Consiglio Nazionale delle Ricerche

CONSORZIO LaMMA

Regione Toscana

An adaptable modelling approach to the management of toxic microalgal bloom events in coastal areas

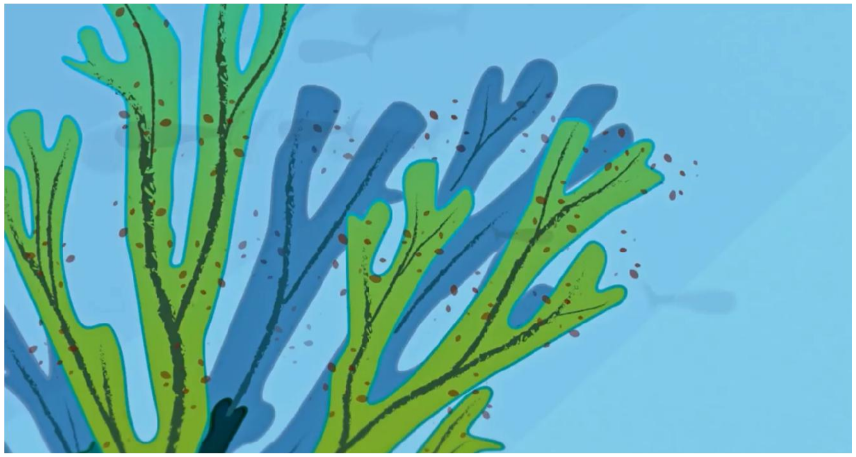
V.Asnaghi[1], D.Pecorino[2], E.Ottaviani[3], A.Pedroncini[4], R.Bertolotto[5], M.Chiantore[1]

(1) University of Genova - DISTAV; (2) MESA S.r.l.; (3) On Air S.r.l.; (4) DHI S.r.l.; (5) Environmental Agency of Liguria Region - ARPAL

LATEST MODELLING TECHNIQUES FOR SHALLOW SEAS

OCTOBER 17 - 19 JONSMOD 2018

Ostreopsis *ovata*



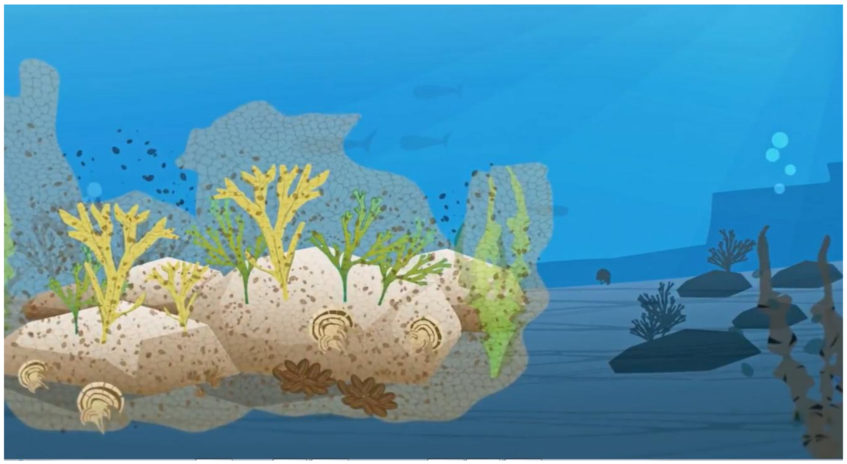
https://youtu.be/XNzZRvP_TlM

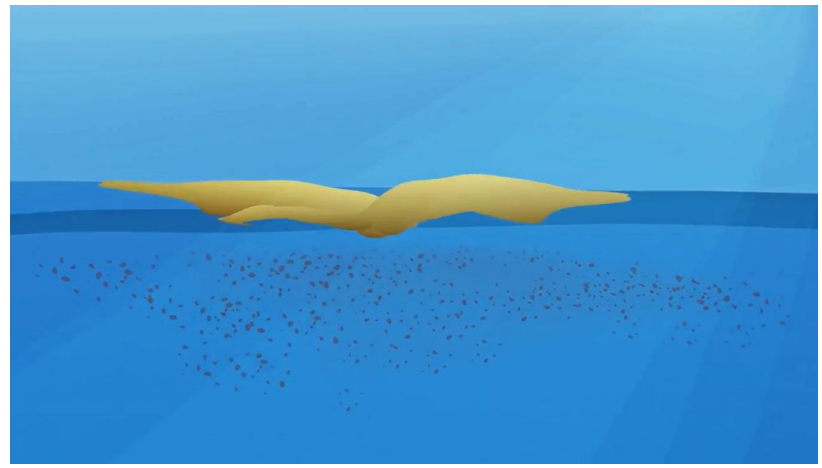
Consiglio Nazionale delle Ricerche

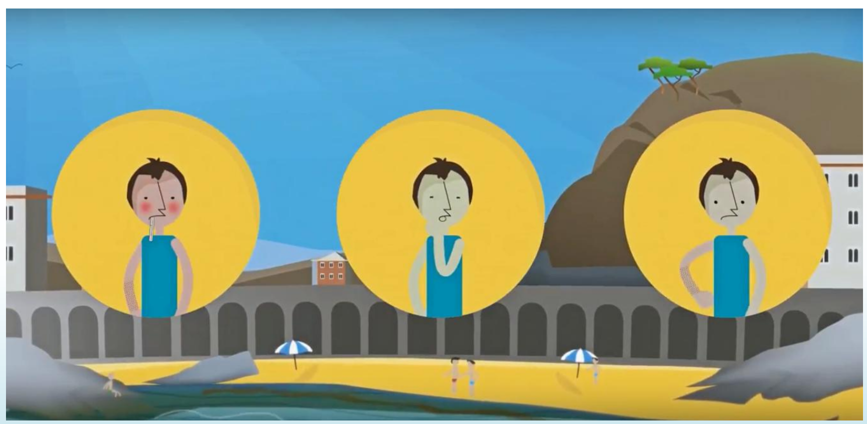
CONSORZIO LaMMA

Regione Toscana

Ostreopsis *ovata*







https://youtu.be/XNzZRvP_TlM

OCTOBER 17 - 19 JONSMOD 2018

Ostreopsis *ovata*

Lack of a robust theoretical model of correlation between the algal bloom development and the main chemical-physical parameters of the system.

Seawater parameters:

- Temperature, salinity
- Nitrites, Nitrates, Phosphates

Atmospheric parameters:

- Air temperature
- Wind speed and direction
- Wave height and direction
- Barometric pressure
- Rain / dry

Site description:

- Substrate
- Macroalgal species

Other:

- Week of the year
- Last measured concentration of Ostreopsis

Exploration of Historical Datasets: Genoa (Quarto) 2006-2013

- ❖Surface Seawater Temperature (SST)
- ❖Salinity

directly measured *in situ* with a probe or from models

- ❖Wind Speed
- ❖Wind Direction
- ❖Wave Height
- ❖Barometric Pressure

From meteorological stations or models

❖*Ostreopsis* cells/l (Log_{10})

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots.. + \varepsilon$$

Ostreopsis cells/l ~ 38 + 0.24 SST*** + 0.003WindDirection – 0.04 Barometric_pressure

$R^2=0.18$; p-value: < 0.0002 AIC = 283.88

Exploration of Historical Datasets: Genoa (Quarto) 2006-2013

- ❖Surface Seawater Temperature (SST)
- ❖Salinity

directly measured *in situ* with a probe or from models

- ❖*Ostreopsis* cells/g FW (Log_{10})

- ❖Wind Speed
- ❖Wind Direction
- ❖Wave Height
- ❖Barometric Pressure

From meteorological stations or models

- ❖*Ostreopsis* cells/l (Log_{10})

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \varepsilon$$

Ostreopsis cells/l ~ 38 + 0.24 SST*** + 0.003WindDirection – 0.04 Barometric_pressure

$R^2=0.18$; p-value: < 0.0002 AIC = 283.88

Ostreopsis cells/l ~ -0.53 + 0.80 *Ostreopsis* cells/g *** - 0.37 Wave Height* + 0.001WindDir

$R^2=0.79$; p-value: < 2.2e-16 AIC = 137.36

Consiglio Nazionale delle Ricerche

CONSORZIO LaMMA

Regione Toscana

Development of the predictive tool

Lack of a robust theoretical model of correlation between the algal bloom development and the main chemical-physical parameters of the system.

Seawater parameters:

- Temperature, salinity
- Nitrites, Nitrates, Phosphates

Atmospheric parameters:

- Air temperature
- Wind speed and direction
- Wave height and direction
- Barometric pressure
- Rain / dry

Site description:

- Substrate
- Macroalgal species

Other:

- Day of the year
- Last measured concentration of Ostreopsis

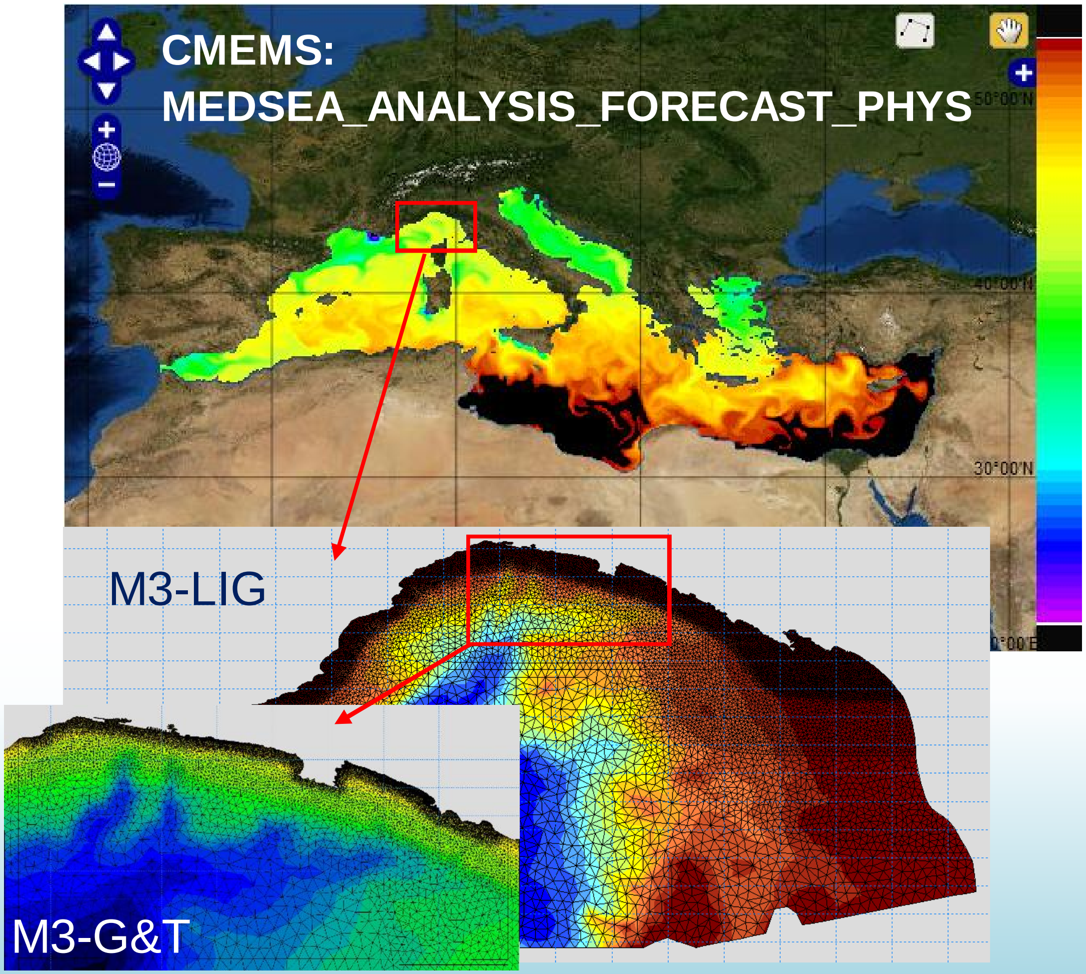
Consiglio Nazionale delle Ricerche

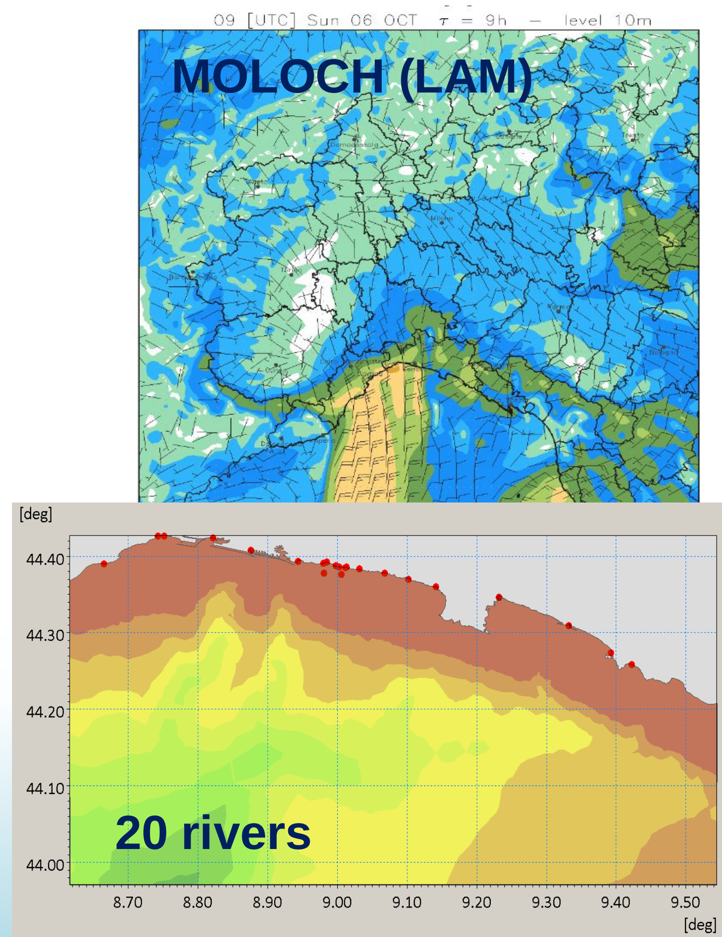
CONSORZIO LaMMA

Regione Toscana

Generation of meteo-marine data

Operational chain @ARPAL



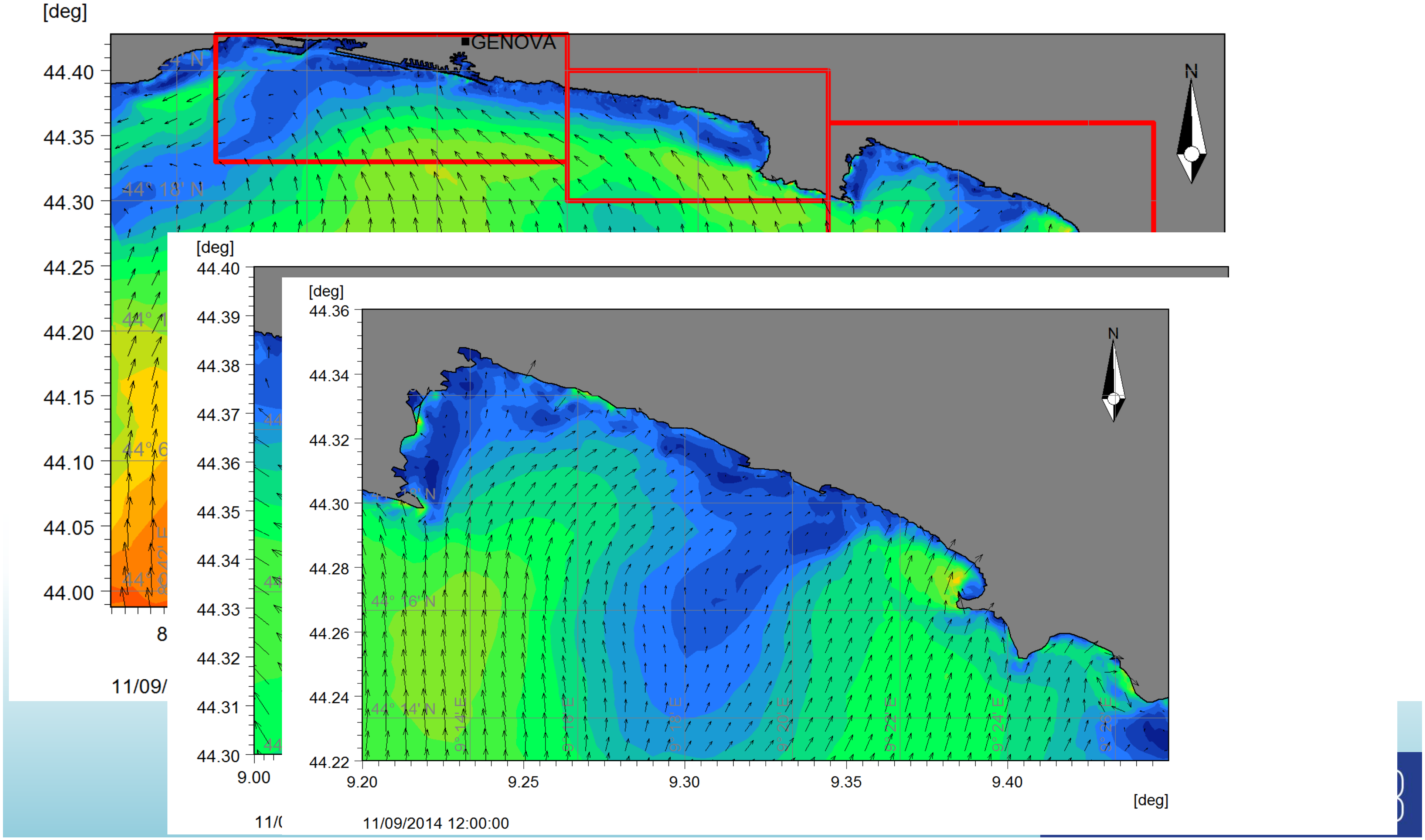


Consiglio Nazionale delle Ricerche

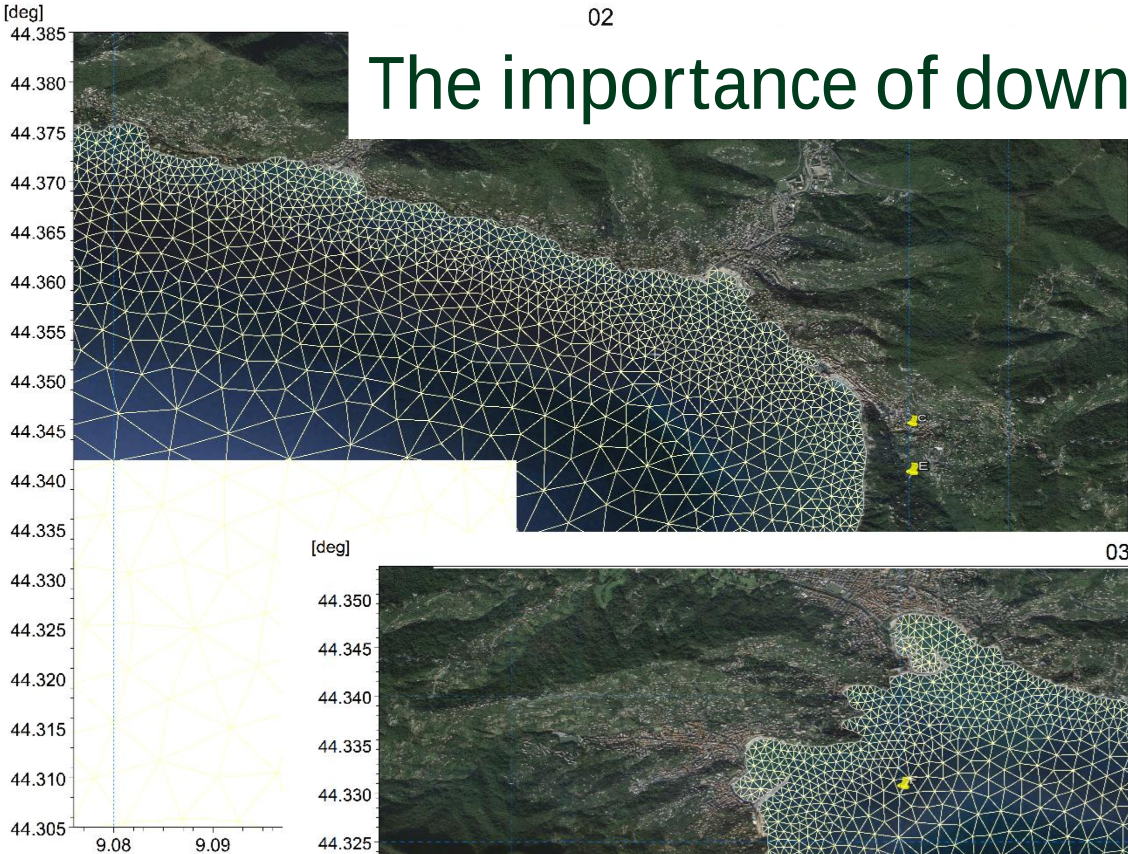
CONSORZIO LaMMA

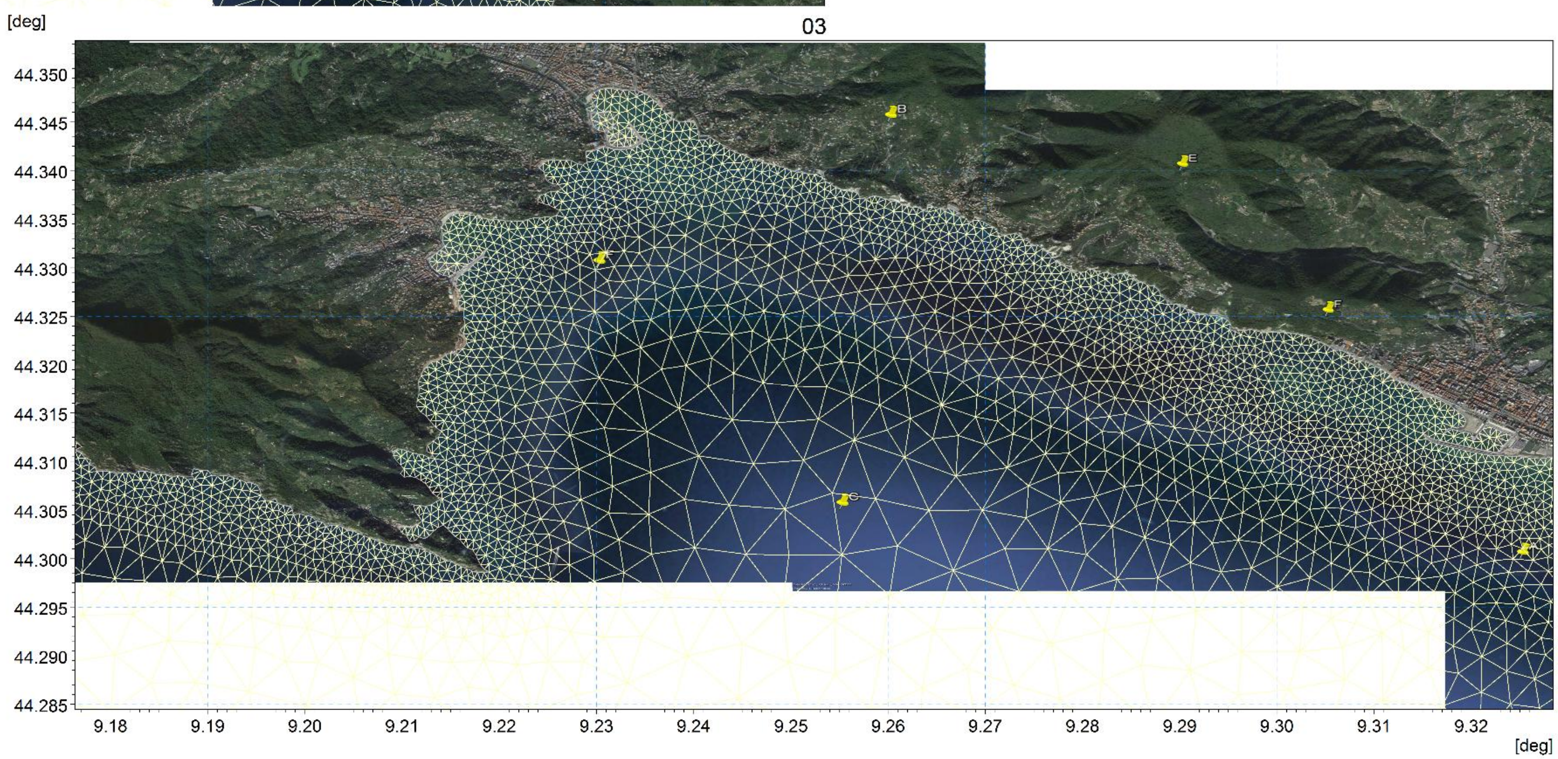
Regione Toscana

The importance of downscaling

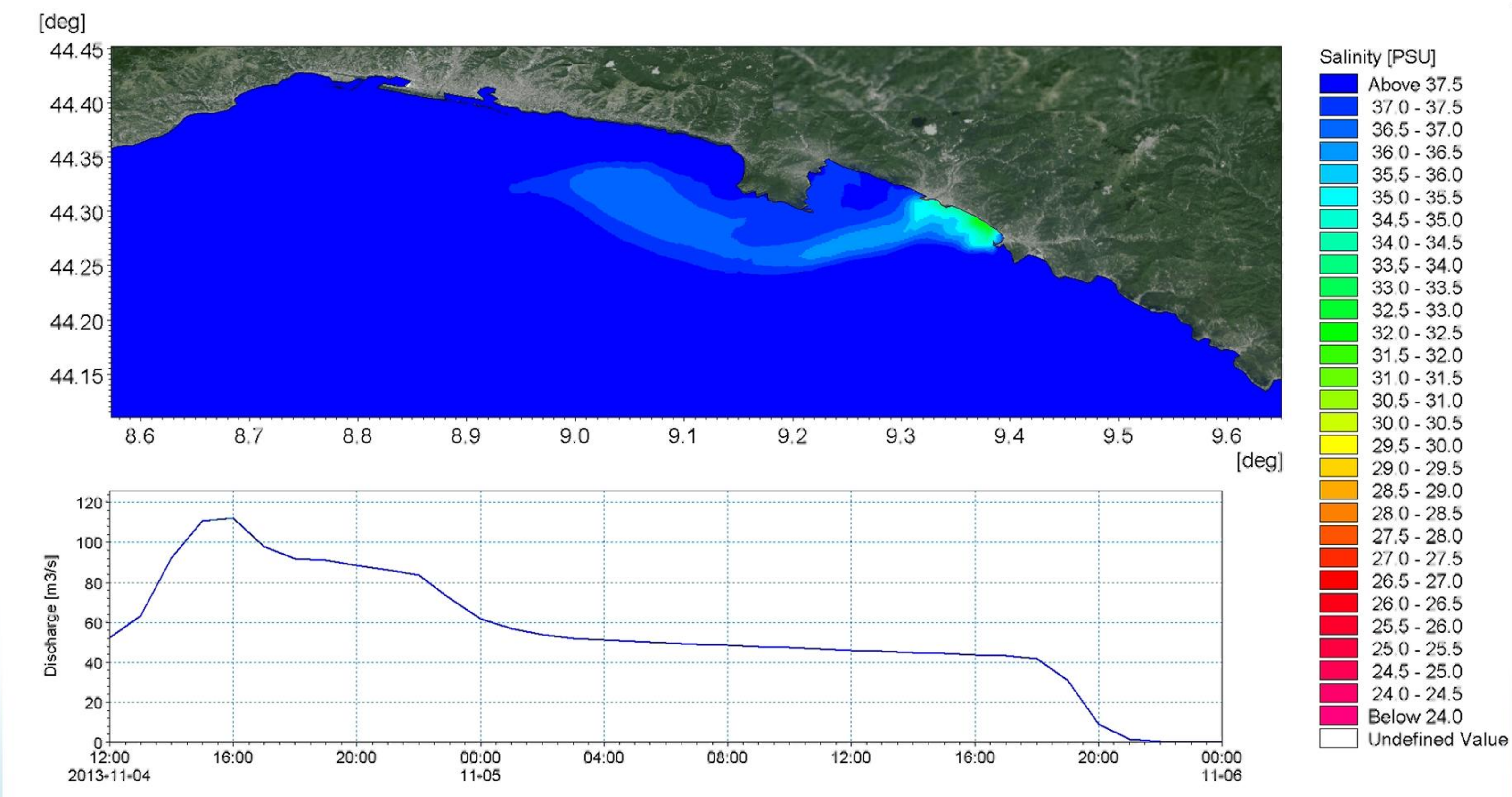


The importance of downscaling





The importance of downscaling



Development of the predictive tool

Lack of a robust theoretical model of correlation between the algal bloom development and the main chemical-physical parameters of the system.

Seawater parameters:

- Temperature, salinity
- Nitrites, Nitrates, Phosphates

Site description:

- Substrate
- Macroalgal species

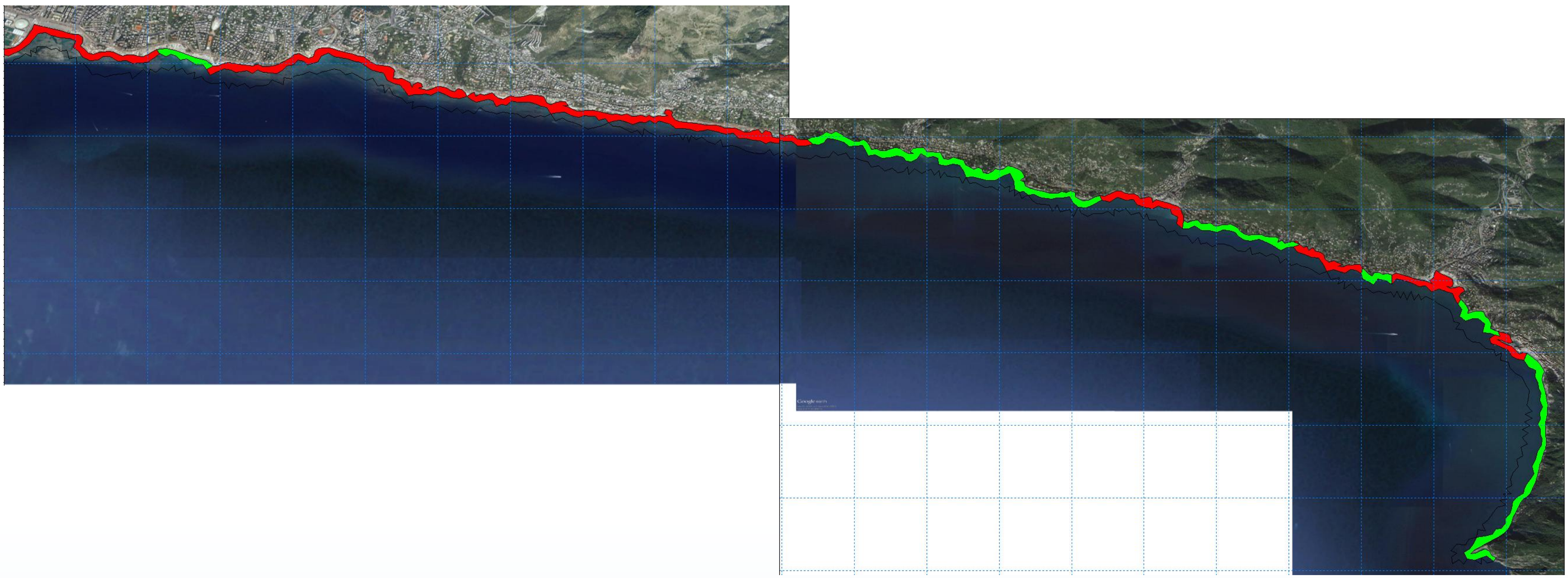
Atmospheric parameters:

- Air temperature
- Wind speed and direction
- Wave height and direction
- Barometric pressure
- Rain / dry

Other:

- Day of the year
- Last measured concentration of Ostreopsis

Local characteristics are parametrized



Consiglio Nazionale delle Ricerche

CONSORZIO LaMMA

Regione Toscana

Development of the predictive tool

Dataset construction:
280 entries present
Sampling sites:
7 sites in the Genoa area (Ligurian Sea)
Time frame:
2006 - 2013, (July-September)
Data:
Samples of *O.* cf. *ovata* into the water column (cells/l) collected by Genoa University and ARPAL

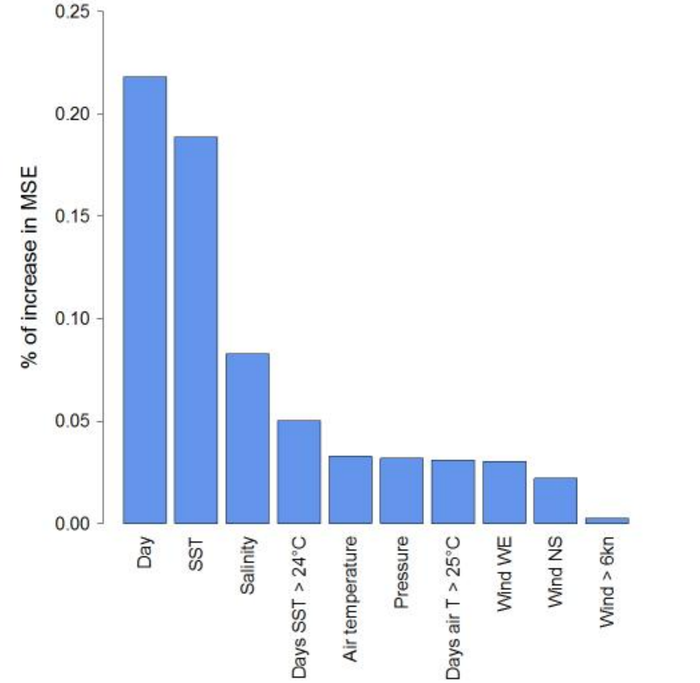
Selected covariates:

- day of the year
- sea surface temperature
- days before sampling when sea surface temperature exceeded 24°C
- salinity
- air temperature
- days before sampling when air temperature exceeded 25°C
- atmospheric pressure
- North-South component of the wind velocity
- East-West component of the wind velocity
- presence/absence of wind above 6 m/s

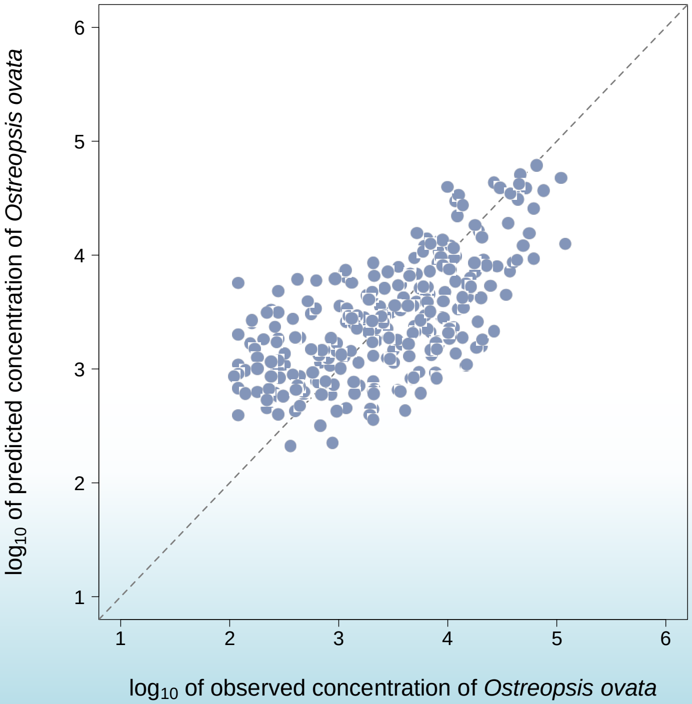
Selected approach: Quantile Regression Forests

The Quantile Regression Forest approach (Meinshausen, 2006) is based on the Random Forest method (Breiman, 2001), but allows for a whole distribution of predicted values to be generated for each combination of input features.

Non linear regression problem: 10 meteorological predictors → Log_{10} *Ostreopsis* cells/l

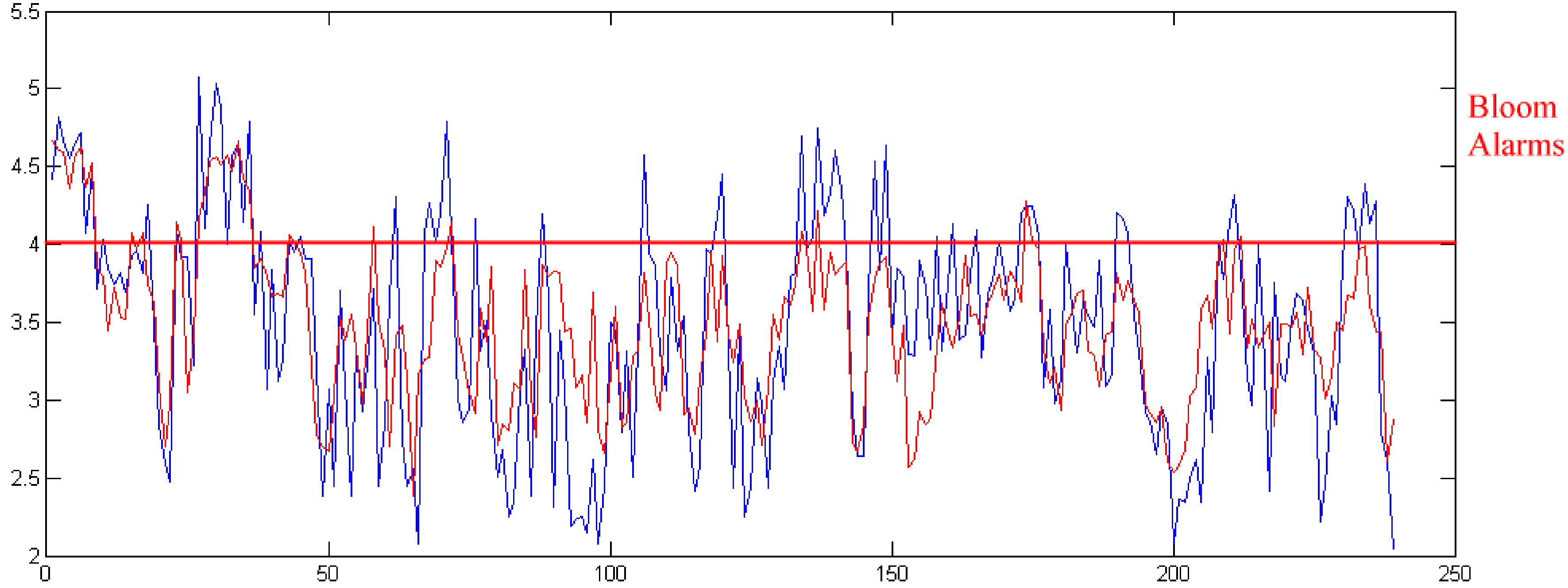


$R^2 = 0.43$
Pearson's $\rho = 0.66$



Quantile Regression Forests for bloom detection

Bloom detection can be performed by analyzing the data distribution of the predicted values from the trees and evaluating the probability of exceeding a given threshold



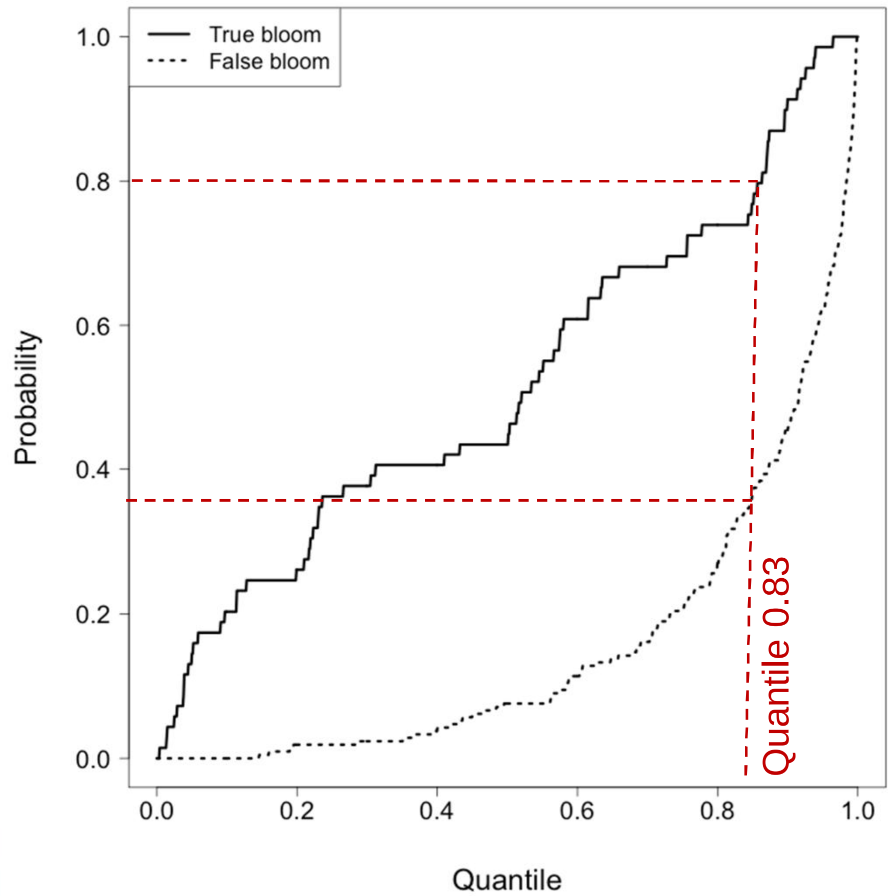
Bloom threshold set at 10 000 cells/l

Consiglio Nazionale delle Ricerche

CONSORZIO LaMMA

Regione Toscana

ROC curve for classification accuracy evaluation



True bloom: predict a bloom when it actually occurred
False bloom: predict a bloom that did not occur

Assuming 80% detection rate we need to select the 0.83 quantile. This imply 38% false alarm rate
Other quantiles give different tradeoffs

Consiglio Nazionale delle Ricerche

CONSORZIO LaMMA

Regione Toscana

Validation of the predictive tool

The QRF model built on 2006-2013 data was validated by generating predictions of concentrations of *Ostreopsis cf. ovata* in the water column during summer 2015:

- meteorological data generated weekly from June to September 2015 by DHI

- predicted values of cells/l compared to actual measurements of *O.* cf. *ovata* performed on samples collected in Genoa-Quarto in the same period

- 3 blooms successfully detected (and 9 false bloom but very close to the threshold)

In a management perspective, two threshold (10.000 and 30.000 cells/l, Health Ministry Guidelines, 2007) was considered to discriminate between bloom and no-bloom conditions.

Automatic production of predictive maps

Using the same model, we generated daily predictive maps for the whole Genoa area, using forecasted meteo-marine data provided by DHI

Output:

- concentration map for *Ostreopsis ovata* at 10:00 every day
- probability of exceeding the 10.000 cells/l threshold
- probability of exceeding the 30.000 cells/l threshold

The values provided by the model were corrected for the different locations based on the characterization of the sites. The correction factors have been calculated on the basis of historical data available in different locations along the Ligurian coast and have been attributed to the different portions of coast based on their characteristics.

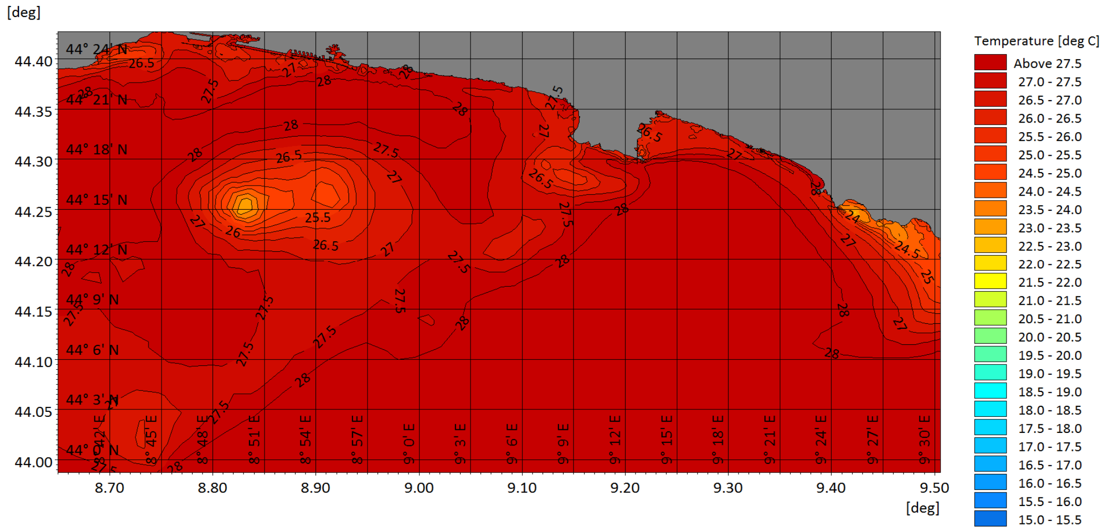
Consiglio Nazionale delle Ricerche

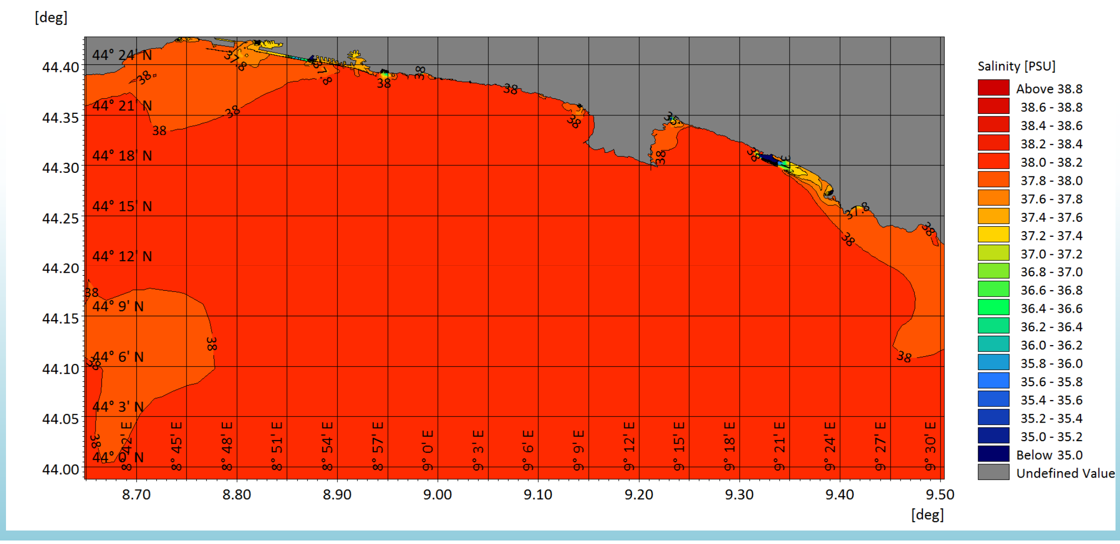
CONSORZIO LaMMA

Regione Toscana

Example

15 July 2015 h.10.00: Sea surface temperature and salinity





OCTOBER 17 - 19 JONSMOD 2018

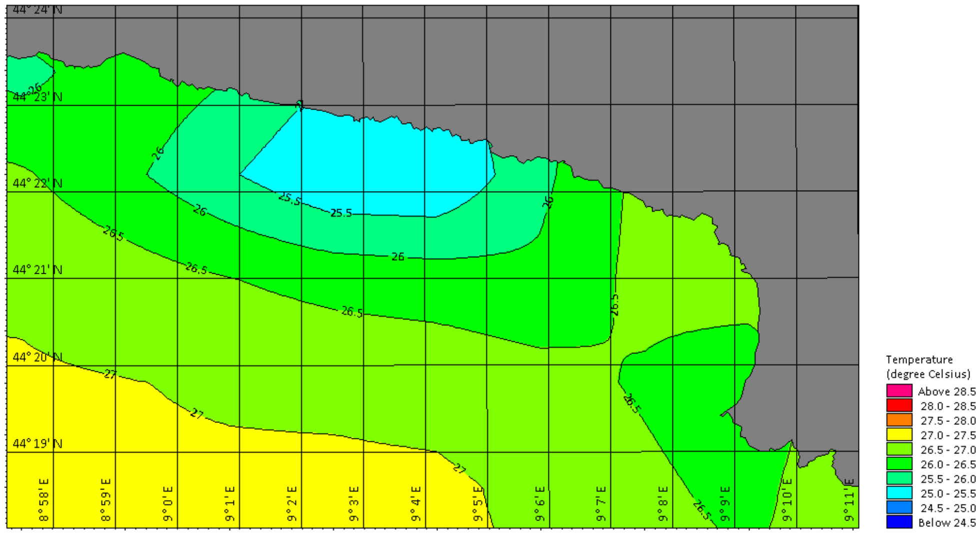
Consiglio Nazionale delle Ricerche

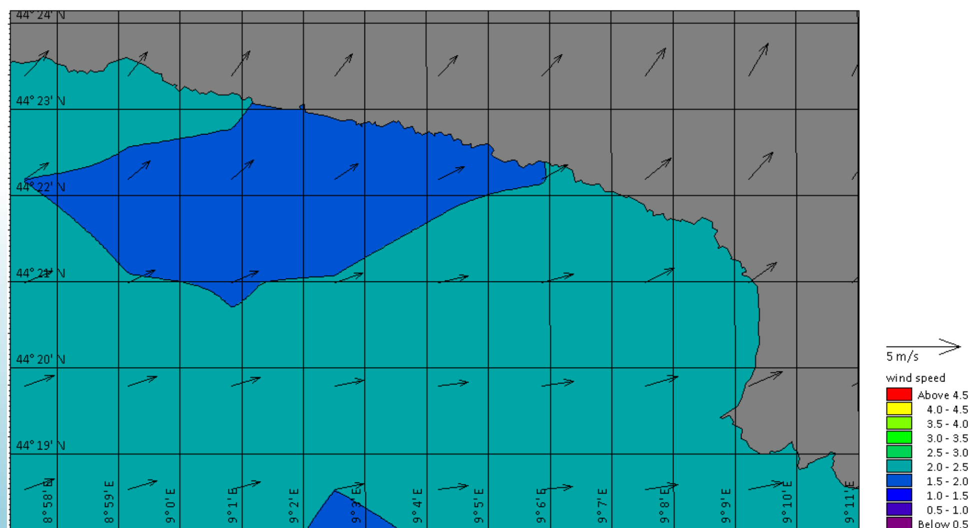
CONSORZIO LaMMA

Regione Toscana

Example

15 July 2015 h.10.00: Air temperature and wind speed / direction





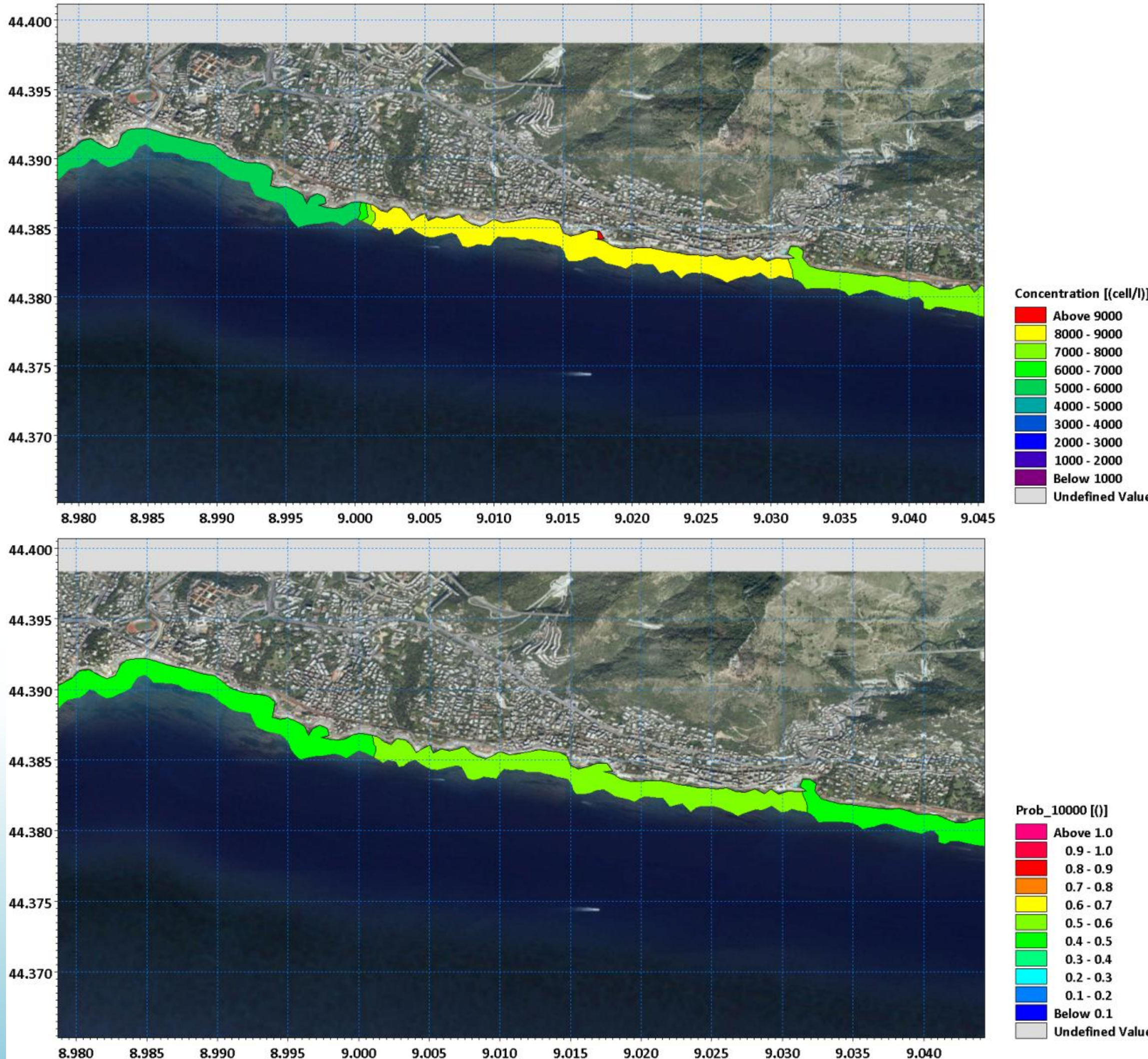
Consiglio Nazionale delle Ricerche

CONSORZIO LaMMA

Regione Toscana

Example

15 July 2015 h.10.00: Concentration and exceeding 10.000 threshold maps



ML algorithms for spatial and temporal prediction

Random forest and other Machine Learning algorithms can be adapted to generate spatial and temporal predictions, adding time/space dependent explanatory variables to the set of environmental covariates, to incorporate seasonal and geographical effects into the prediction process of the target variable.

Advantages

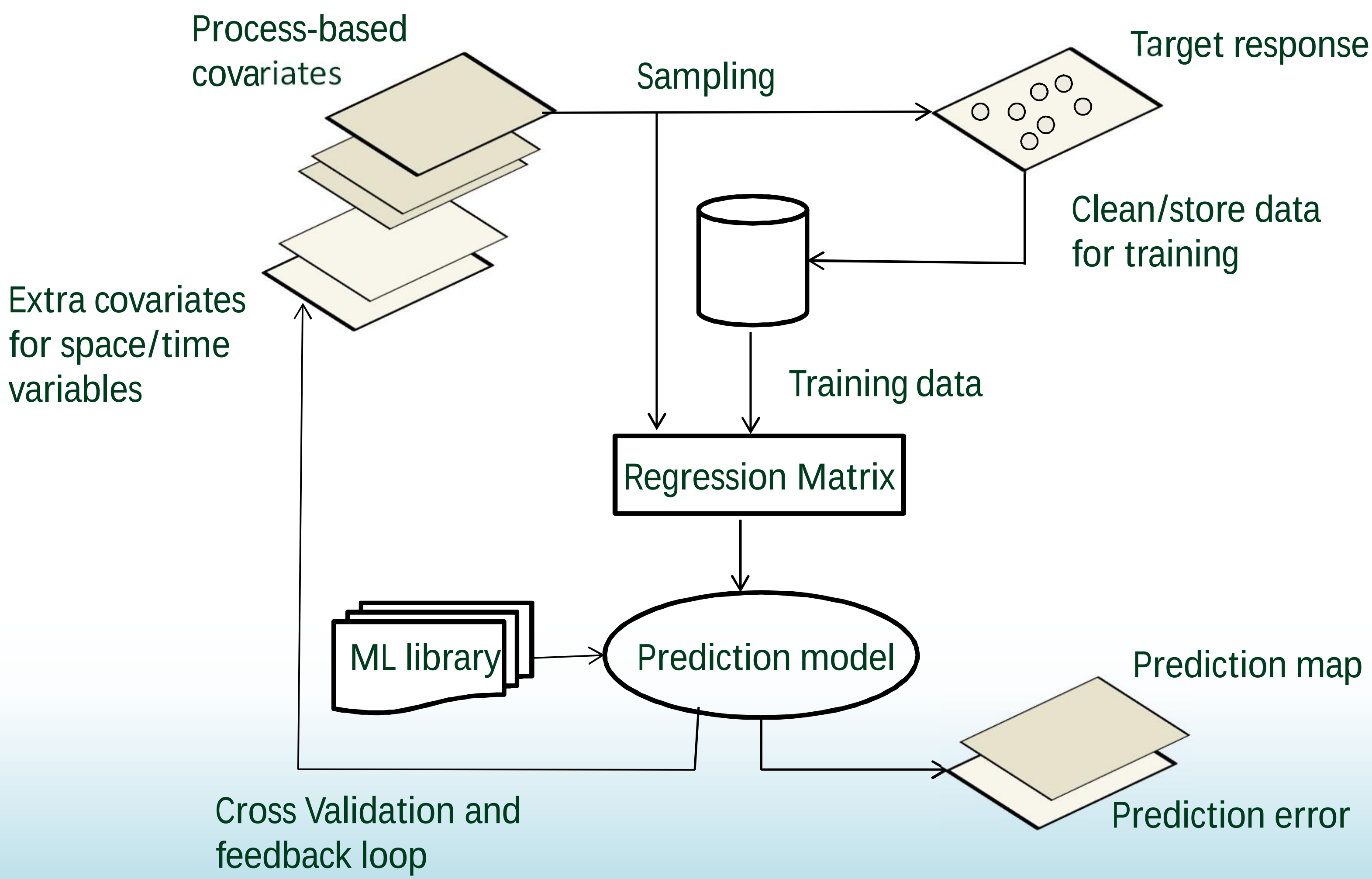
- no rigid statistical assumptions about the distribution and the stationarity of the target variable,
- more flexible towards incorporating, combining and extending covariates of different types,
- more informative maps characterized by the prediction error,
- attractive for building multivariate spatial prediction models that can be used as “knowledge engines” in various fields.

Consiglio Nazionale delle Ricerche

CONSORZIO LaMMA

Regione Toscana

ML prediction architecture



Thank you for your attention!

Andrea Pedroncini

DHI S.r.l.

+39 328 6341889

anp@dhigroup.com

Ennio Ottaviani

On Air S.r.l.

+39 348 3206619

ennio.ottaviani@onairweb.com