AIQUAM: Artificial Intelligence-based water QUAlity Model

C. G. De Vita[1], G. Mellone[1], D. Di Luccio[1], S. Kosta[2], A. Ciaramella[1], R.Montella[1]
[1]University of Naples “Parthenope”, Science and Technologies Department, Napoli, Italy
[2]Aalborg University, Department of Electronic System, Copenhagen, Denmark

Abstract

Monitoring the impact of the pollutants on the sea is a crucial issue for coastal human activities, such as aquaculture. However, leveraging a continuous microbiological laboratory analysis is unfeasible for costs and practical reasons. Fish and mussel farms are critically sensitive to seawater quality and thus require continuous monitoring to enforce food security and prevent any possible disease affecting human health [1]. Here we present a novel methodology finalized to predict water quality as categorized indexes leveraging an integrated approach between computational components and artificial intelligence techniques. As a paradigm demonstrator, we couple WaComM++ with **AIQUAM**.

The use case presented is an application of AIQUAM in the Bay of Naples (Campania Region, Italy) for predicting bacteria contaminants in mussel farms. The results are encouraging as the model reached a correct prediction rate of **93%**.

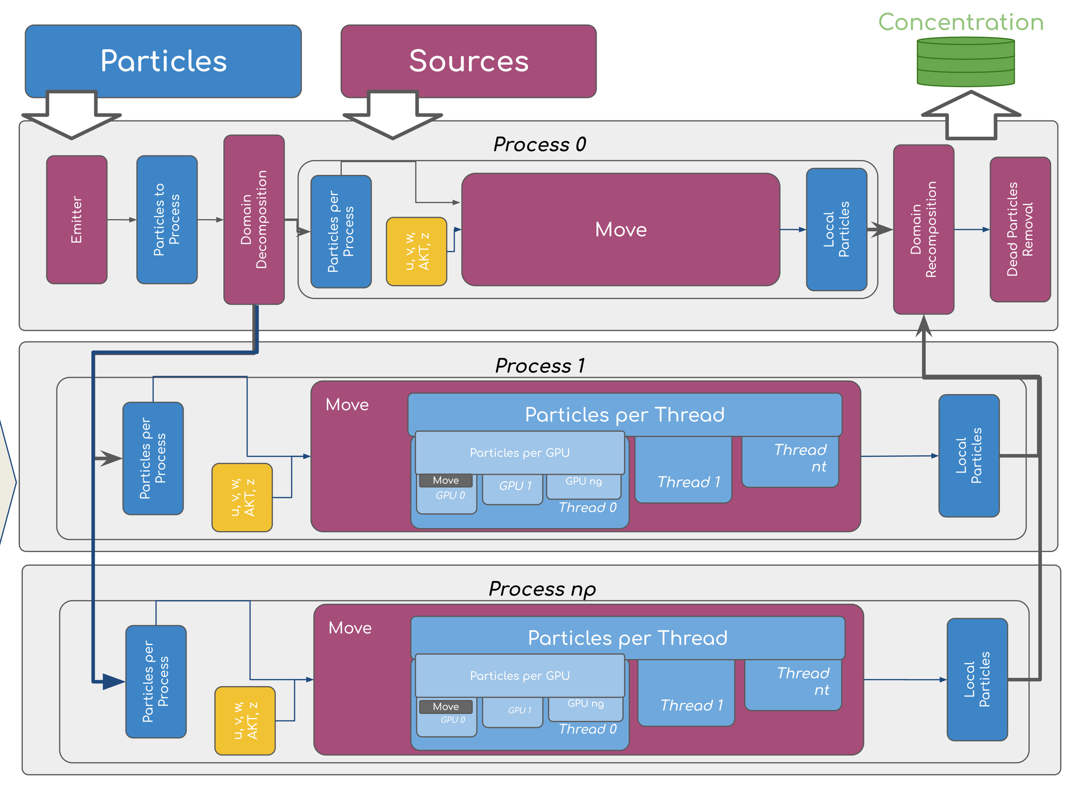
WaComM++

WaComM++ [2] is a Lagrangian model for inert tracers transport and diffusion simulation and forecast. It works with an ocean dynamics model as ROMS.
It supports the **hierarchical** and **heterogeneous** parallelization scheme.

For each iteration, the particles to be processed are divided between processes and threads. If one or more GPUs are available, each thread distributes its computational burden on each available GPU.

WaComM++ works on any combination of processes, threads, and GPUs. It has been designed with computational elasticity in mind [3] to benefit from High-Performance Cloud Native Computing environments.

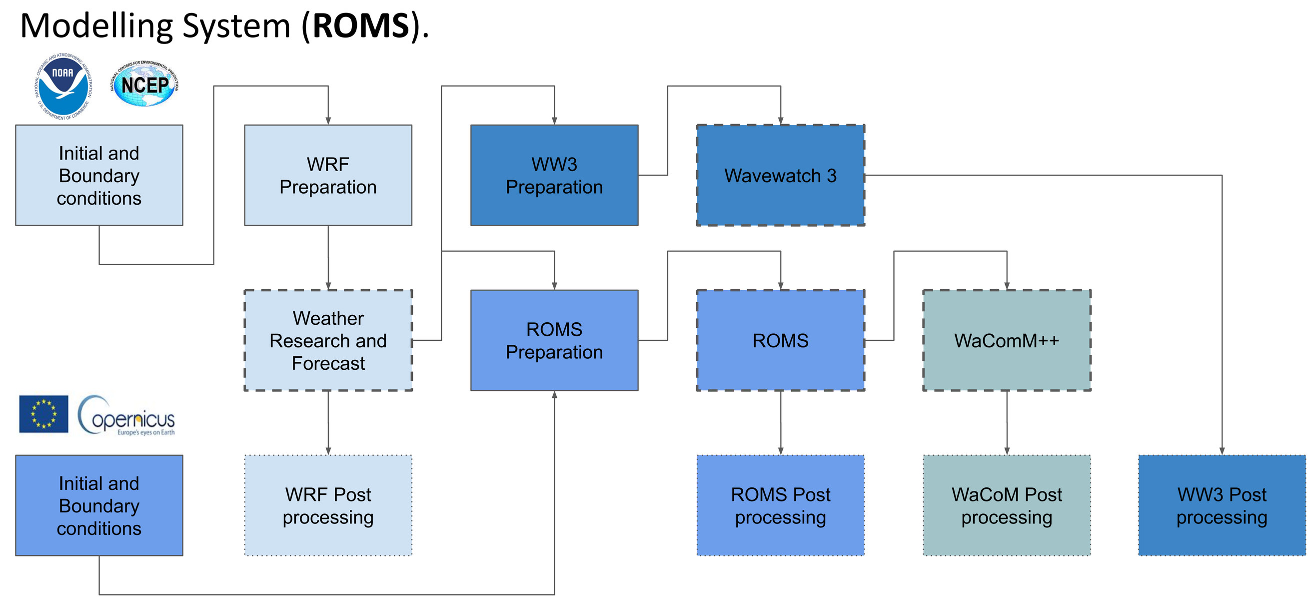
https://github.com/CCMMMA/wacommplusplus



WaComM++ parallelization schema

Forecasting System

AIQUAM is an application of WaComM++ that is a part of model chain based on Weather Research and Forecast (**WRF**) to compute the wind conditions that are one of the forcing of sea surface current forecasted by Regional Oceanic Modelling System (**ROMS**).



https://github.com/CCMMMA/wacommplusplus

AIQUAM

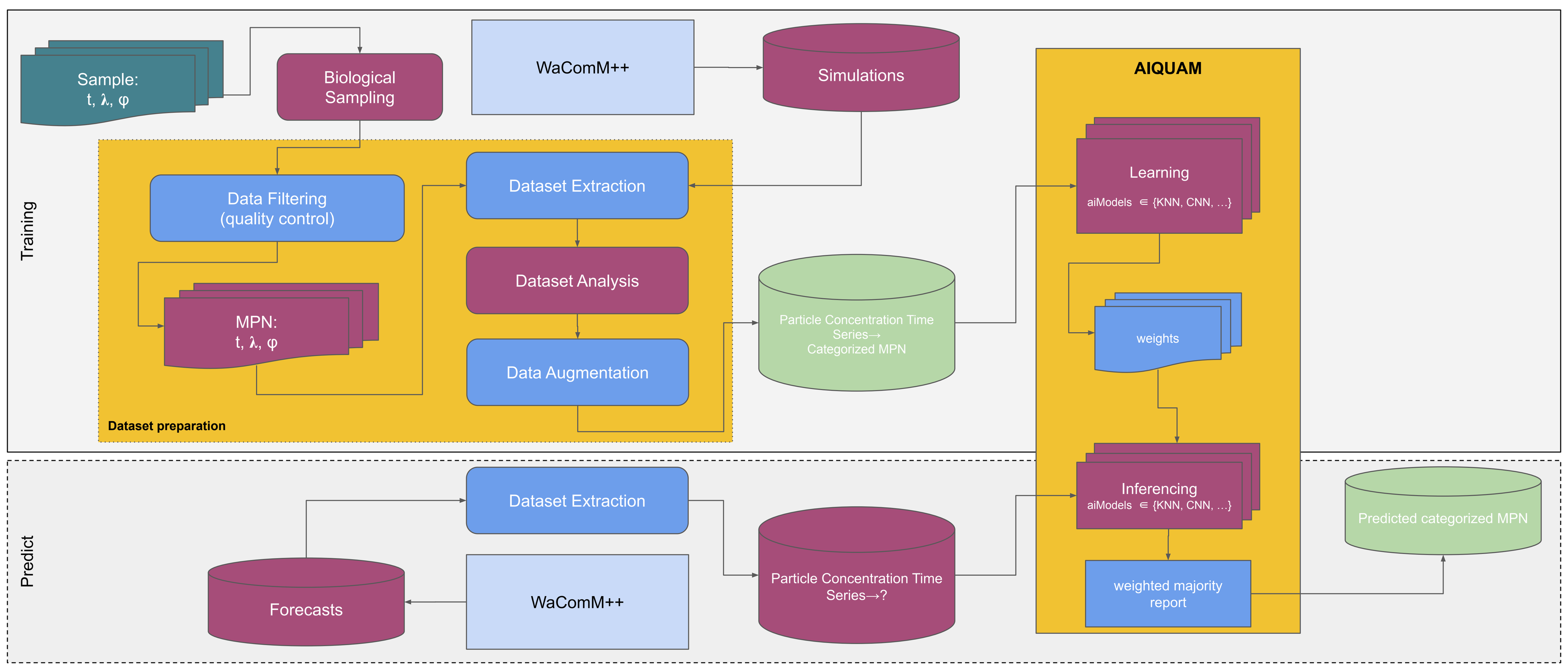
AIQUAM implements an AI model for seawater quality predictions.

The model performs **time series classification** leveraging various and different algorithms and then performs a weighted majority report for predicting the best result.

The time series classification (TSC) problem can be approached as *distance based* and *feature based*.

It consists of training a dataset classifier to map possible inputs to a probability distribution over the class variable values (labels). We tested three ML models:

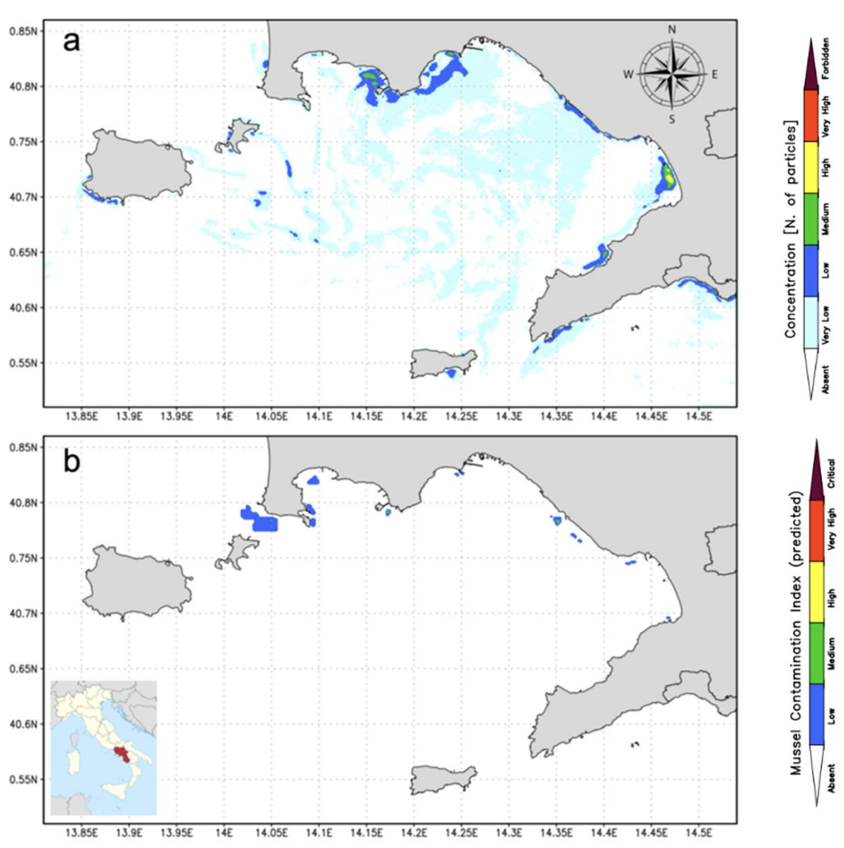
- **KNN**
- **KNN + DTW**
- **CNN**



AIQUAM is trained using the categorized concentration of bacteria in seawater samples as labels of events represented by the concentration time series from WaComM++.

At the prediction stage, for each WaComM++ forecasted time iteration, AIQUAM estimates the category of each grid point.

Use case



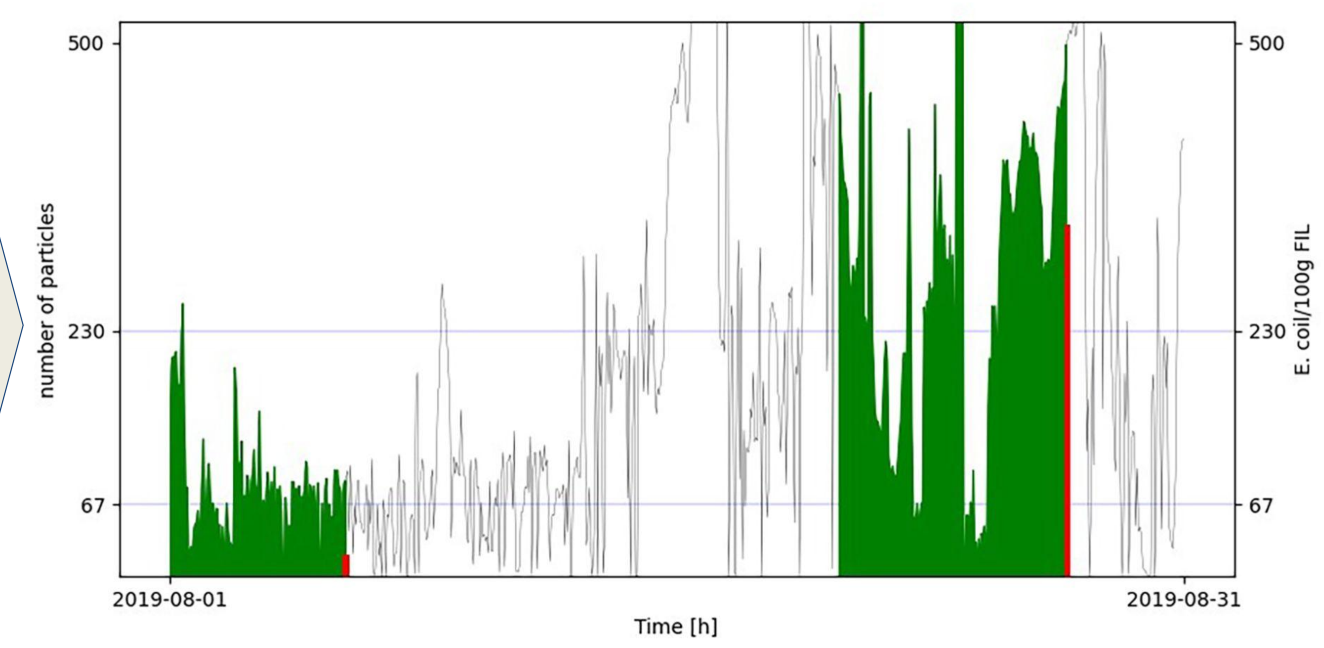
WaComM++ (a) and an AIQUAM (b) typical result in terms of hourly particle concentration and predicted categorized MPN of bacteria, respectively.

We applied **WaComM++ and AIQUAM** to assess and predict farmed mussels quality in the Bay of Naples, Italy, in terms of the categorized amount of most probable number (MPN) of *Escherichia Coli bacteria*.

The mussel assimilates bacteria by the hourly concentration of particles from the time t0 (microbiological sampling) to the **previous t0-72 hours**.

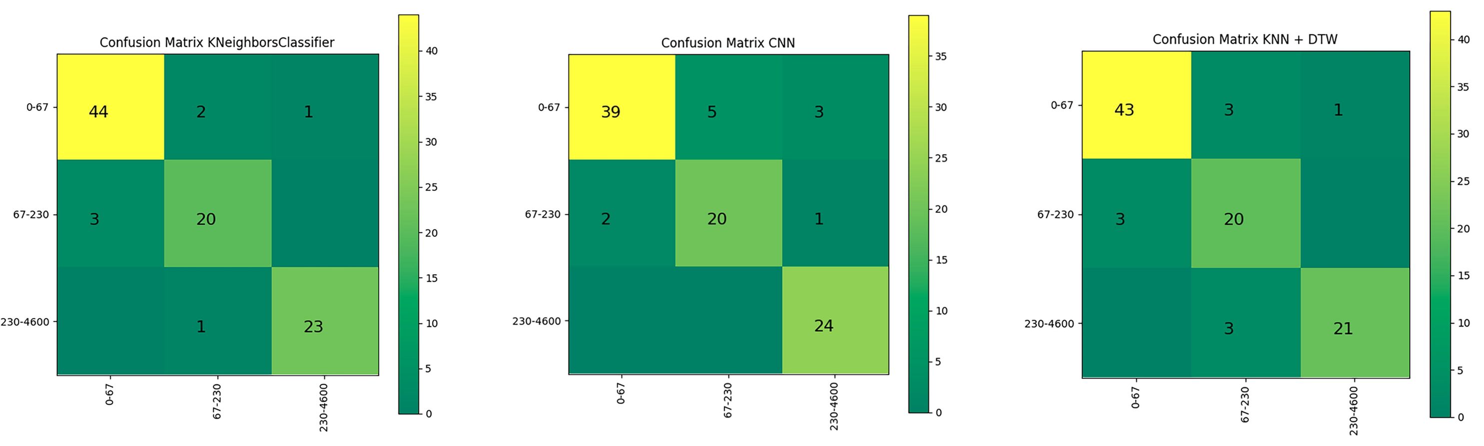
We used the in-situ microbiological sampling provided by the local government agency, and the results carried out by WaComM++ to train AIQUAM.

The dataset feature space of each event is represented in **green**.



The microbiological measurements are categorized into three classes and used as label space: class 0 (0-67 MPN/100 g); class 1 (67-230 MPN/100 g), and class 2 (230-4600 MPN/100 g). The labels are represented in **red**.

The number of samples in each class is unbalanced, so a data augmentation technique has been applied. The final dataset has 221 samples: 111 in class 0, 55 in class 1, and 55 in class 2.



Confusion matrix for the used ML models: the best results are obtained using the KNN model in which 44/47 samples in class 0, 20/23 samples in class, and 23/24 samples in class 2 were correctly classified, with an accuracy of approx. **93%**.

TRY WACOMM++

TRY AIQUAM

Conclusion

A novel ML methodology has been proposed to predict the concentration of pollutants in mussels. The best results were obtained with the KNN method. AIQUAM needs to be re-trained with a larger dataset to be used operationally.

This methodology can be used to improve the final quality of the meteo-oceanographic models, adding a predictive algorithm to the forecast algorithm.

References

[1] A. Galletti, R. Montella, L. Marcellino, A. Riccio, D. Di Luccio, A. Brizius, and I. Foster. Numerical and implementation issues in food quality modeling for human diseases prevention. In Healthinf, pages 526–534, 2017.

[2] R. Montella, D. Di Luccio, P. Troiano, A. Riccio, A. Brizius, and I. Foster. Wacomm: A parallel water quality community model for pollutant transport and dispersion operational predictions. In 2016 12th International Conference on Signal-Image Technology Internet-Based Systems (SITIS), pages 717–724, 2016.

[3] G. Martin, M. Marinescu, D. E Singh, and J. Carretero. Flex-mpi: an mpi extension for supporting dynamic load balancing on heterogeneous non-dedicated systems. In European Conference on Parallel Processing, pages 138–149. Springer, 2013.