{x}^f_{i+1})(\boldsymbol{x}^{ef}_{i+1} - \boldsymbol{x}^f_{i+1})^T$

$$\boldsymbol{K}_{i+1} = \boldsymbol{P}^f_{i+1}\boldsymbol{H}^T_{i+1}(\boldsymbol{H}_{i+1}\boldsymbol{P}^f_{i+1}\boldsymbol{H}^T{}_{i+1} + \boldsymbol{R}_{i+1})^{-1}$$

Analysis (^{a})

KF	EnKF
$\boldsymbol{x}^a_{i+1} = \boldsymbol{x}^f_{i+1} + \boldsymbol{K}_{i+1}(\boldsymbol{y}^o_{i+1} - \boldsymbol{H}_{i+1}\boldsymbol{x}^f_{i+1})$ $\boldsymbol{P}^a_{i+1} = \boldsymbol{P}^f_i - \boldsymbol{K}_{i+1}\boldsymbol{H}_{i+1}\boldsymbol{P}^f_{i+1}$	$\boldsymbol{x}^{ea}_{i+1} = \boldsymbol{x}^{ef}_{i+1} + \boldsymbol{K}_{i+1}(\boldsymbol{y}^o_{i+1} - \boldsymbol{H}_{i+1}\boldsymbol{x}^{ef}_{i+1} - \boldsymbol{\epsilon}_{j,i})$

Problem Statement

Estimating accurately the state variables in a sub-mesoscale process is very difficult, particularly for physical ocean models, which are highly nonlinear and require a dense spatial discretization in order to correctly reproduce the dynamics.

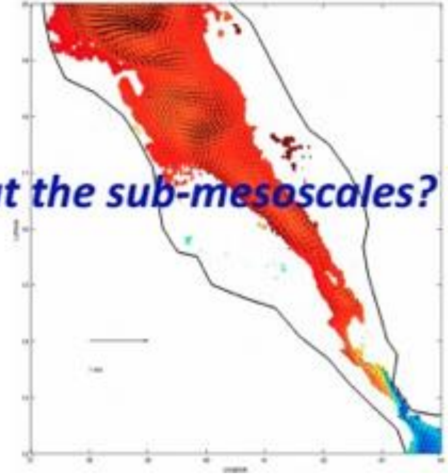
1. High computational cost incurred by a high-resolution numerical model
2. The efficiency of Kalman Filter in sub-mesoscale processes is unknown
3. Sensitivity of the model to perturbation
4. Resolution and Instrument error can affect the forecast

Problem Statement

- EnKFs are being proven efficient for ocean data assimilation at the mesoscale, typically ~O(10km)

We are observing finer faster scales

and our models can now resolve these scales



- ***Will the EnKF be as efficient at the sub-mesoscales?***

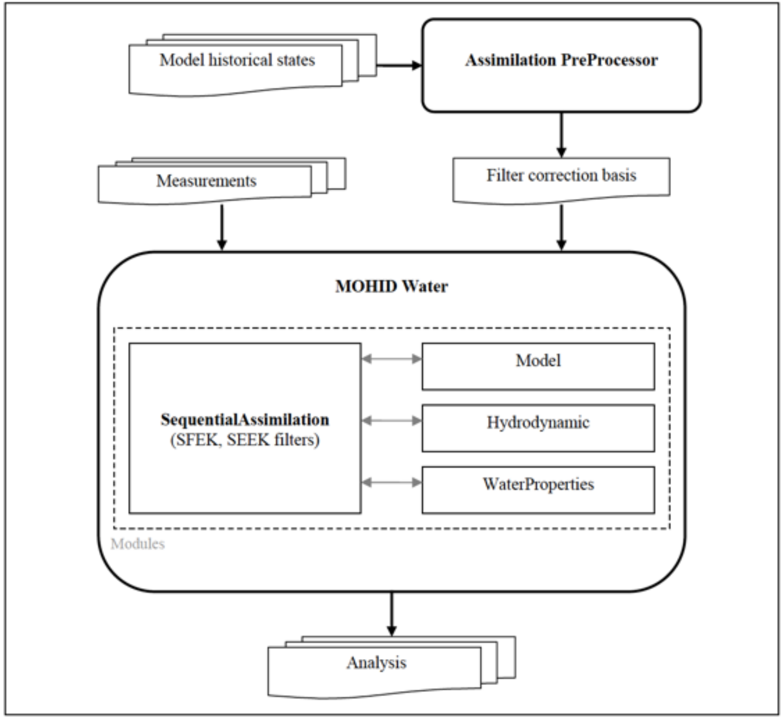
Hoteit, TAMOS workshop NCAR 2015.

How are we going to do this?

Continue Canas, 2009 work?

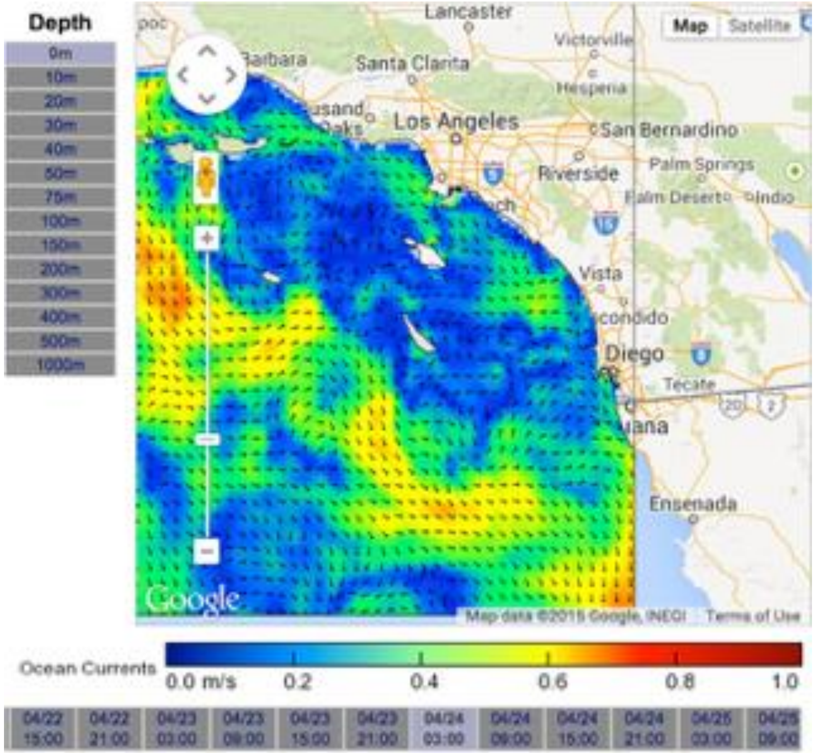
MOHID the Singular Evolutive Extended Kalman Filter (SEEK), developed by Pham et al. (1998a), and the Singular Evolutive Interpolated Kalman Filter (SEIK), developed by Pham et al. (1998b) these schemes are chosen for this first data assimilation implementation since they have reported applicability in non linear models, have small computational cost and algorithm complexity and are easily scalable to other more advanced schemes articularly adequate for data assimilation in non linear models

Figure VI-1: General scheme of the implementation of sequential data assimilation in MOHID Water framework. In grey arrows are presented the interaction between MOHID Water modules.



Another option: To implement 3DVAR/4DVAR into MOHID.

3DVAR has been very successfully implemented into ROMS with several operational applications.



Possibles Solutions

DART	OPENDA	PDAF
Data Assimilation Research Section	Open Data Assimilation	Parallel Data Assimilation Framework
NCAR	Deltares & Delft University	Computing Center of the Alfred Wegener Institute
Supported Models: WRF, CAM, MIT, AM2 COAMPS, POP ROMS + 30 more	Supported Models: Deltft3D, SWAM, open FOAM, DCSM	N/A
Assimilation Methods: EAKF, EnKF, ELTKF, Particle Filter	Assimilation Methods: EnKF, EnSR, Particle filter, 3DVAR,DuDEnFK*	**Assimilation Methods**: LETKF, LSEIK, SEEK, SEIK, ESKTF, ETKF, EnKF
	Parameter Estimation: DuD, Simplex, Powell, GLUE, Conjugate Gradient	

[10] *Data Assimilation Research Testbed - DART*.
[11] *The OpenDA data-assimilation toolbox*.
[12] Nerger and Hiller 2013.

OpenDA

From Deltare

Application of a Three-Dimensional Hydrodynamic Model for San Quintin Bay, B.C. Mexico. Validation and Calibration using OpenDA.

SAN DIEGO STATE UNIVERSITY

Mariangel Garcia[1], Isabel Ramirez[2], Martin Verlaan[3], Barbara Bailey[1] and Jose Castillo[1]

(1) Computational Sciences Research Center, San Diego State University. (2) Centro de Investigacion Cientifica y Educacion Superior de Ensenada B. Cfa. Mexico. (3) Deltares, Delft, The Netherlands.

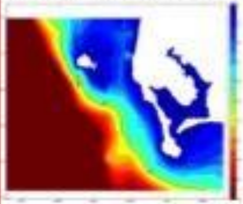
CSRC

Abstract

A 3D hydrodynamic model (Delft3D) was developed for San Quintin Bay (SQB). Calibration and validations were conducted, using measured bathymetry, water surface elevation, velocities, and temperature. The calibration period was taken in the winter season of 2010. Model predictions were evaluated graphically and statistically against field observations to quantify the accuracy of model predictions and evaluate the success of model calibration. Comparisons between model predictions and field observations of water surface elevations at interior stations indicated that the model was successfully calibrated and model predictions were highly correlated with observed water surface elevations. Agreement between observed and simulated values was based on graphical comparisons, root-mean-square errors, and principal components analysis. The objective of this study was to show that OpenDA can be used to rapidly calibrate a hydrological model.

Keywords: Calibration, San Quintin Bay, OpenDA, Delft3D.

Study Region: San Quintin Bay



The most important biological process for cultivating organisms like the Japanese oyster, "Crassostrea Gigas" is the reproductive cycle which is principally regulated by the amount and quality of the food as well as by temperature and salinity. Furthermore, the hydrodynamic state has to be known with high accuracy.

Fig 1. SQB is Located in the northwestern coast of the Baja California Peninsula (Mexico). SQB is a coastal lagoon covering an area of approximately 42 km²

Delft 3D Model

Delft3D-FLOW is a multi-dimensional (2D or 3D) hydrodynamic (and transport) simulation program which calculates non-steady flow and transport phenomena that result from tidal and meteorological forcing on a rectilinear or a curvilinear, boundary fitted grid. In 3D simulations, the vertical grid is defined following the sigma or z co-ordinate approach.

Delft3D-FLOW solves the Navier Stokes equations for an incompressible fluid, under the shallow water and the Boussinesq assumptions. In the vertical momentum equation the vertical accelerations are neglected, which leads to the hydrostatic pressure equation. In 3D models the vertical velocities are computed from the continuity equation. The set of partial differential equations in combination with an appropriate set of initial and boundary conditions is solved on a finite difference grid. [1]

Figure 2. Example of σ-grid

Model Set Up

The model grid is in Cartesian, 15 layers sigma coordinates, with 3 seconds resolution in the horizontal (~90 m). Since the main force on the bay is coming from tide [3], one major boundary forcing functions applied at the west side of the domain, the global inverse tide model (TPXO 7.2) is used to initiate Delft3D. To start the computations, it is necessary to specify initial conditions for elevation, velocity, Salinity and temperature. Initial water surface was assumed horizontal, and velocity components were set to zero through the model domain.

Monitoring Data

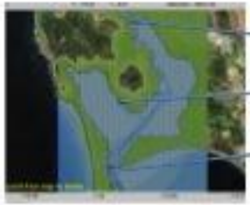


Figure 3. Curvilinear grid.

- **North Station:**
  - Water Level from Model MAREG 2010 CICESE Mexico.
- **Wet Station:**
  - Air Temperature, Wind Velocities Components, Atmospheric Pressure, humidity and rain.
- **Bay Entry Station:**
  - Pressure, water temperature, currents from Navy Secretary of states.

Data from a field campaign during the wet season (August – September 2010) was used to establish the numerical scenario to simulate the hydrodynamics of SQB.

Model Skill Assessment

The model was calibrated by adjusting the depth, the bottom frictional drag coefficient, the semidiurnal M2 and S2 tide coefficients. These coefficients are adjusted to reproduce measured tidal elevations and currents.

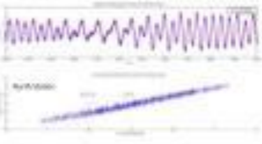


Figure 4a. Comparison of Instantaneous Water Levels at the **North Station**.

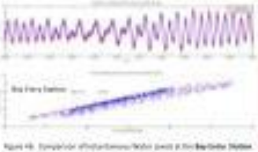


Figure 4b. Comparison of Instantaneous Water Levels at the **Bay Entry Station**.

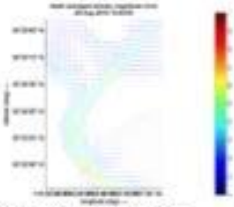


Figure 5a. Depth Average Velocity Magnitude during low tide.

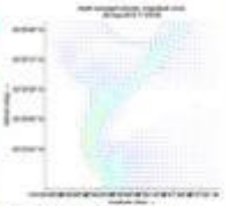


Figure 5b. Depth Average Velocity Magnitude during high tide.

Open DA Algorithms

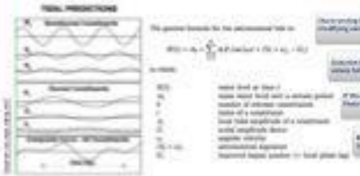
OpenDA is an open interface standard for (and free implementation of) a set of tools to quickly implement data-assimilation and calibration for arbitrary numerical models [3]

Calibration: The aim is to tune a set of parameters that is fixed in time.

Ex: Depth, Amplitude, Phase, roughness

Data Assimilation: aims to improve the starting position of the model for a forecast, so the estimates are different each cycle.

Calibrating tidal constituents



Dud Algorithm

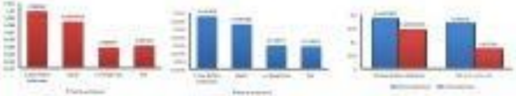


Fig 6. Root Mean Square Error comparison before and after the parameter calibration: Depth, friction and the tidal constituents M2 S2 K1 O1

CONCLUSION

The model results so far show that the Delft3D model is capable of simulation the essential processes in the San Quintin Bay, and can be forced by the tidal model.

Calibration using OpenDA, did improve the first model implementation, getting better results in the north station than the Bay Entry Station, after 14 days calibration. Results show this tool has the potential to deliver real time forecasting/nowcasting capabilities in this region.

References

[1] I. Ramirez, R. Blanco, et all. The simulation of the circulation of San Quintin Bay. Submitted to Ocean Dynamics 2012

[2] http://oss.deltares.nl/web/delft3d

[3] http://www.openda.org

Acknowledge.

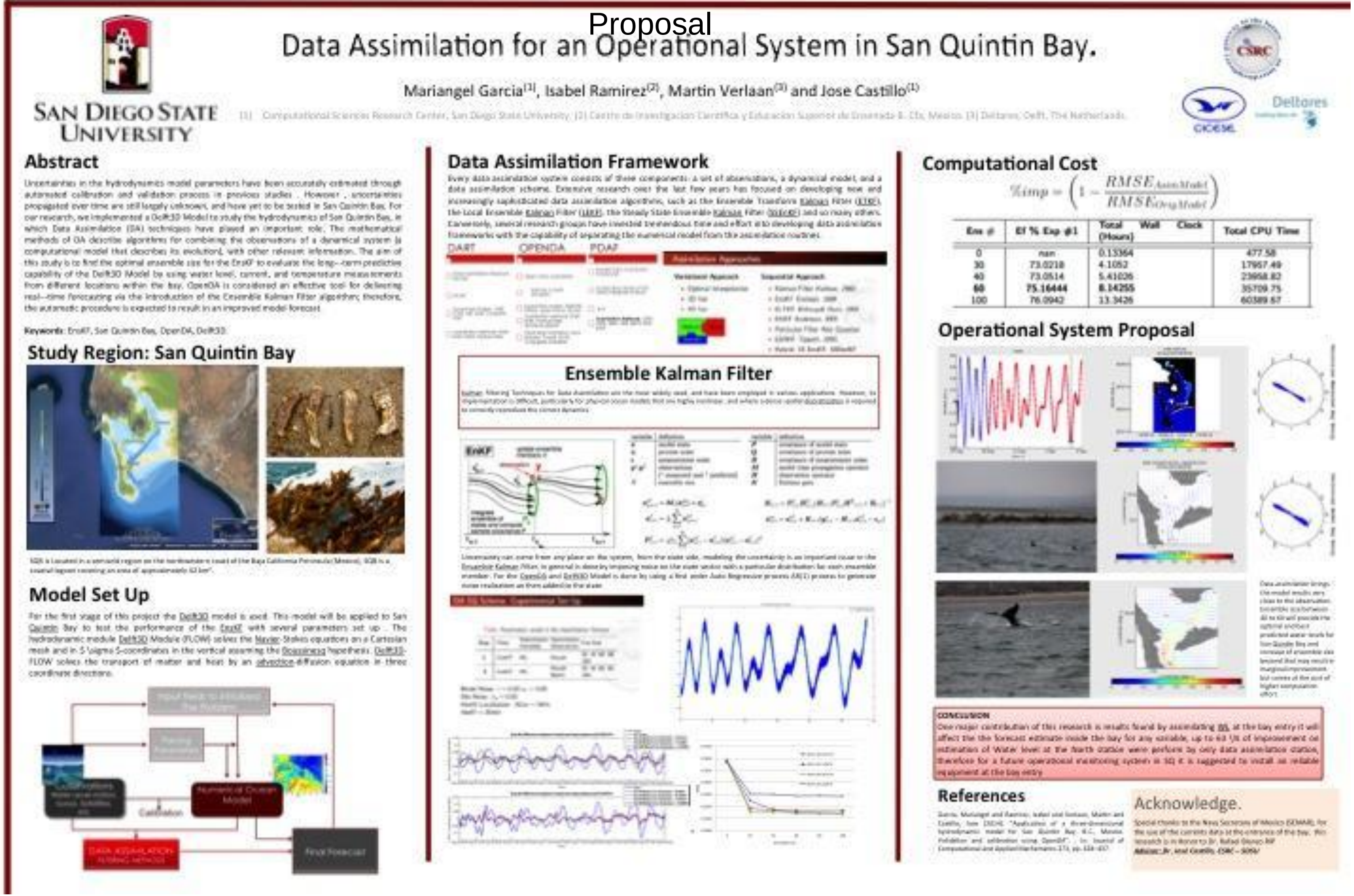
Special thanks to Abdulla Mohammad for the introduction to the Delft3D model. To the Navy Secretary of Mexico (SEMAR), for the use of the currents data at the entrance of the bay and Rafael Blanco for the wonderful images of San Quintin Bay.

This research is supported by the Computational Science Research Center at San Diego State University.

Advisor: Dr. José Castillo, CSRC – SDSU

Contact: mgarcia@sciences.sdsu.edu

Proposal



Data Assimilation for an Operational System in San Quintin Bay.
Mariangel Garcia(1), Isabel Ramirez(2), Martin Verlaan(3) and Jose Castillo(1)
SAN DIEGO STATE UNIVERSITY
CSRC
CICESE
Deltares
Abstract
Keywords: EnKF, San Quintin Bay, OpenDA, Delft3D
Study Region: San Quintin Bay
Model Set Up
Data Assimilation Framework
DART
OPENDA
PDAF
Ensemble Kalman Filter
EnKF
Computational Cost
Ens #
Ef % Exp #1
Total Wall Clock (Hours)
Total CPU Time
0
nan
0.13364
477.58
30
73.0218
4.1052
17957.49
40
73.0514
5.41026
23958.82
60
75.16444
8.14255
35709.75
100
76.0942
13.3426
60389.67
Operational System Proposal
CONCLUSION
References
Acknowledge.

DART

From NCAR

GCCOM_DART Ensemble Data Assimilation Analysis System for the 3D-Non-Hydrsotatic General Curvilinear Coastal Ocean Model (GCCOM)

Mariangel Garcia(1), Tim Hoar(2), Mary Thomas(1), Barbara Bailey(3), Jose Castillo(1)
mgarcia@sciences.sdsu.edu

(1) Computational Sciences Research Center, San Diego State University. (2) National Center for Atmospheric Research, Data Assimilation Research Section (NCAR-DAReS). Department of Mathematics & Statistics, San Diego State University.

Abstract

In this project we demonstrate how data assimilation can be used, with a non-hydrostatic coastal ocean model, to study sub-mesoscale processes and accurately estimate the state variables. Its implementation in non trivial, particularly for physical ocean models, which are highly nonlinear, very sensitive to perturbations and require a dense spatial discretization in order to correctly reproduce the dynamics. The major challenge which we address here, is the high computational cost typically incurred by a high-resolution numerical model with a three-dimensional data assimilation scheme in a complicated stratified system. Having the General Curvilinear Coastal Ocean Model (GCCOM) interfaced with the faster data assimilation framework, the Data Assimilation Research Testbed (DART-NCAR), allowed us to assimilate very high resolution observations (10th of meters) into the system. A perfect model experiment in a very steep seamount test case is presented. This experiment, allow us to explore the proper initial ensemble members for the model, estimate the observation error variance needed to reproduce the dynamics in a turbulent flow experiment and, to analyze the impact of localization in such small processes.

Keywords: GCCOM, Non-hydrostatic, 3D, Curvilinear coordinates, High Resolution, Data Assimilation, EnAKF, DART.

UCOAM System

The San Diego State University (SDSU) General Curvilinear Environmental Modeling (GCEM) Group has been developing the Unified Curvilinear Ocean and Atmospheric Model (UCOAM) with the goal to resolve problems using a high-resolution mesh in curvilinear coordinates in 3 dimensions, with Ocean-Atmospheric Interaction, that allows us to simulate stratified ocean currents over uneven terrains and provides better performance and a faster response. The current UCOAM, solve the 3D Navier Stokes equations within the Boussinesq approximation. The full non-dimensional momentum equation is solve using Large Eddy Simulation (LES) technique and a sub grid scale (SGS) model.

UCOAM

UCOAM: Unified Curvilinear Ocean Atmosphere Model

1. Primitive 3D Navier-Stokes equations using Boussinesq approximation.
2. Nondimensionalization and scaling of the NavierStokes equations.
3. Large Eddie Simulation (LES)
4. Fully written in FORTRAN 90.
5. Uses General Curvilinear Coordinates.
6. Using Fully Non-Hydrostatic Pressure Equation.
7. Using UNESCO Equation of State for density.

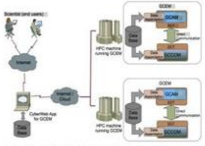


Fig 1. UCAOM General Framework

UCOAM is designed to be an ultra-high resolution model, i.e. UCOAM is capable of performing simulations with spatial resolution as low as 5m. To ease the calculation, all equations are transformed into a uniform curvilinear grid.

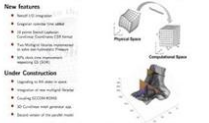
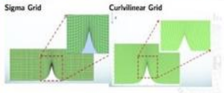


Fig 2. Curvilinear transformation from the physical domain to the computational domain.

Fig. 3 Temporal and spatial scales of ocean processes. Modified after Dickey (2001). Source: http://www.theseusproject.eu/

Sigma Vs Curvilinear



Differences

- Other models make use of sigma coordinate in vertical directions
- The sigma coordinate is not able to handle non convexity in the physical domain.
- Sigma-Coordinate produces artificial forces, particulary in the steep regions

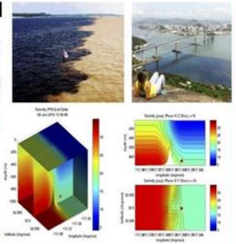
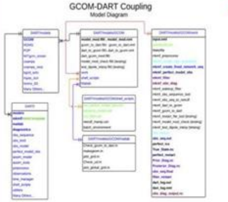


Fig 4. Motivation of some of the sub-mesoscale processes GCCOM model is able to explore. Mass of fresh meeting with the ocean, internal waves, internals bores, etc.

Acknowledgements:
Special thanks to SIAM travel awards program to support this research.
GCCOM is possible thanks to our multidisciplinary collaborators:
Choboter, Paul Coupling, GCEM-ROMS project. - Smirniotis, Colette, DA Algorithms.
Bucciarelli, Randy, GCEM 3DGGrid Generator App. - Abouali Mohammad, GCCOM-NS Model.

DART-GCCOM Ensemble Data Assimilation Analysis System

DART employs a modular programming approach to the application of the Ensemble Kalman Filter, which impels the underlying model into a state that is more consistent with information from a set of observations. It utilizes the Ensemble Adjustment Kalman Filter (Anderson, 2009) by default, but can also execute other filters. Everything is driven by a Fortran namelist, as well as the presence or absence of observations. A Fortran executable named 'filter' reads a namelist, an initial state for the ensemble, and a file containing the observations then begins to work. More detailed information can be found in www.image.ucar.edu



Filter Module

most common namelist settings and features built into DART

- **Ensemble Size:** ensemble sizes between 20 and 100 seem to work best.
- **Localization:** To minimizes spurious correlations and reduce the spatial domain of influence of the observations . Also, for large models it improves run-time performance because only points within the localization radius need to be considered.
- **Inflation:** The spread of the members in a systematic way to avoid problems of filter divergence.
- **Outlier Rejection:** Can be used to avoid bad observations.
- **Sampling Error:** For small ensemble sizes a table of expected statistical error distributions, corrections accounting for these errors are applied during the assimilation.

DART has been compiled using many Fortran 90 compilers, and has run on linuxcompute-servers, linuxclusters, OSX laptops-desktops, SGI Altixclusters, IBM supercomputers (based on both Power and Intel CPUs), and Cray supercomputers. Its structure and prior successful use in global and regioanal ocean models made this particular framework the best candidate for use in the GCCOM.

Model Set Up

A classical seamount under neutral conditions was chosen (i.e homogenous density, buoyancy and Coriolis forces were neglected). In this case, the bottom bathymetry was selected to be very steep, in order to show model to works under a very sensitive (numerically speaking) curvilinear mesh. The domain size is 3.6 km x 2.8 km, where the depths varied between 1km at it deepest and 0.5 km at the origin. The grid size of the domain is 97x32x32 with overall resolution of 30 m x 30 m; however, to better represent the bottom variation, more grid points were chosen in the middle of the domain. A total run of 6 hours in GCCOM requires 21600 iterations. All terms for the non-hydrostatic pressure and all 3 velocities components (u, v, w) are stored every 10 minutes, which yields a single 120MB GCCOM Netcdf output file. For more details of this experiment, refer to the GCCOM user Manual (Abouali and Castillo, 2010).

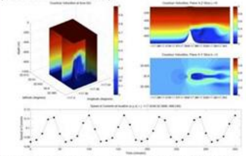


Fig. 5. 3D speed visualization from the True State of the model for the OSSE experiment. Observation control at grid point (nx,ny,nz)=(64,16,10).

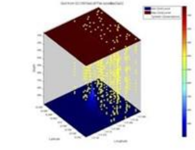


Fig. 6. 50 monitoring observation at 11 random depths, observation are available every 10 minutes for 6 hours.

OSSE Perfect Model Experiment

The primary strategy is to use Observing System Simulation Experiments (OSSEs) to evaluate the impact of new or planned observing systems. Here we introduce the first set of experiments designed to evaluate the new DART-GCCOM framework, the methodology involved, the generation of the initial ensemble, the amount and type of observation assimilated (Fig.6), the ensemble size (30, 60 and 90) and localization parameters (250 m, 1000 m and 2000 m). A sensitivity analysis of observation error variance (1.0-0.1) was performance, with the aim to reproduce the turbulent flow from the true state experiment (Fig. 5).

San Francisco Observing System. (Credit: NOAA)

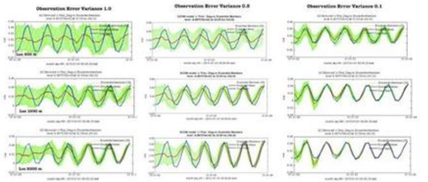
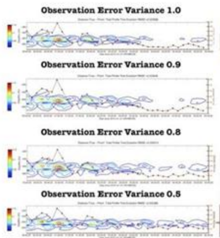


Fig. 7. Ensemble mean (red), true state (blue), ensemble member (green), at control location (nx,ny,nz)=(64,16,10). Observation error variance 1.0, 0.5, 0.1 (first, second, third column respectably) with localization 500m, 1000m, 2000m (row 1, 2 and 3 from top to bottom)



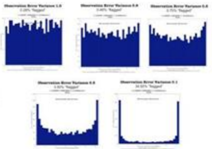


Fig. 8. Last 3 hours, forecast rank histogram. 30 Ensembles members at depth -898.80 m, cutoff 0.00320. Rank histogram with the correct spread observation error variance 1.0, 0.9, 0.8 (top). Rank histogram with small spread, observation error variance 0.5 and 0.1 (bottom).

Fig. 9. Vertical profile time evolution of U_Component at location (nx,ny,nz)=(64,16,[1:32]). Distance =|True – Forecast| observation error variance 1.0, 0.9, 0.8, 0.5 localization 2000 m.

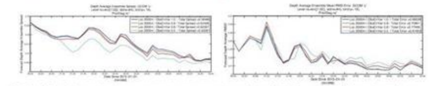


Fig. 10. Left figure: Forecast depth average root mean square error time evolution for U_Component. Tigth figure: Time evolution depth average ensemble spread.

	6 hours simulation / 10 minutes Assimilation				
nodes:processes	Ensemble Size	Wall-clock time (hours)	Output Size	Total RMSE	Total Spread
2:15	30	6.40	3.93 GB x 2	0.69536	0.56466
4:15	60	7.40	7.38 GB x 2	0.68237	0.56336
6:15	90	8.10	10.84 GB x 2	0.6779	0.56255

Table 1. Computational cost associated to the increase of ensemble member from 30 to 90. Localization 2000 m, observation error 1.0. Total spread and rmse are computed from the priori at control profile.

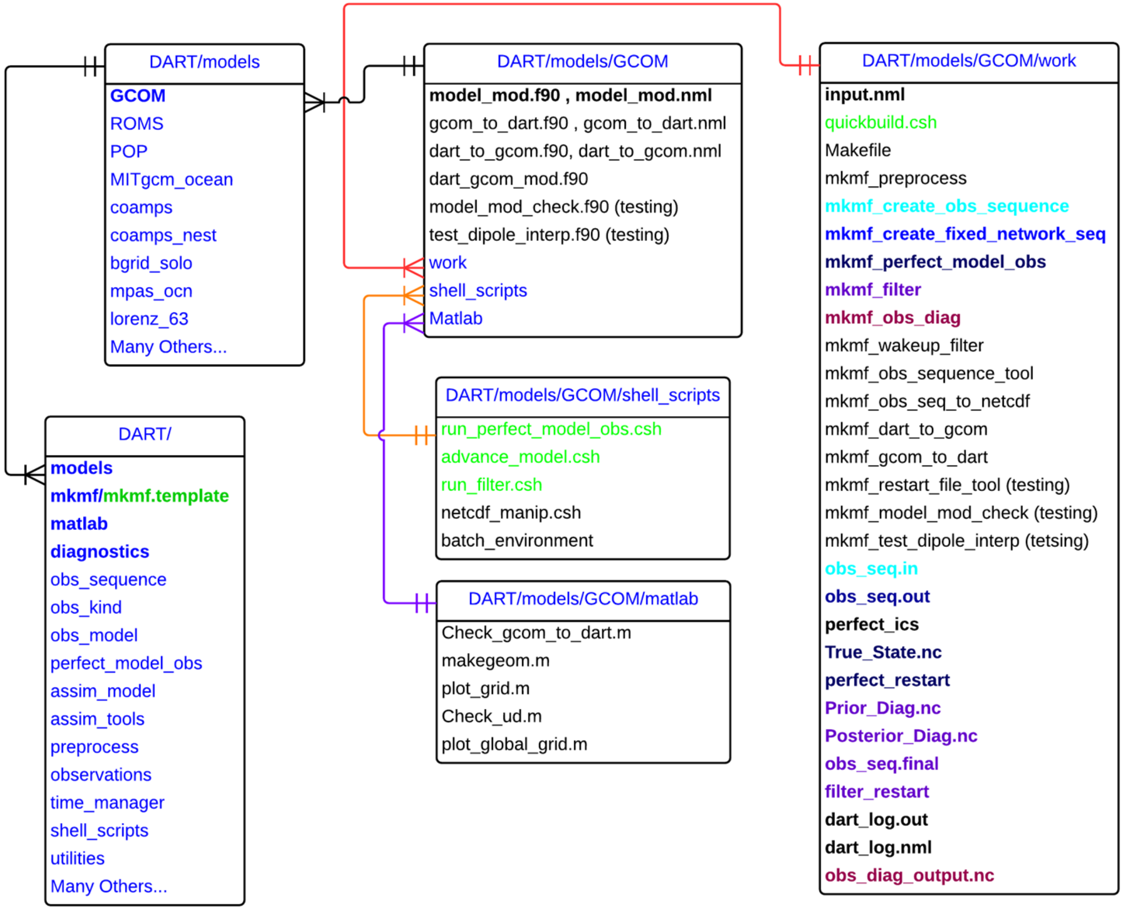
Conclusions

The ensemble adjustment Kalman filter (EAKF) successfully assimilate very high resolution data in small time window of 10 minutes. With only 30 ensemble members exhibits relatively high forecast skill. Increasing the ensemble size from 30 to 100 was not crucial for the current prediction. We found for small domain of couple of kilometers, every observation impact every state variable, however the spread of ensembles tend to reduce over time. The assimilation system although exhibited some sensitivity to observation error variance, in general it can handle large observation error variance from 0.8-1.0. All these results suggest that this ensemble-based system is able to extract the dynamically important information from the model to provide reliable statistics to map the information from the observations into the model space and generate a good initialization for the ensemble forecast.

References
Anderson, J., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Arellano, 2009. The Data Assimilation Research Testbed: A Community Facility. *Bulletin of the American Meteorological Society:* **90** (9), 1283-1296.
M Abouali, JE Castillo 2013. Unified curvilinear ocean atmosphere model (ucoam): A vertical velocity case study. Mathematical and Computer Modelling 57 (9), 2158-2168.

GCOM-DART Coupling

Model Diagram




```
[mgarcia@node10 models]$ pwd
/home/mgarcia/UCOAM-DART/UCOAM/models
[mgarcia@node10 models]$ ls
9var         dynamo            lorenz_04          mpas_atm     POP
am2          ECHAM             lorenz_63          mpas_ocn     ROMS
bgrid_solo   forced_barot      lorenz_84          NAAPS        rose
cam          forced_lorenz_96  lorenz_96          NCOMMAS      simple_advection
CESM         GCOM              lorenz_96_2scale   noah         sqg
clm          gitm              MITgcm_annulus     null_model   template
coamps       ikeda             MITgcm_ocean       PBL_1d       tiegcm
coamps_nest  LMDZ              model_mod.html     pe2lyr       wrf
[mgarcia@node10 models]$
```

Filter Module

most common namelist settings and features built into DART

- **Ensemble Size:** ensemble sizes between 20 and 100 seem to work best.
- **Localization:** To minimizes spurious correlations and reduce the spatial domain of influence of the observations . Also, for large models it improves run-time performance because only points within the localization radius need to be considered.
- **Inflation:** Modify the spread of the members in a systematic way to avoid problems of filter divergence.
- **Outlier Rejection:** Can be used to avoid bad observations.
- **Sampling Error:** For small ensemble sizes a table of expected statistical error distributions, corrections accounting for these errors are applied during the assimilation.

I can go on forever....

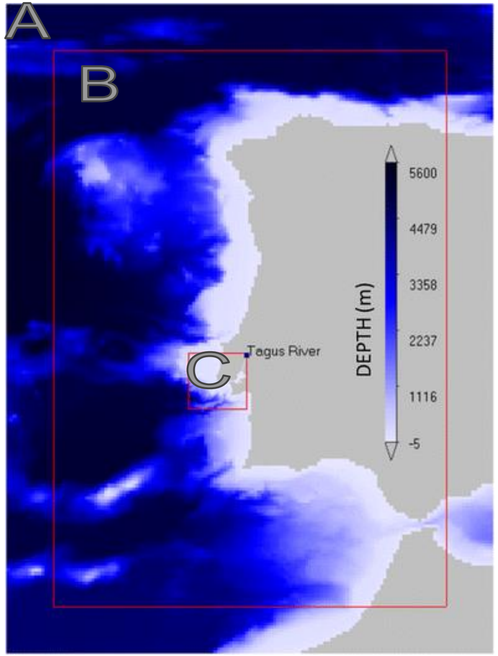
How to proceed?

01	What forecasting system we want to improve?	• Define study region • Define analysis season • Identify the challenges
02	What type of observation we are going to assimilate?	• Find spatial and temporal scales • UQ for observation errors • Model Validation
03	Define the DA Methodology for MOHID.	• Define Interpolation Methods • Initial Ensemble Member • Nr. of Ensemble • Assimilation window • OSSE

What forecasting system we want to improve?

PCOMS

Operational since 2011



Campuzano F, Brito D, Juliano M, Fernandes R, de Pablo H, Neves R. Coupling watersheds, estuaries and regional ocean through numerical modelling for Western Iberia: a novel methodology. Ocean Dynamics. 2016; 66(12): 1745–1756. DOI: 10.1007/s10236-016-1005-4.

DOMAIN	(A) West Iberia (2D)	(B) Portugal (3D)+ the Iberian Atlantic coast	(C) Tagus Estuary 3D
Grid Corner	(33.48 °N, 45.90 °N) (4.20 °W, 13.50 °W)	34.38 °N, 45.00 °N 5.10 °W, 12.60 °W	38.16 °N, 39.21 °N 38.5-39.1 °N
Horizontal Dim	207 x 155	177 x 125	120x145
Vertical Dim	----	7 Sigma Layer (0- 8.68) 43 Cartesian layers	7 Sigma Layer (0-8.68) 43 Cartesian layers
Delta x	0.06° (≈ 5.2 km)	0.06° (≈ 5.2 km)	≈ 2 km off the coast up to 250 m
Delta t			
Tides	FES2004 (Lyard et al., 2006)	From (A)	From B
Atmosphere	WRF (Skamarock et al., 2005) 12 km ???? http://www.meteogalicia.es/	MM5 (Grell et al., 1994) 9km provided by IST	WRF (Trancoso, 2012) 3 km http://meteo.tecnico.ulisboa.pt/
Discharge	---	---	Almourol
Modules		Baroclinic hydrodynamic, ecological	Baroclinic hydrodynamic, ecological

A Validated Model

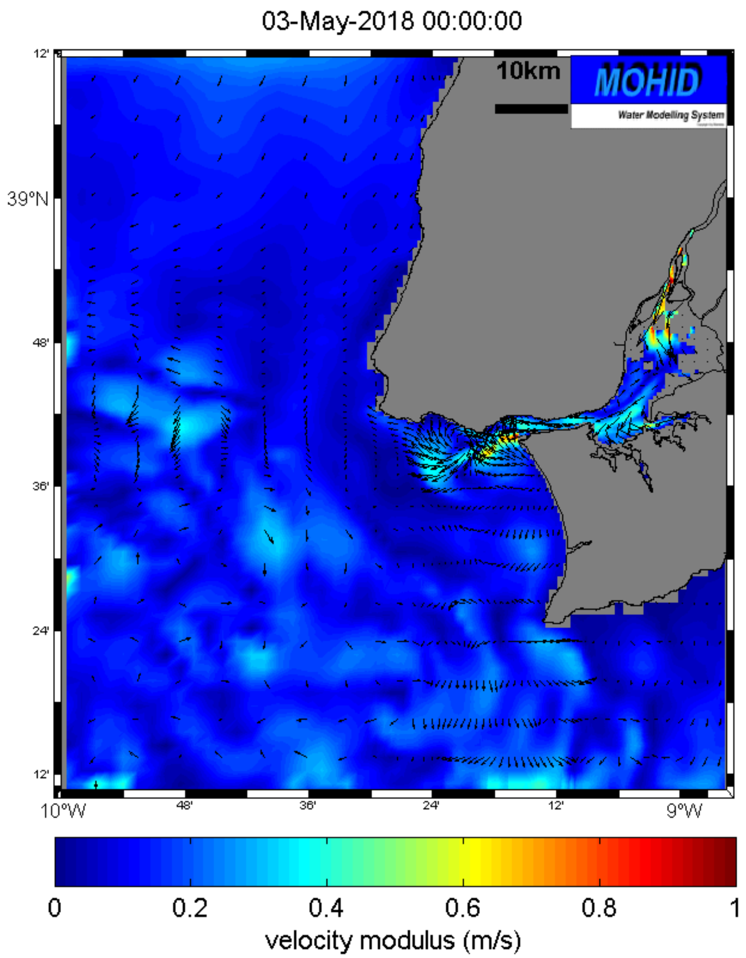
Operational since 2011

- Data available every 3 hours
- Salinity, Temperature, Velocities, SSH, Dissolved Oxygen, Chlorophyll-a, Zooplankton, Nitrate, Phosphate,suspended particulate Matter.

Campuzano, et al. (2016). Coupling watersheds, estuaries and regional ocean through numerical modelling for Western Iberia: a novel Methodology . Ocean Dynamics. 2016; 66(12): 1745–1756. DOI: 10.1007/s10236-016-1005-4.

Rodrigues, J. (2015). The Tagus estuarine plume variability: impact in coastal circulation and hydrography. Master Thesis, Universidade de Aveiro.

Add more references here



Mohid data source: : http://forecast.maretec.org/maps_tagusmouth.asp

HF Radar
&
Mohid Tagus Model

Exploratory Analysis

by Mariangel Garcia;

Francisco Campuzano; Paulo Chambel Leitão and Ramiro Neves

Motivation

Data Assimilation Project

Develop a data assimilation module for MOHID to improve forecast capabilities along Tagus River

Interreg Atlantic Area
European Regional Development Fund
EUROPEAN UNION
MyCoast

GOAL

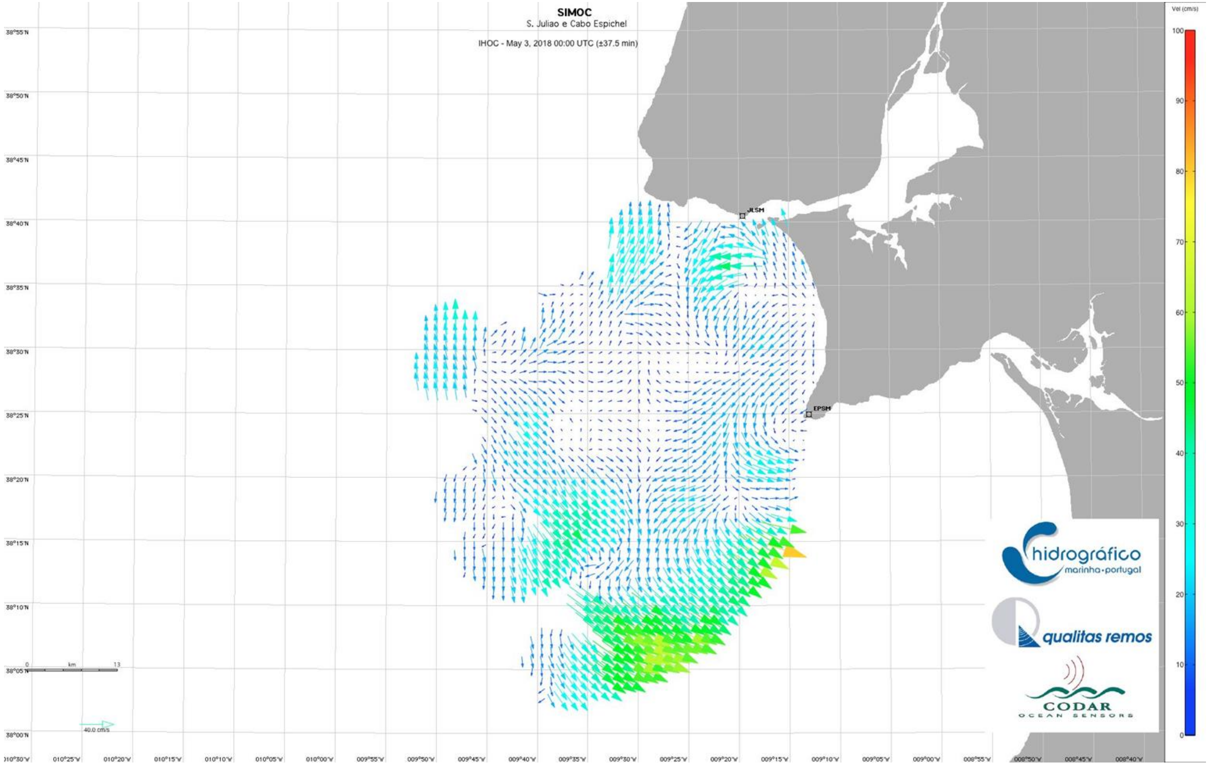
1. Explore the quality of the real time data available along Tagus River
2. Validate Mohid Tagus Model with HF Radar data
3. Determine the most accurate region to interpolate HF Radar Data into Mohid Model
4. Propose the methodology to assimilate observations into Mohid Model

Exploring HF

17 days exploratory Analysis
May, 01-17, 2018

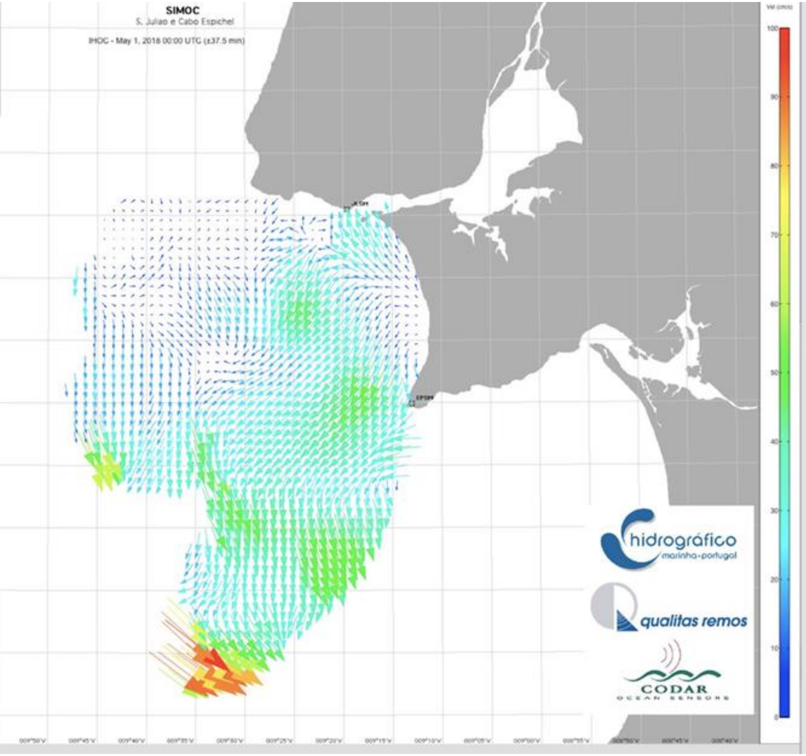
HF data

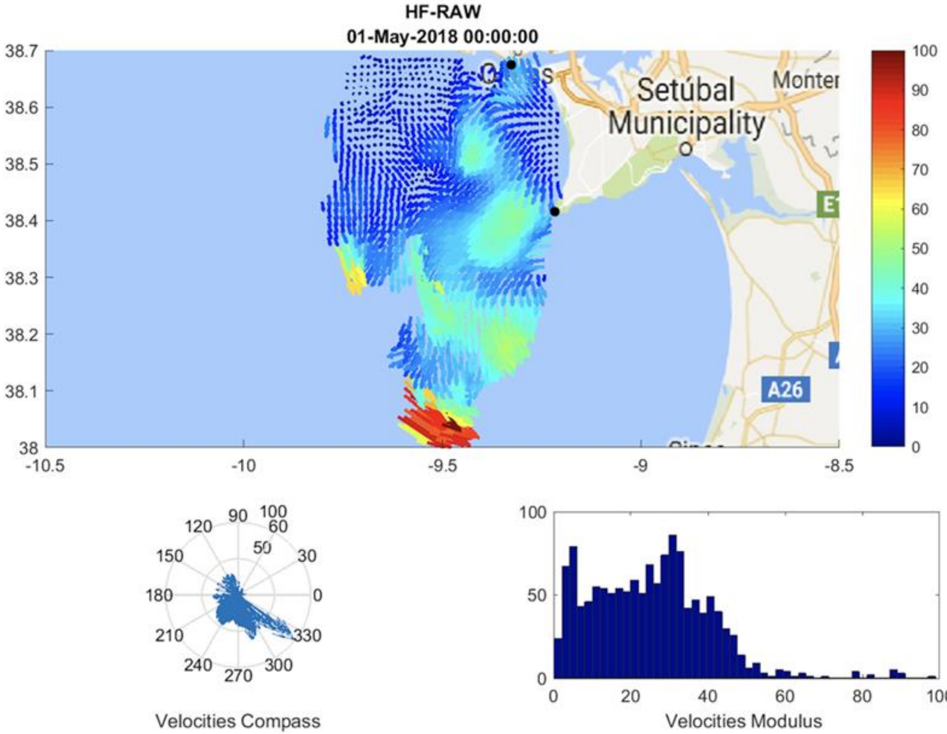
- Resource: Hidrografico
- Grid Spacing: ~1.4 Km
- Frequency: every hour
- Format .tuv (ASCII file)
- The output is already pre-processed by SeaDisplay 6.7.8
- Averaging Radius: 4.000 km
- DistanceAngularLimit: 20.0
- CurrentVelocityLimit: 100.0 cm/s



HF Data source:http://www.hidrografico.pt/simoc.php

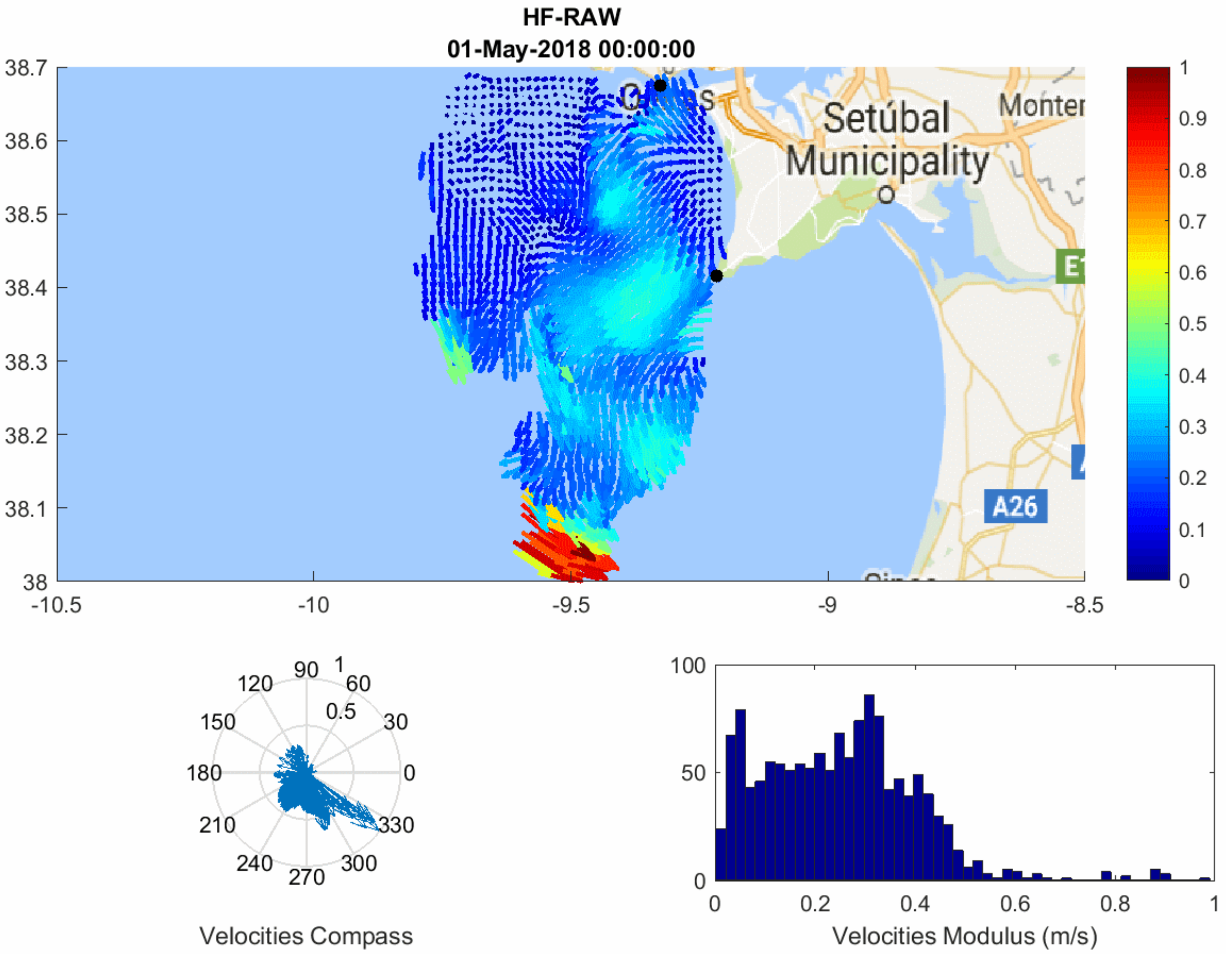
Sanity Check



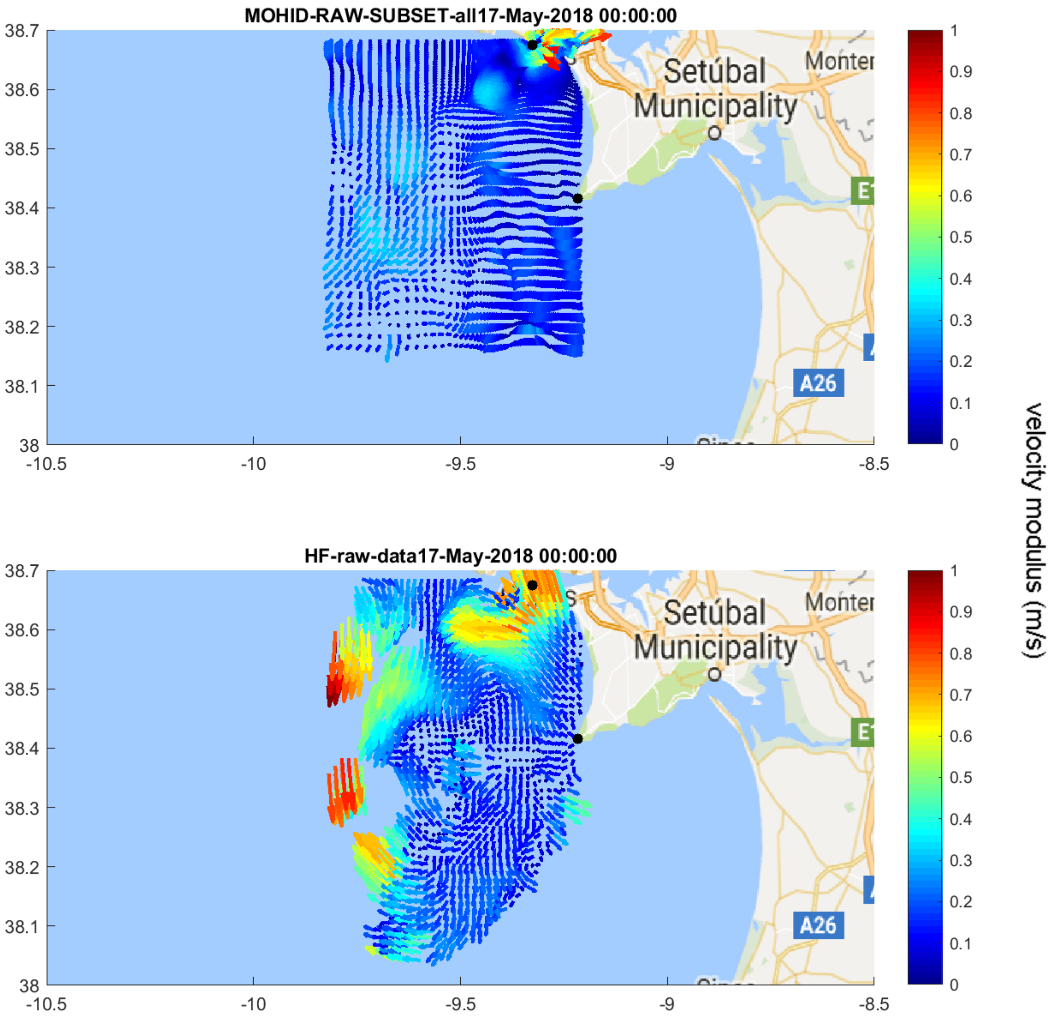


HF Quick Overview

- Mesh changes over time
- Higher velocities at boundaries



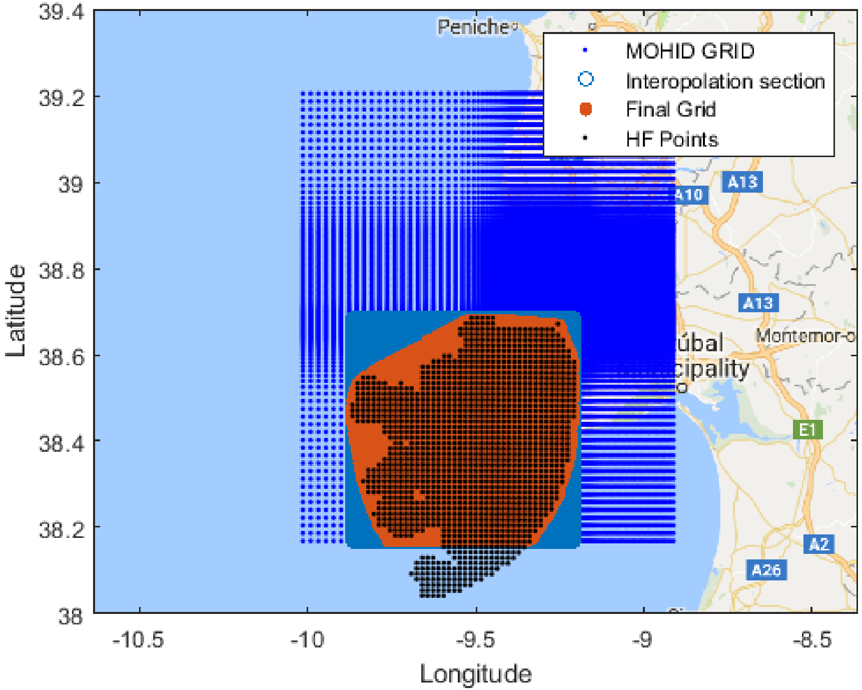
Mohid and HF first look at the data



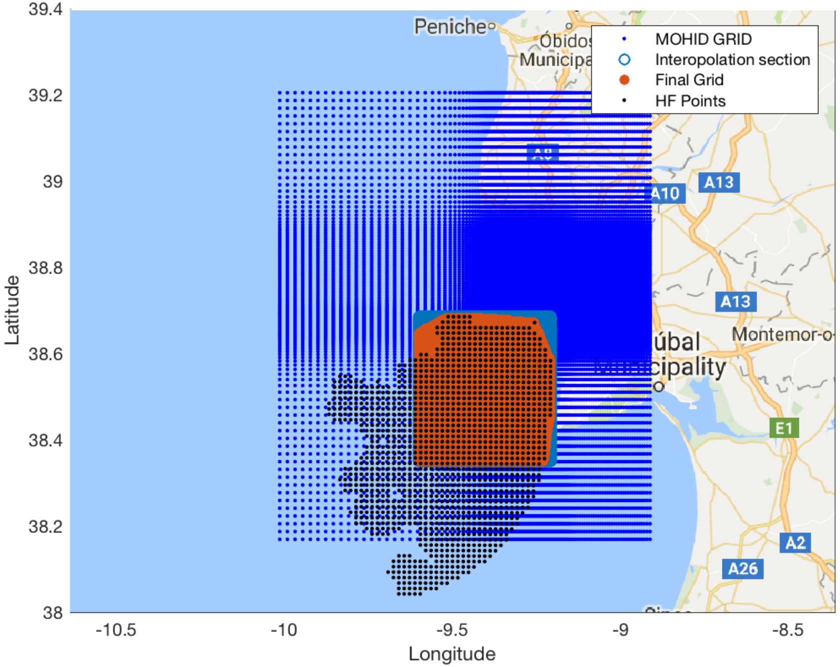
Grid Interpolation

From Observation to Grid

Grid Version 1



Grid Version 2



Grid Version 2:

145x120

xmin=-9.6 ymin=38.35;

xmax=-9.2079 ymax=38.6855

Interpolation Method:

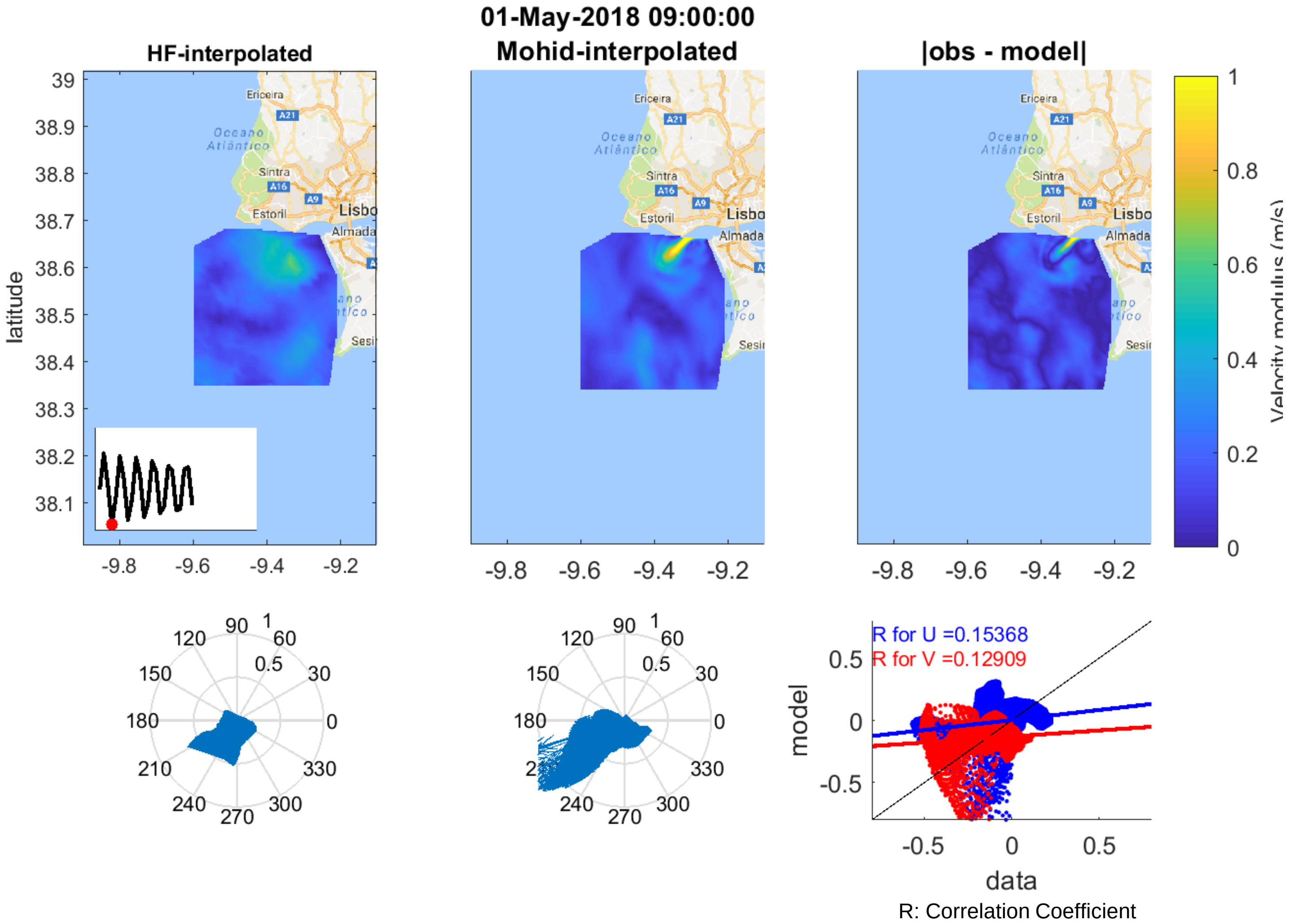
uses a Delaunay triangulation of the scattered sample points to perform interpolation.

[1] Amidror, Isaac. “Scattered data interpolation methods for electronic imaging systems: a survey.” *Journal of Electronic Imaging*. Vol. 11, No. 2, April 2002, pp. 157–176.

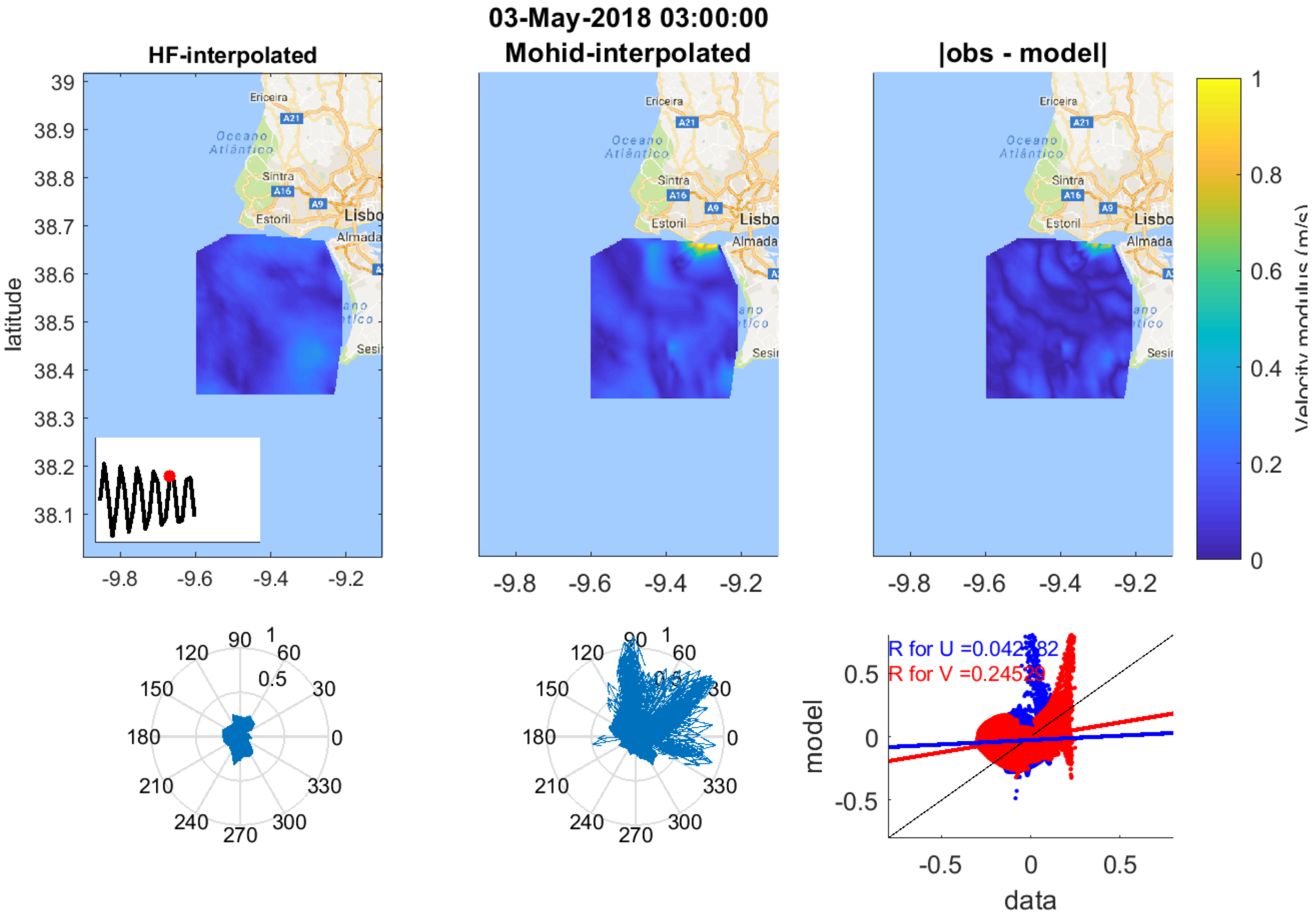
Validation with HF

17 days exploratory Analysis
May, 01-17, 2018

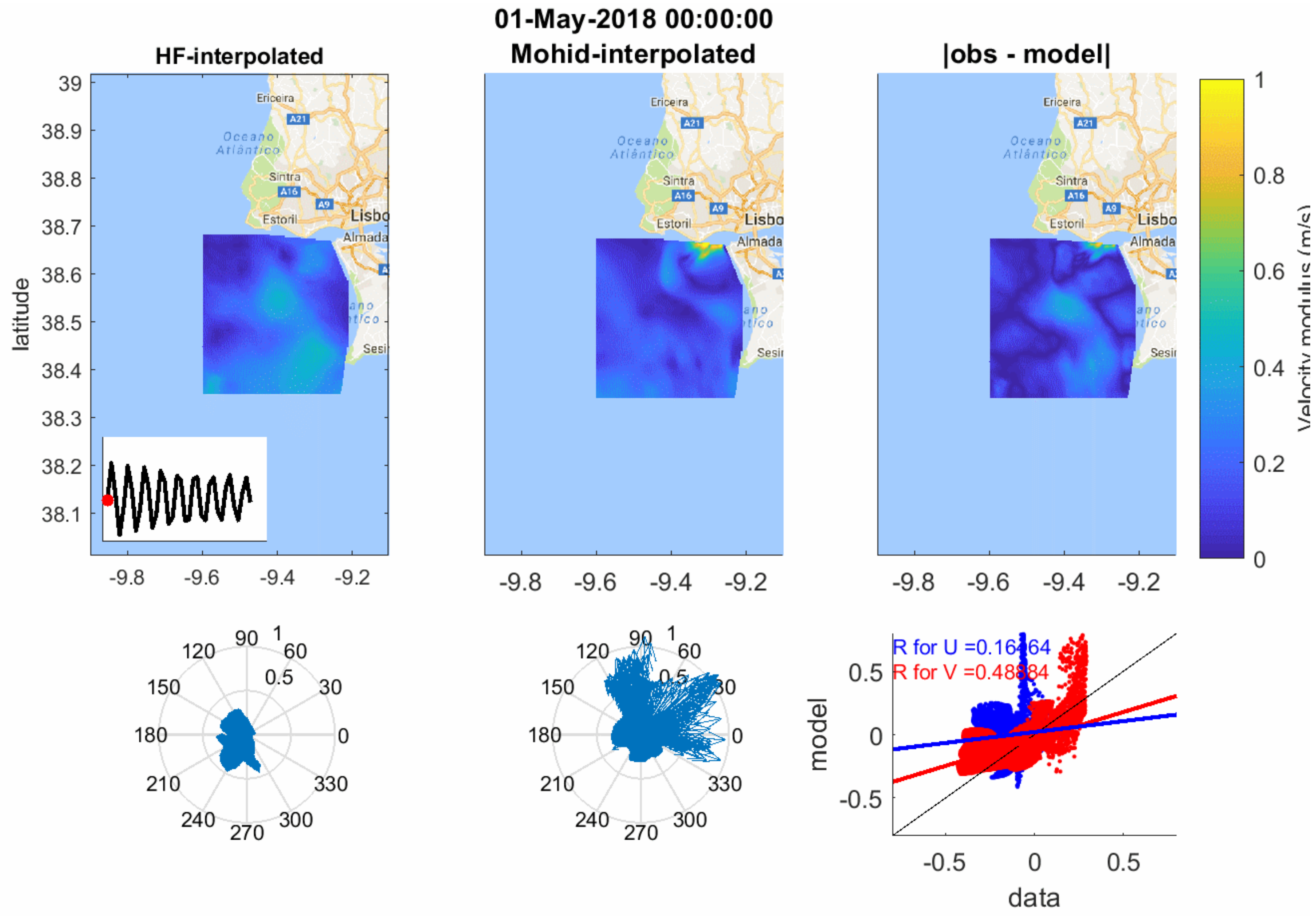
HF Vs Model during Low Tide



HF Vs Model during High Tide

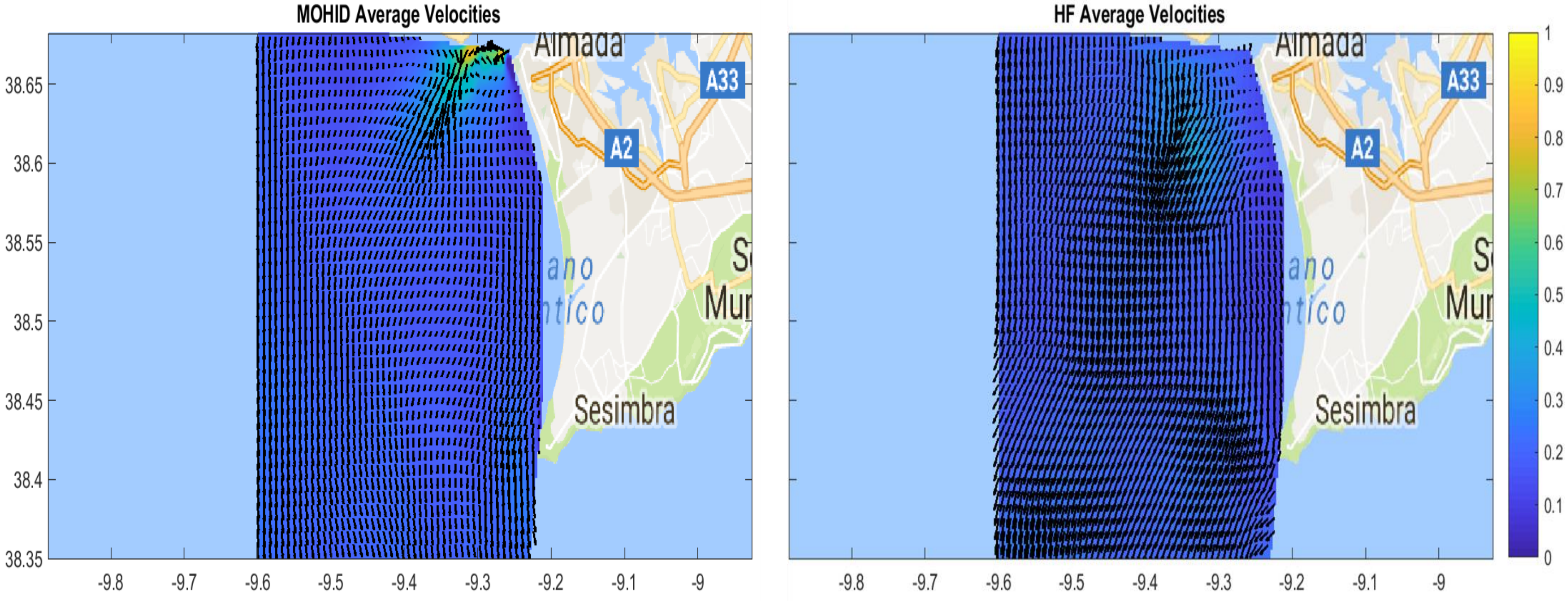


R: Correlation Coefficient

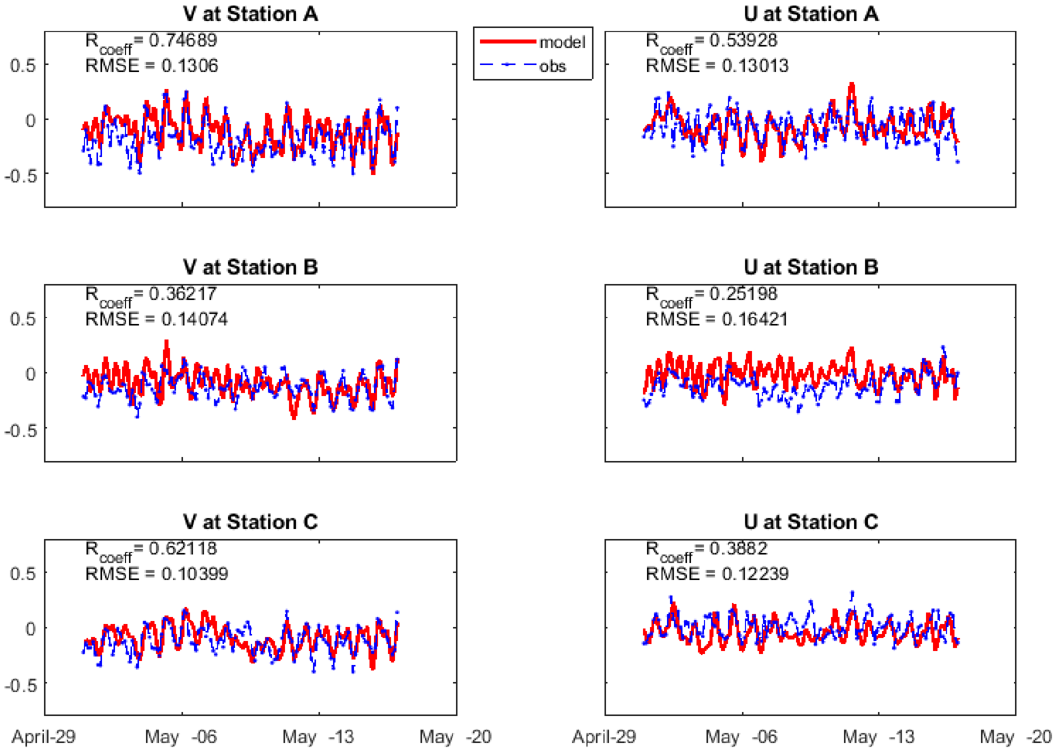


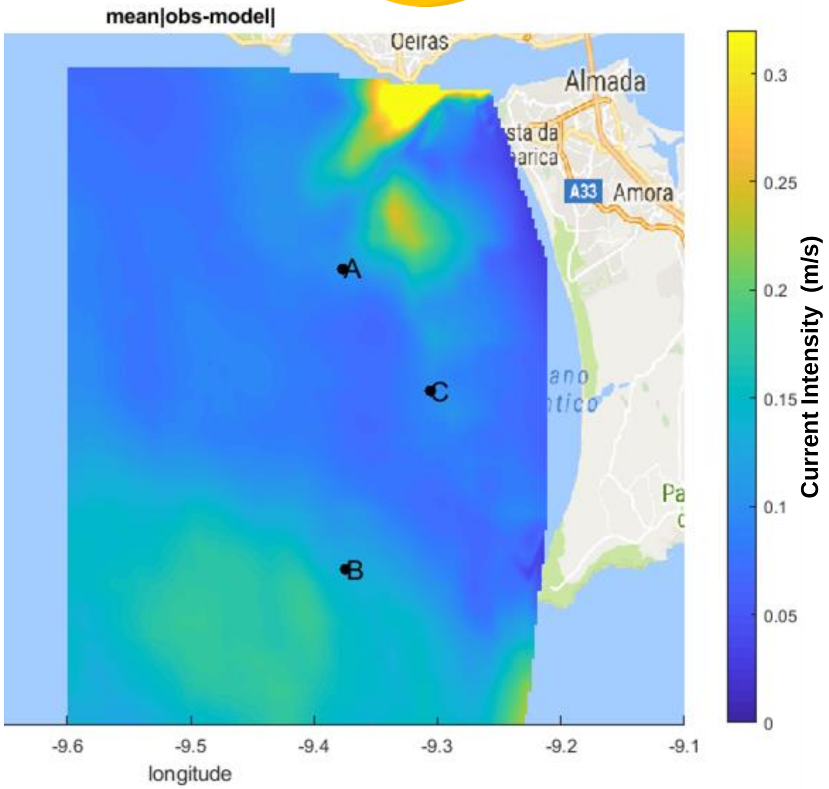
R: Correlation Coefficient

Average velocities over time (17 days)

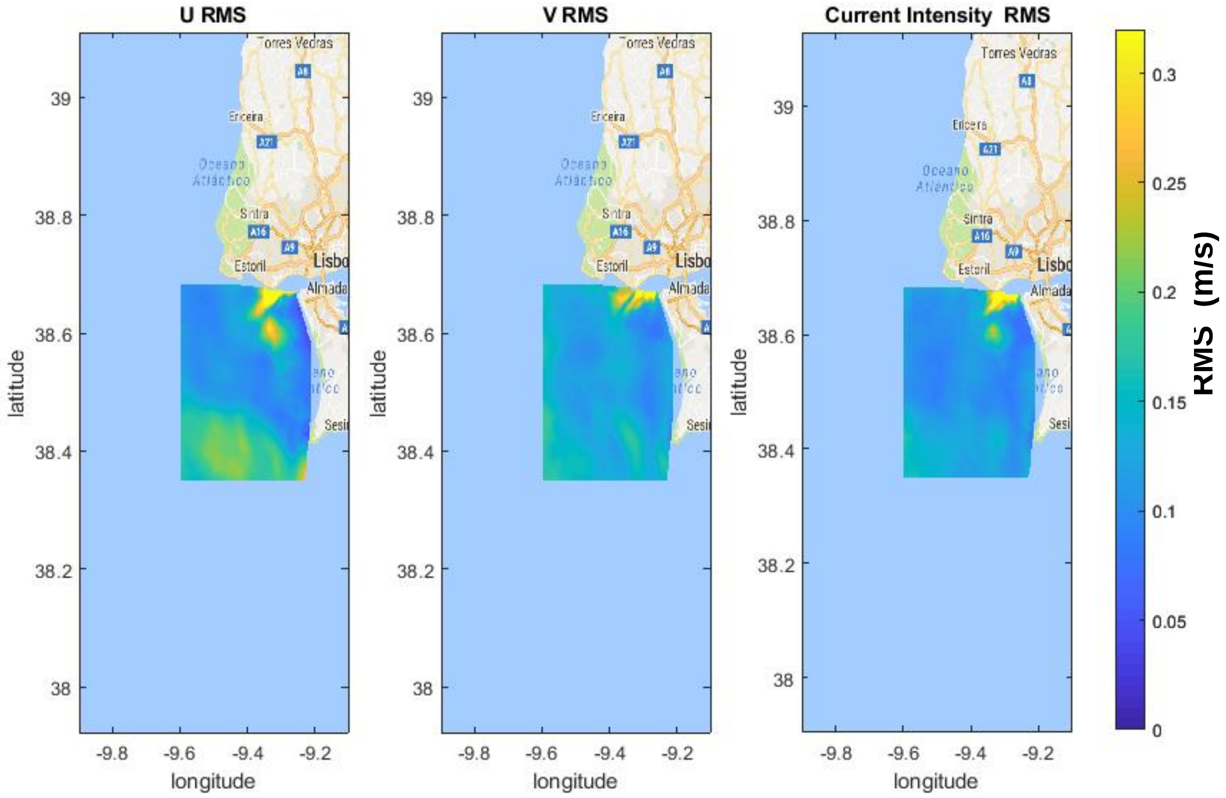


17 days Time Series analysis
May 2018

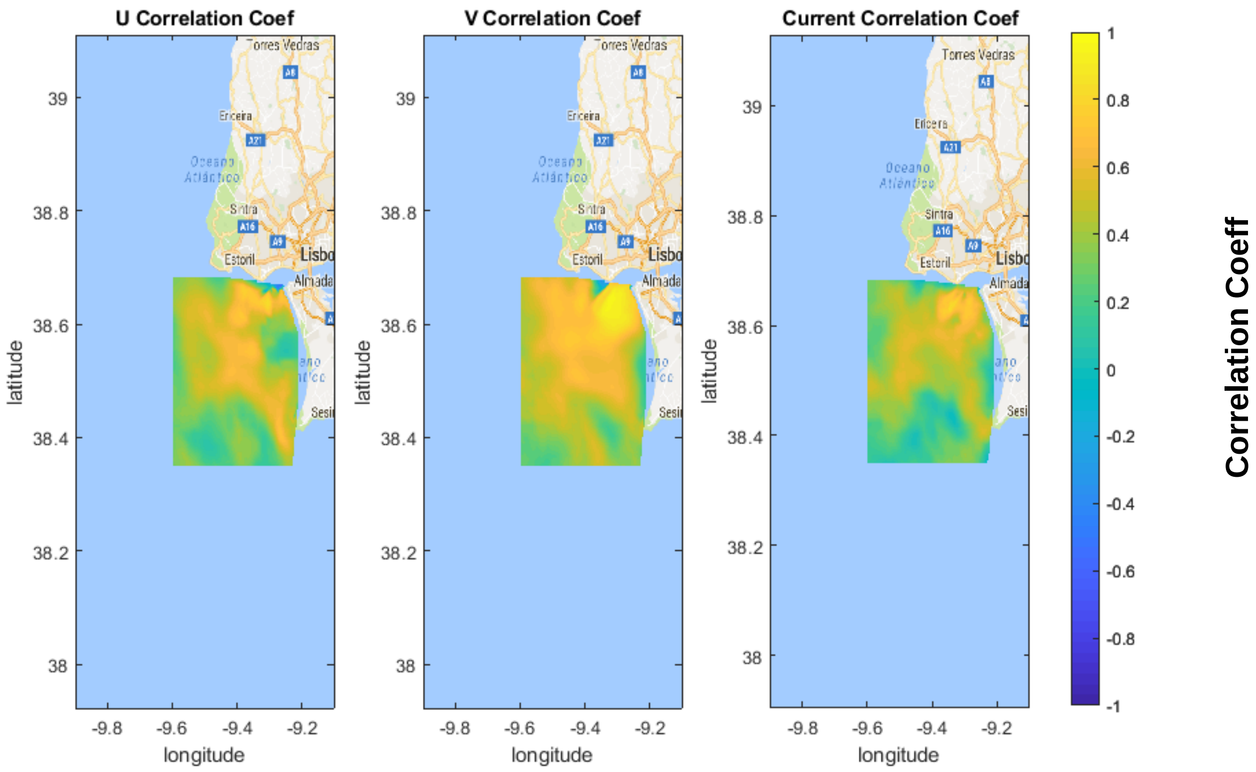




Time Average RMS for the 17 days



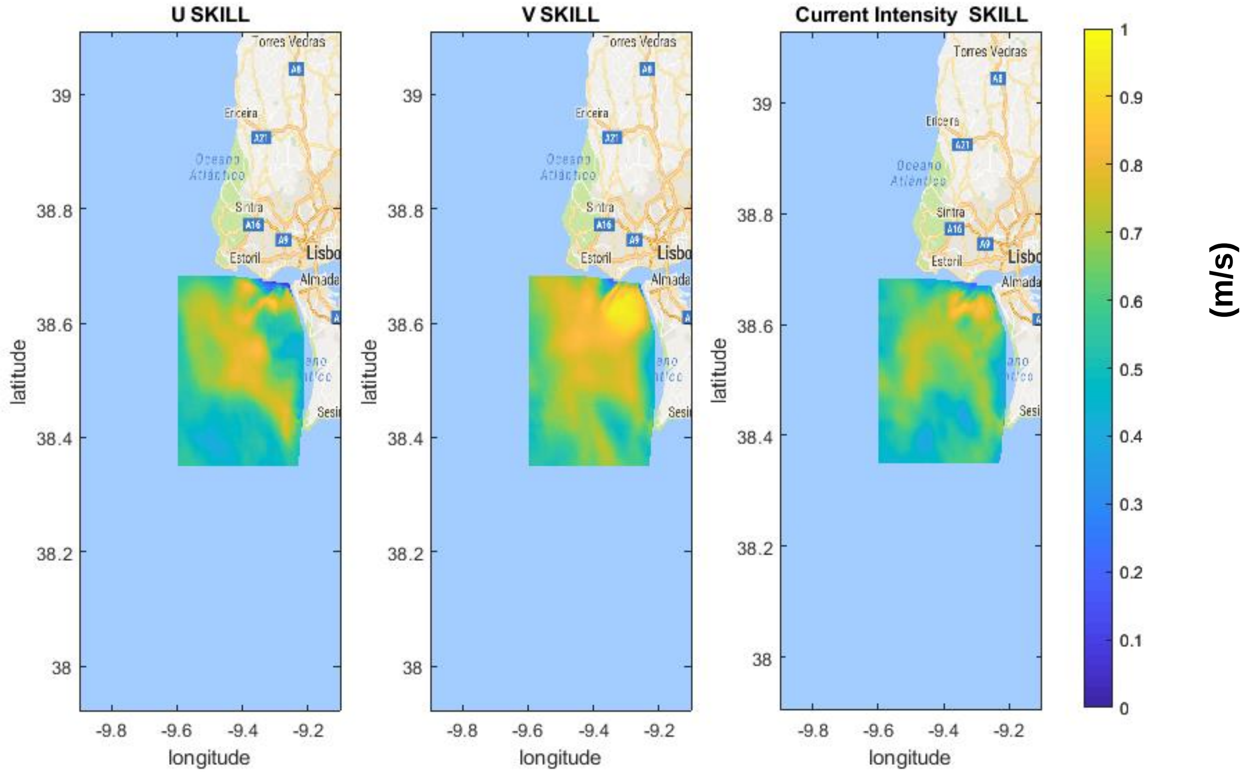
Time Average correlation Coeff for the 17 days



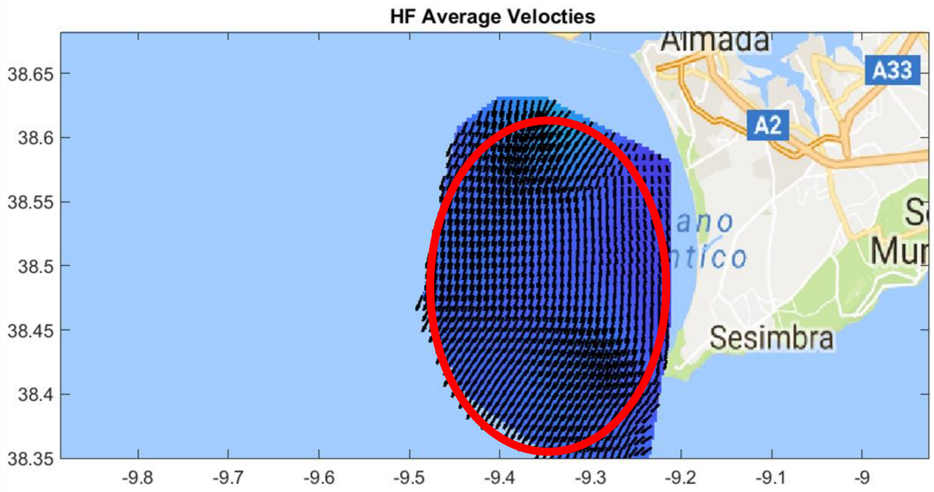
Skill of the model

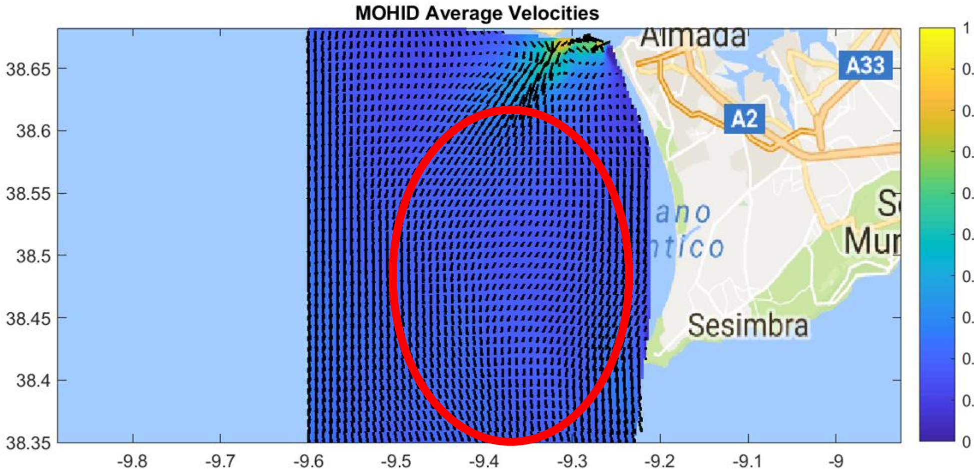
$$SKILL = 1 - \frac{\sum |X_{model} - X_{obs}|^2}{\sum \left(\left| X_{model} - \bar{X}_{obs} \right| + \left| X_{obs} - \bar{X}_{obs} \right| \right)^2}$$

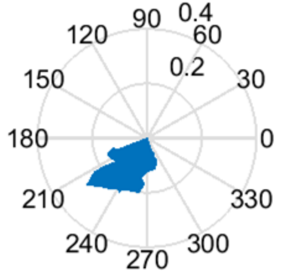
Rodrigues, J. (2015).

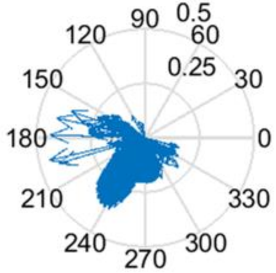


Preliminary region "estimate" to Assimilate HF Radar into MOHID Model









NEXT

For HF Validation and Data

1. Identify sources of weather data.
2. Quick validation of WRF model with one station.
3. Find correlation between wind and currents
4. Find other sources of observations to validate velocities.
5. Propose the best region to assimilate the HF Radar data
6. Explore other sources of observations for Tagus

NEXT

For Data Assimilation

1. Decide the DA approach for MOHID.
2. Find a cluster to do MOHID testing in Linux.
3. Learn Restarting on MOHID
4. Run several perturbed instance of MOHID in parallel.

1. CODE LIKE CRAZY!!

Thank You!

I LOVE
PORTUGAL

Questions?

Weather around Tagus

”The influence of the wind stress on the surface of the ocean has a large impact on surface current variability.

This is quite true in estuarine plume propagation studies.

The adjacent coast of the Tagus Estuary is highly influenced by the wind variability.”

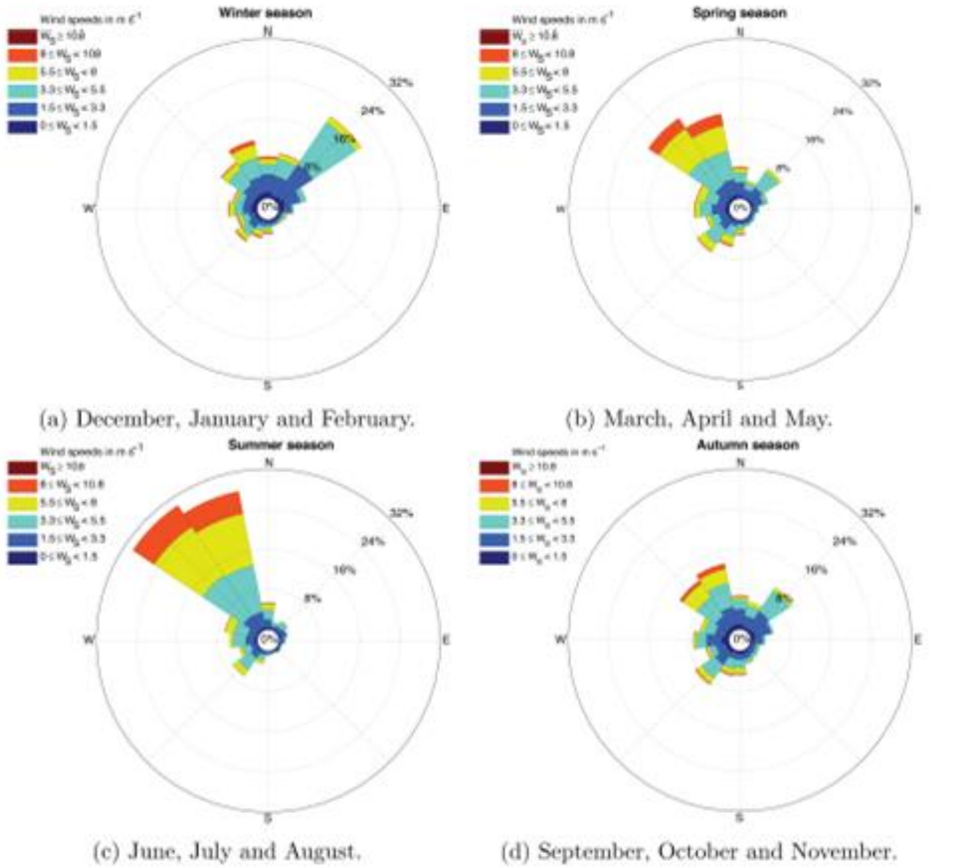


Figure 3.5: Wind roses calculated for (a) winter, (b) spring, (c) summer and (d) autumn, from meteorological airport station data, for the period between January 2001 and December

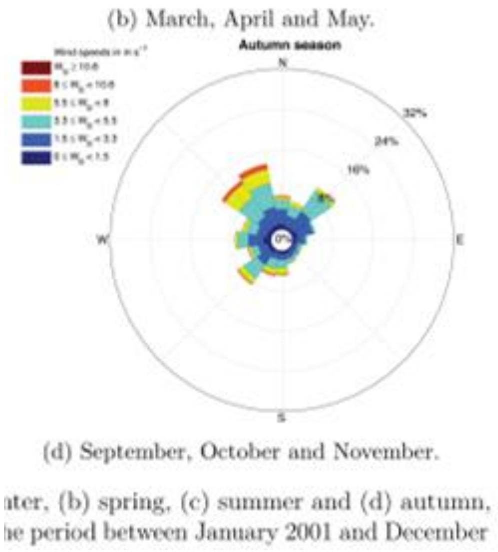
Ref: Rodrigues, J. (2015).

Weather around Tagus

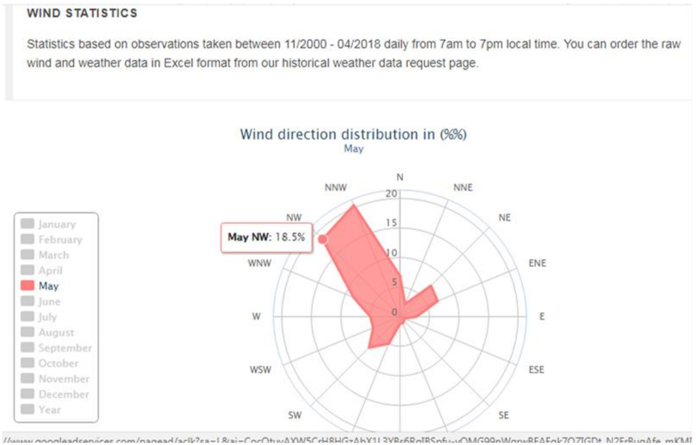
"The influence of the wind stress on the surface of the ocean has a large impact on surface current variability.

This is quite true in estuarine plume propagation studies.

The adjacent coast of the Tagus Estuary is highly influenced by the wind variability."



Ref: Rodrigues, J. (2015).



https://www.windfinder.com/windstatistics/lisboa

WRF3 Validation

EMG: Estação Meteorológica da Guia

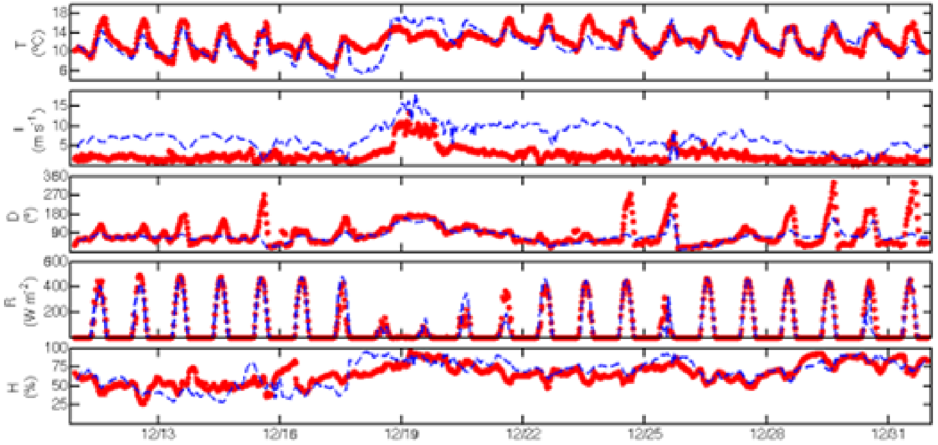


Figure 3.4: Variability of WRF3 data and EMG station data for temperature (Temp), wind intensity (Int) and direction (Dir), humidity (Hum) and radiation (Rad) for 10-31th of December 2007. Red line represents EMG station data and the the blue line the WRF3 predictions.

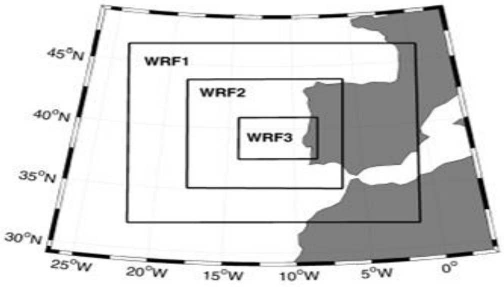


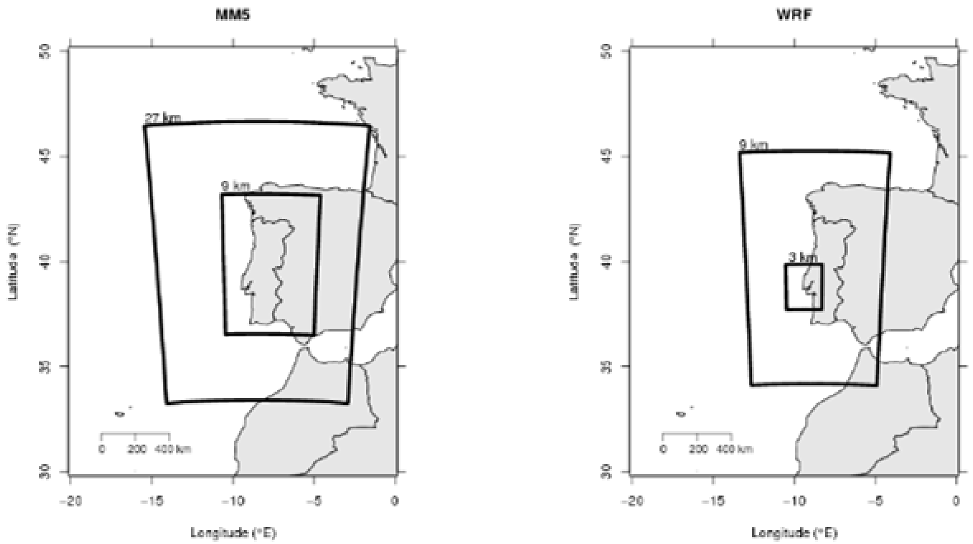
Figure 3.3: Nested domains used for atmospheric simulations.

Ref: Rodrigues, J. (2015).

WRF Model IST Operational since 2007

Table 1.1: *Options used in each model and domain of the IST operational system. GFS - Global Forecast System resolution used for boundary and initial conditions in the outer domain; LSM - Land surface model; PBL - planetary boundary layer. 27 vertical layers were used in all domains.*

Model	GFS	Resol. (km)	NYxNX	Micro-physics	Radiation	LSM	PBL	Cumulus
MM5	1º	81	40x50	Reisner1	Cloud	5-layer soil	MRF	Grell
3.7	-	27	55x40	Reisner2	"	"	"	"
	-	9	82x55	"	"	"	"	"
MM5	0.5º	27	55x40	Reisner1	"	"	"	"
3.7	-	9	82x55	Reisner2	"	"	"	"
WRF	0.5º	9	135x80	3-class	RRTM	Noah	Yonsei Univ.	Kain-Fritsch
3.2	-	3	79x64	"	"	"	"	"



(a) MM5 (b) WRF

Figure 1.3: *Domains in IST 3-day and 7-day forecast system for Portugal, since 2007*

Ref: PhD Thesis, Trancoso 2012

Thanks Marcos Mateus!!!

Source of Meteorological Data

- SNIRH
- IST
- IPMA

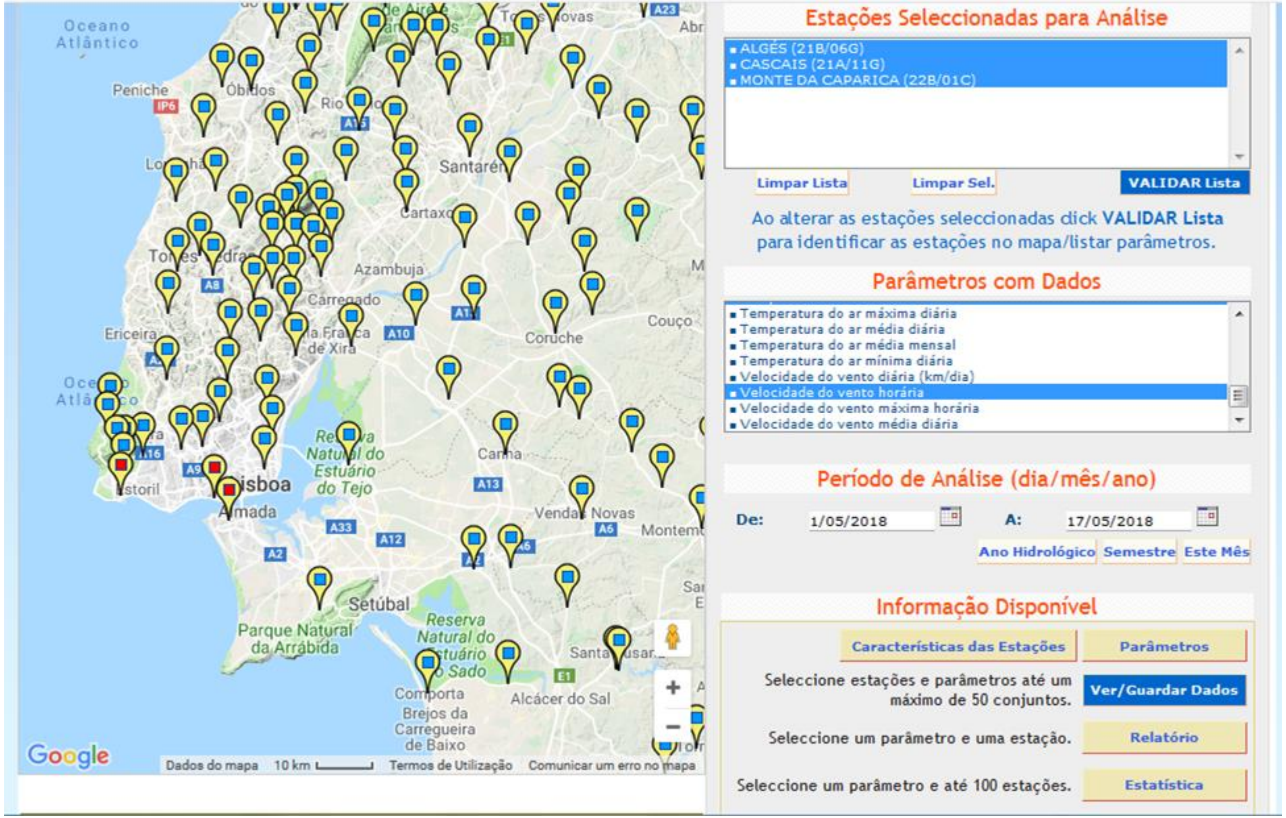
Selected stations:

★ Costa Caparica

★ IST

★ Aeroporto (where??)

★ EMG: Estação Meteorológica da Guia (No data?)



ttps://snirh.apambiente.pt/index.php?idMain=2&idItem=1&objCover=920123704&objSite=920685506

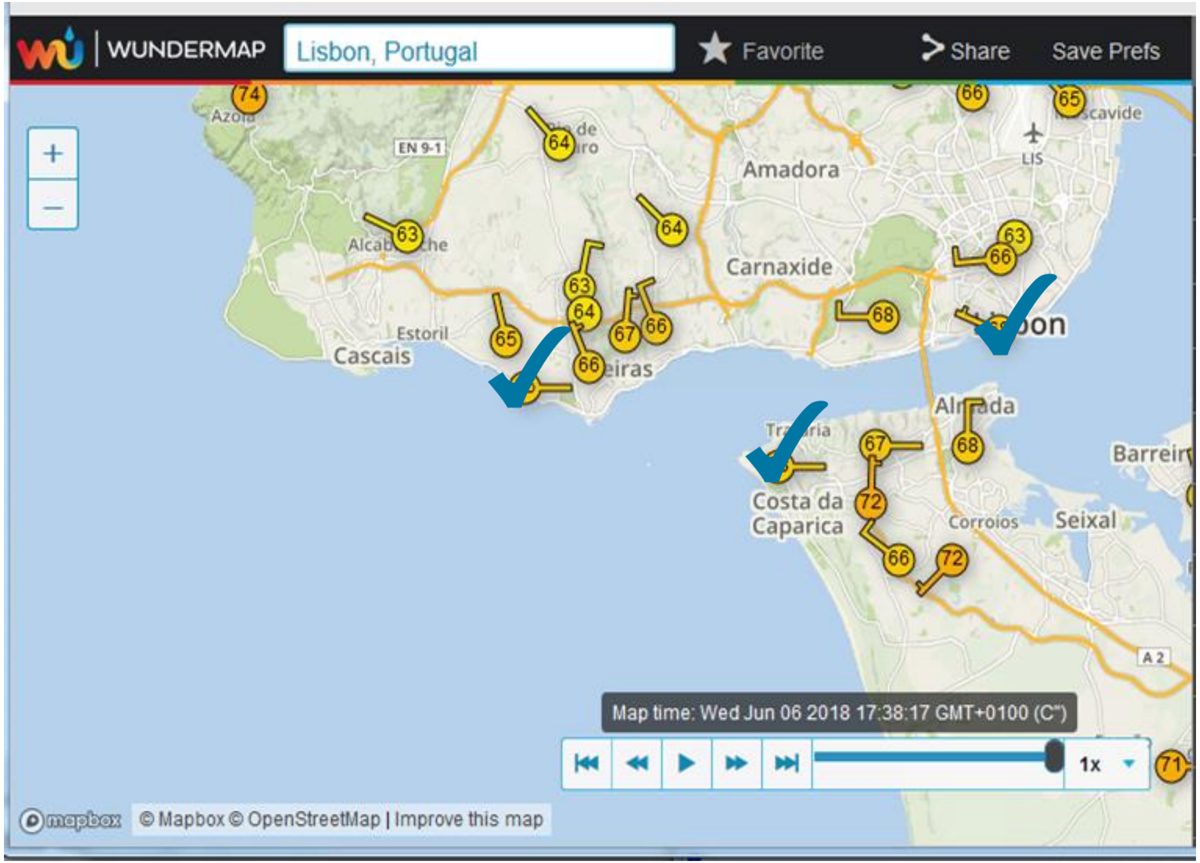
Source of Meteorological Data

Thanks Jorge Palma!!!

Selected stations:

Daily data only

- ★ Cais do Sodre
- ★ Praia Parede
- ★ Caparica S.Joao da Caparica



https://english.wunderground.com/wundermap

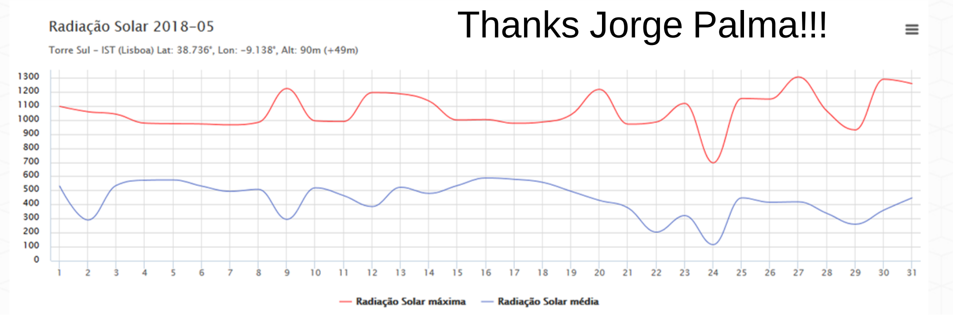
Thanks Jorge Palma!!!

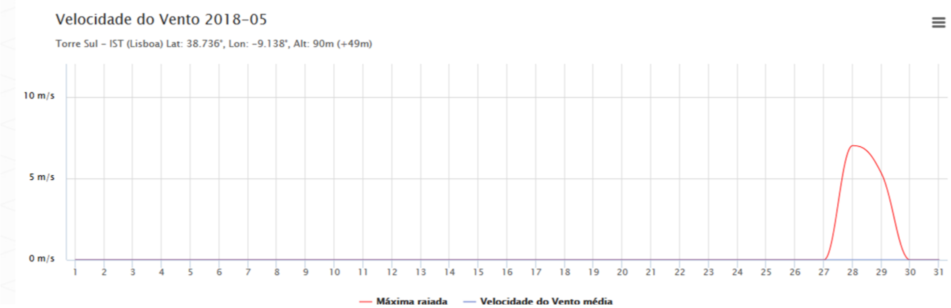
IST Station

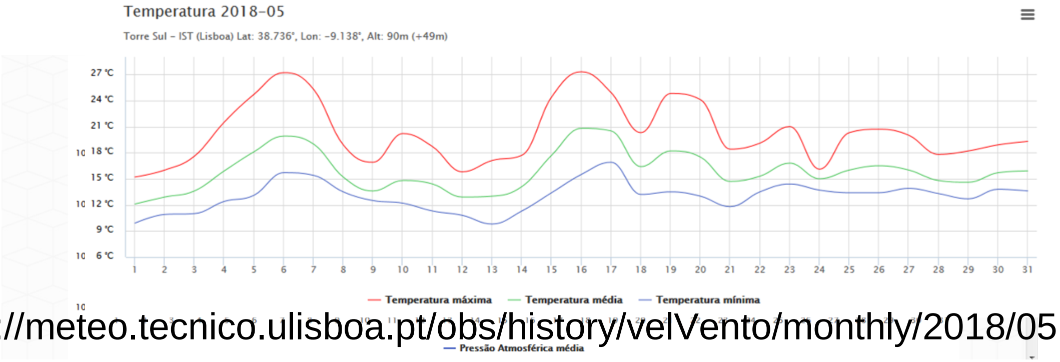
Técnico

Lat:38.736°, Lon:-9.138°, Alt: 90m (+49m)

Data available since 2012 every hour.







http://meteo.tecnico.ulisboa.pt/obs/history/velVento/monthly/2018/05

http://meteo.tecnico.ulisboa.pt/public/emaIST_wind.dat

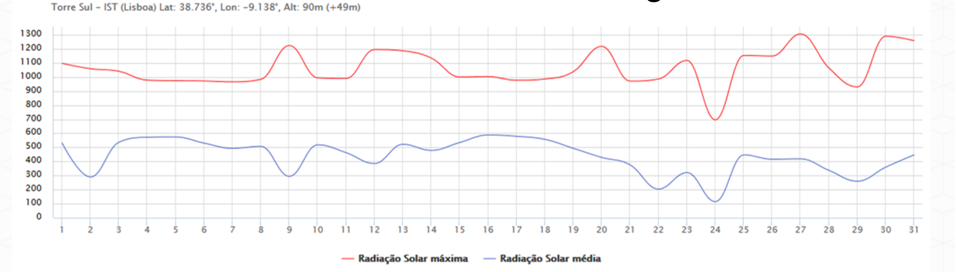
IST Station

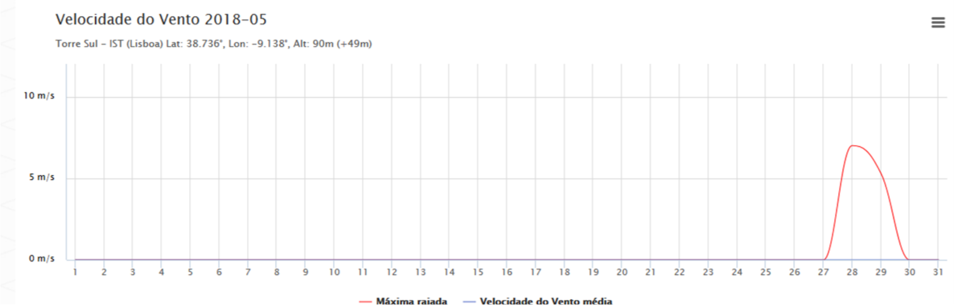
Técnico

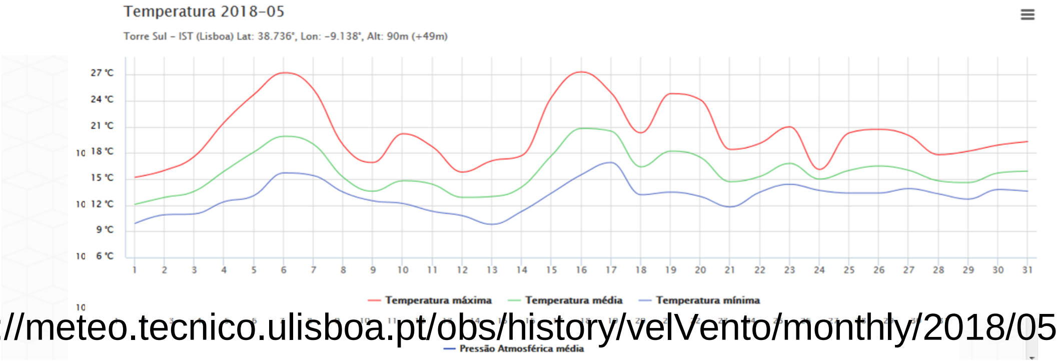
Lat:38.736°, Lon:-9.138°, Alt: 90m (+49m)

Data available since 2012 every hour.

Thanks Jorge Palma!!!





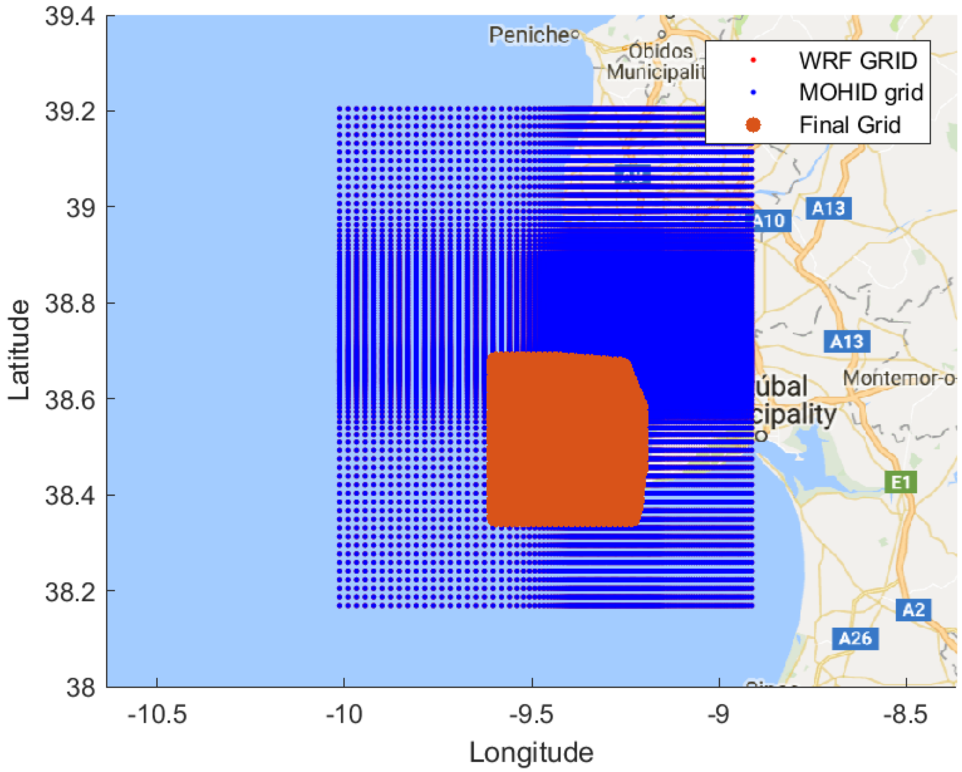


http://meteo.tecnico.ulisboa.pt/obs/history/velVento/monthly/2018/05

http://meteo.tecnico.ulisboa.pt/public/emaIST_wind.dat

WRF grid

- IST
- Cascais
- Aereopuerto



Pre-Processing workflow for comparison

1. **Download HF Data from web**
   Using tjhis tool that download data from Hydrografico website

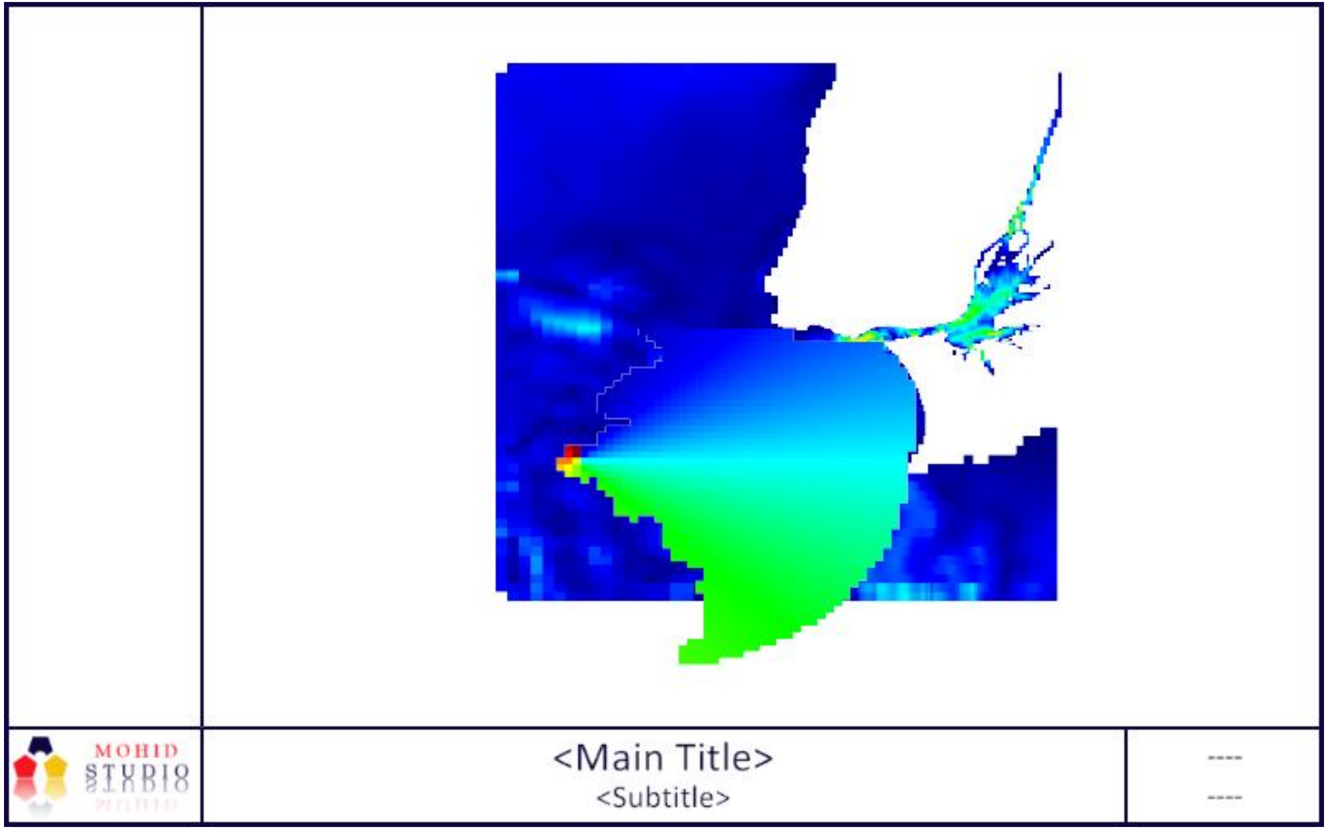
2. **Glue + Convert data to hdf5 and interpolated to a grid**
   Use the convert function with an specifed grid ann interpolated

3. **Glue Mohid Tagus Data to hdf5**
   It use the tool glue to specified the files to be merged

4. **Interpolate Mohid model data to HF grid (or viceversa)**
   Using tjhis tool that download data from Hydrografico website

5. **Skille Model analysis**
   Using tjhis tool that download data from Hydrografico website

MOHID Tools



Trying to use Mohid tools for the first time, I will use the tools later