HPC and AI for Prevision and Prediction of environmental phenomena (Day 2)

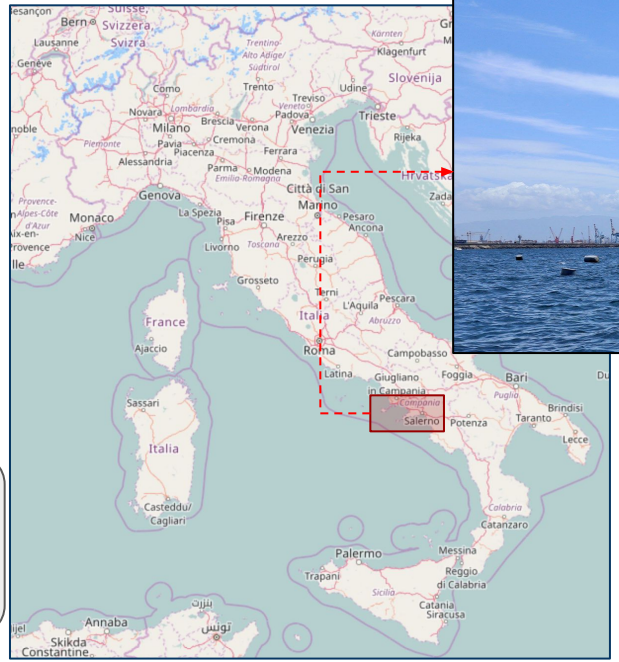
Universidad Carlos III de Madrid
April 20th and 21st, 2023

Prof. Raffaele Montella
University of Napoli "Parthenope"
https://raffaelemontella.it https://www.uniparthenope.it

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
EuroHPC Joint Undertaking
ADMIRE malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

Introduction & Motivation

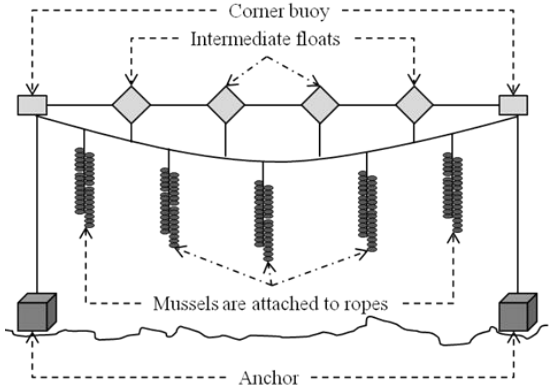
Mussels farming is an outstanding business cornerstone in the most part of Italian coastal regions.



MARKET

- Companies: 263(d)/886(r)
- Tons: ~64235 (⅔ EU prod - ISPRA)
- Euro/Kg: ~1.75 (average)

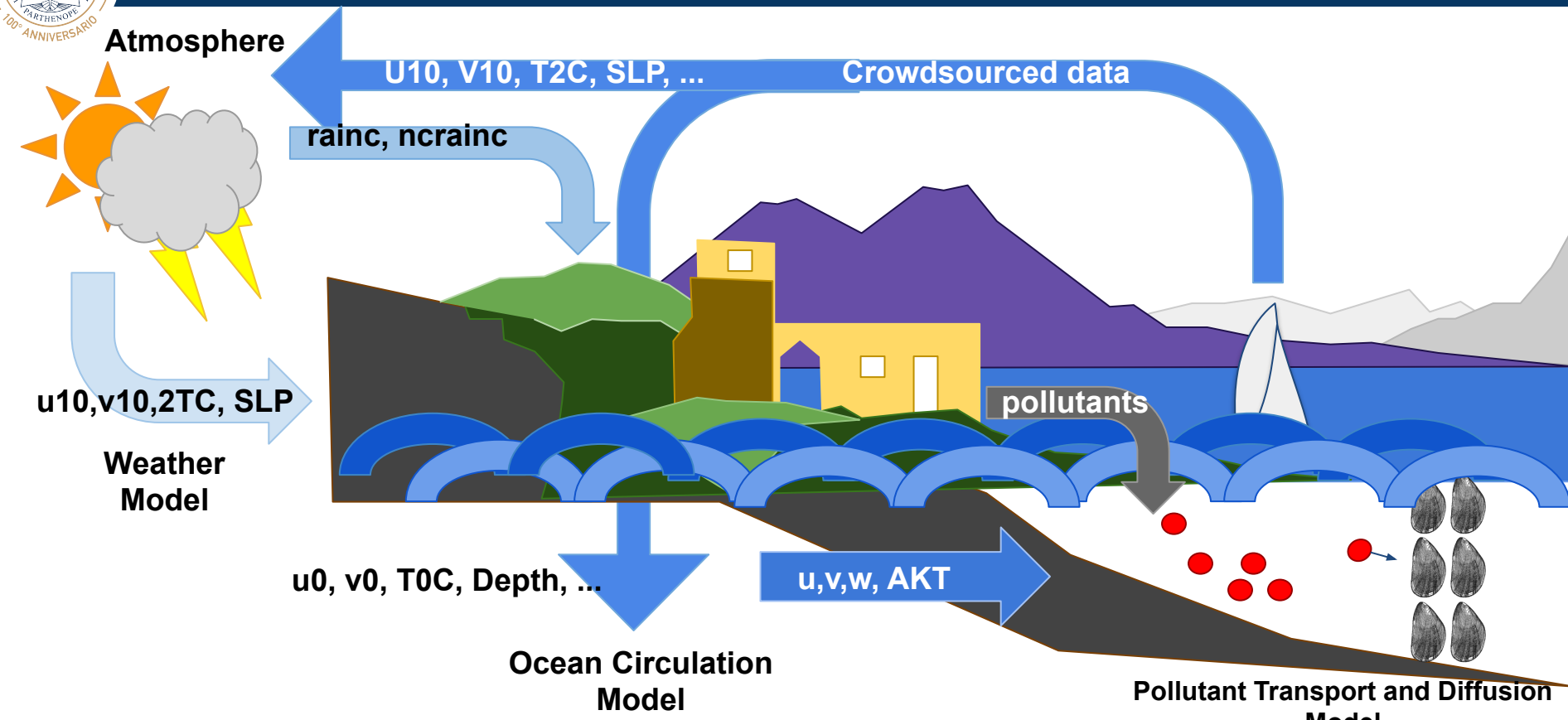
~112M€ (2013, Italy)



Make predictions about the pollutant concentration in mussel farms areas in order to limit human diseases.

EuroHPC Joint Undertaking
ADMIRE
malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

Contextualization



- *A cartoon-like representation of the big picture.*

A highly scalable high-performance Lagrangian transport and diffusion model for marine pollutants assessment

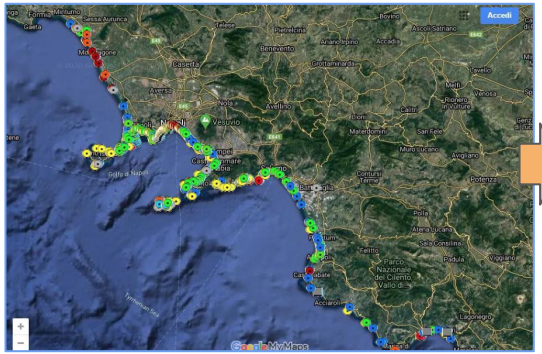
GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN

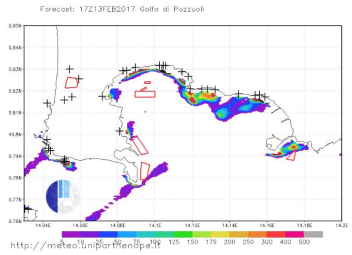
EuroHPC Joint Undertaking

ADMIRE
malleable data solutions for HPC

ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

Water quality Community Model ++





Initial Conditions (ROMS)

Emission Sources

WaCoM restart (previous run)

WaComM++

WaCoM Post processing

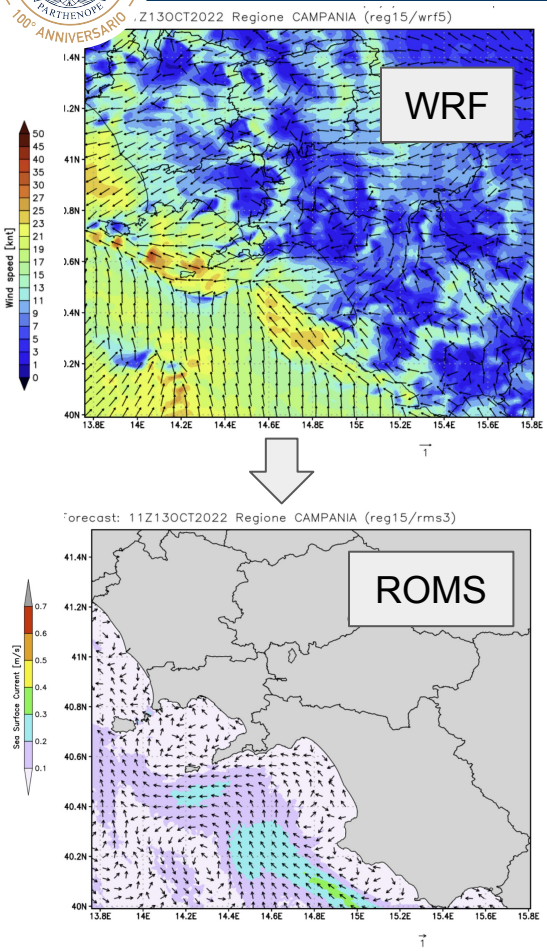
WaCoM restart

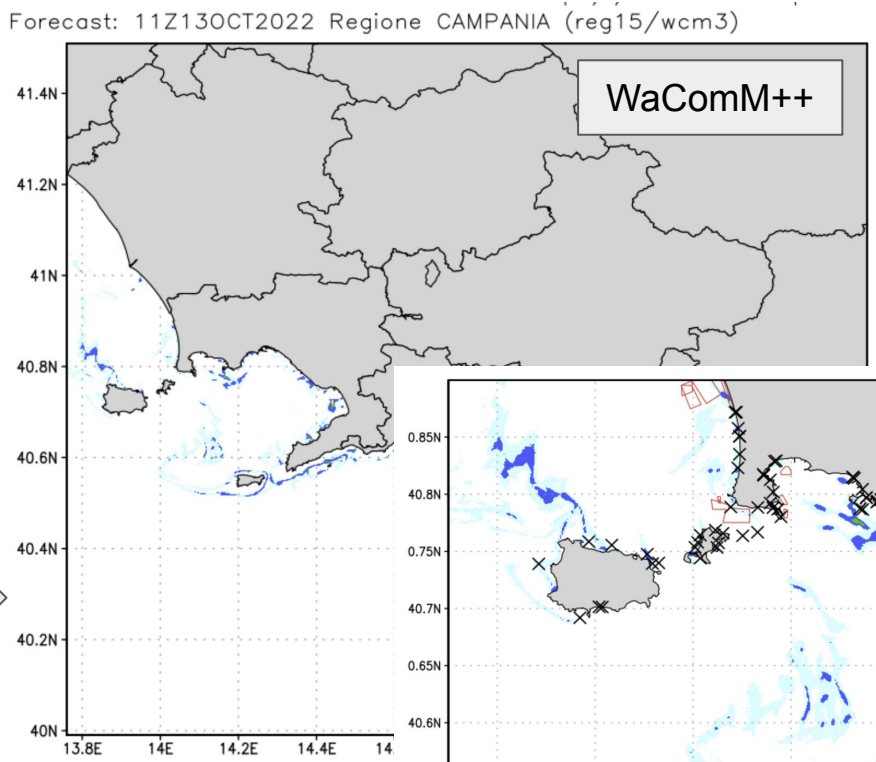
- **Lagrangian** model for the transport and diffusion of passive tracers.
- **Hierarchical** and **Heterogenic** parallelization (Distributed memory (MPI), Shared memory (OpenMP), and GPU (CUDA).
- **Restart** particles and **Sources**.
- Different and numerous applications. (pollutant concentration, search and rescue, ...).
- C++ 17 Language.
- Designed with the computational malleability in mind.

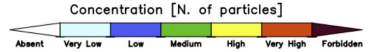
https://github.com/ccmmma/wacommplusplus

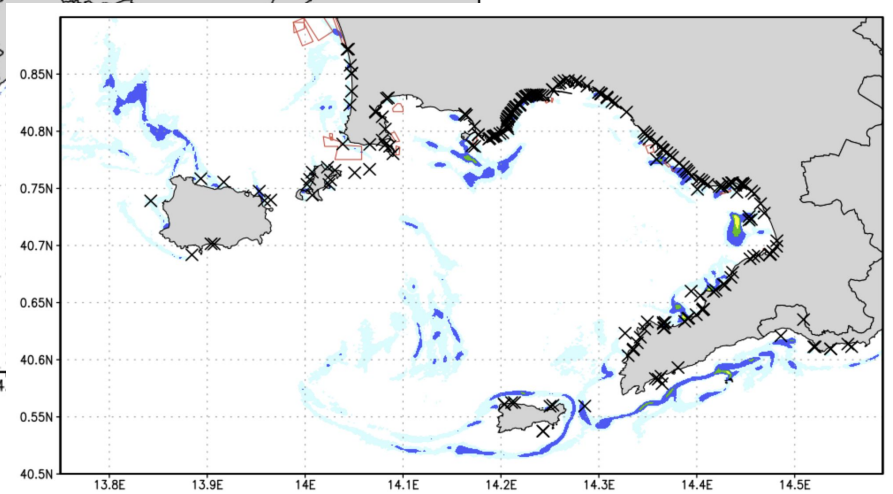
REGIONE CAMPANIA
EuroHPC Joint Undertaking
ADMIRE malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

WaComM++: Input & Output







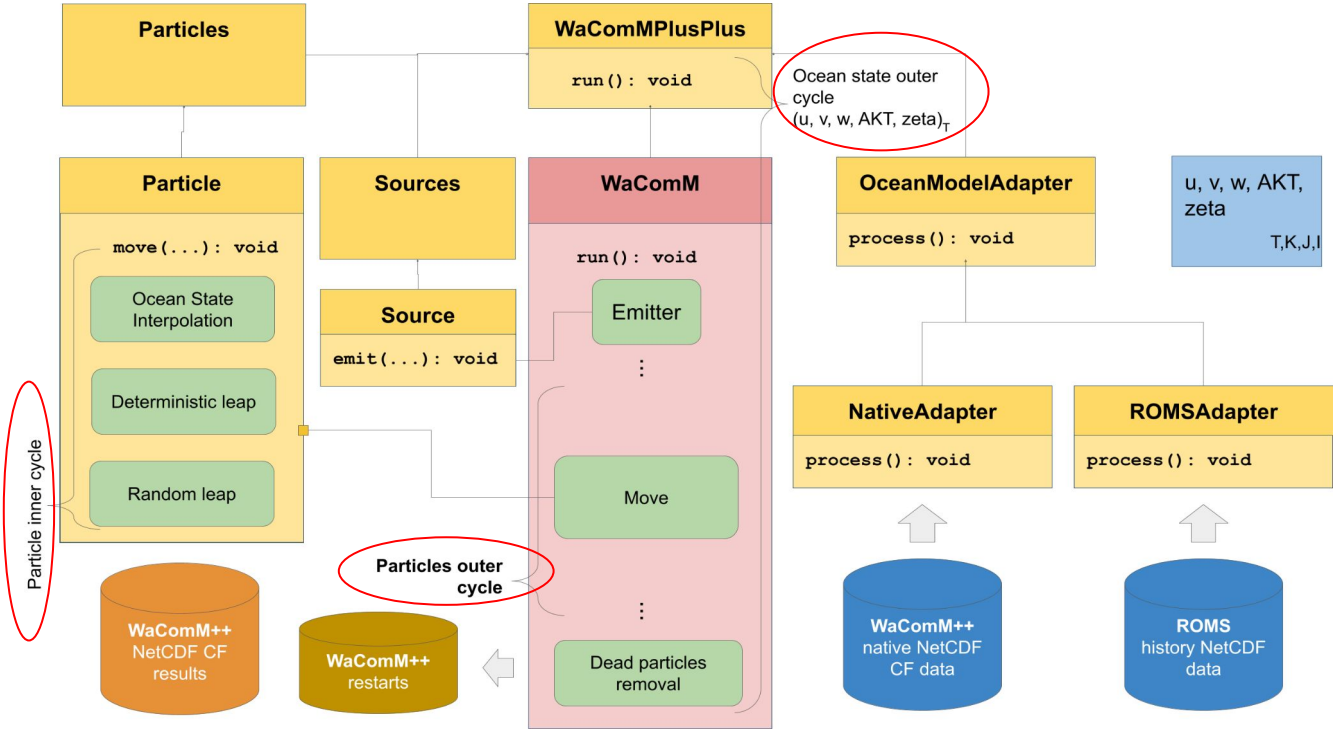


EuroHPC Joint Undertaking

ADMIRE
malleable data solutions for HPC

ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

WaComM++ architecture



The overall computation is performed over three nested cycles:

Ocean state outer cycle: for each time-referenced dataset (usually 1-hour), a WaComM component is instanced.

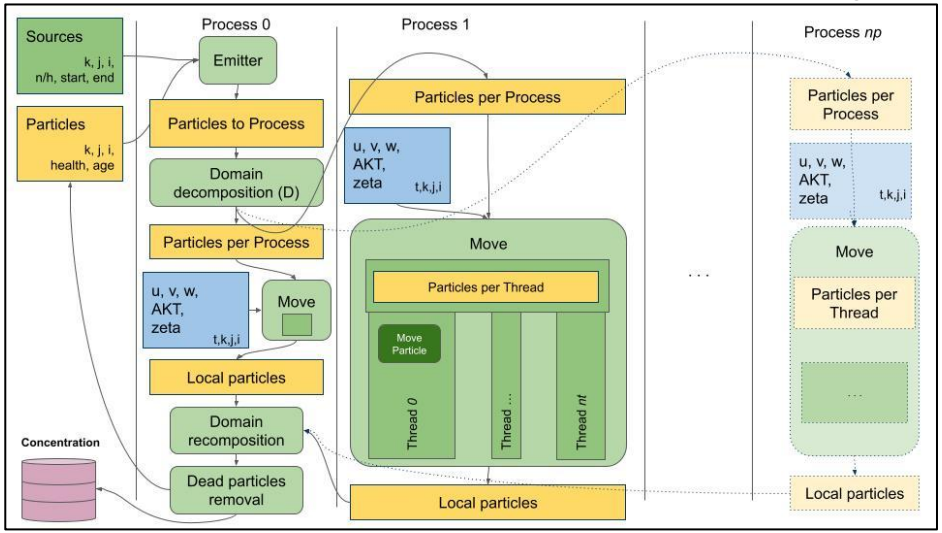
Particles outer cycle: moves the particle to process using ocean data.

Particle inner cycle: moves the particle within the considered time slice, applying the Lagrangian transport and diffusion equations integrated on a given time step.

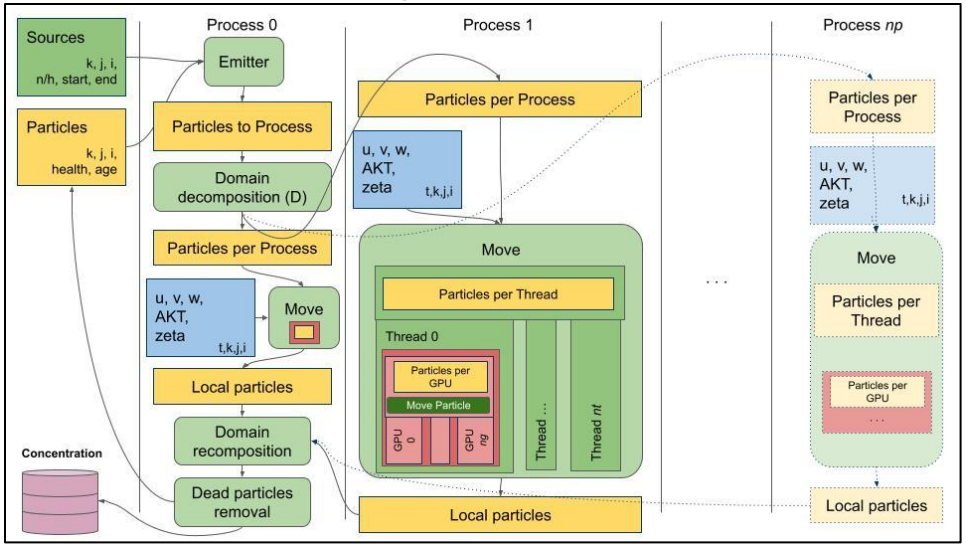
While time-dependent iterations characterize the ocean state outer cycle and the inner particle cycle, the particles' outer cycle has been hierarchically parallelized because each particle movement is independent of the others.

WaComM++ hierarchical parallelization schema

Without multi-GPU paradigm.



With multi-GPU paradigm.



EuroHPC Joint Undertaking
ADMIRE
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE
malleable data solutions for HPC

WaComM++ evaluation configuration

We use the following configurations:

- 25 million particles spilled out by a single coastal source located in the Gulf of Napoli (Campania, Italy)
- no restart mode
- 24 h of simulation
- **Different parallelization schema**

PurpleJeans (HPC Tests):
2 Intel(R) Xeon(R)Gold 5218 CPU@2.30GHz 16 cores each
4 Nvidia Tesla V100SXM232GB 5120 CUDA cores each)

https://rcf.uniparthenope.it

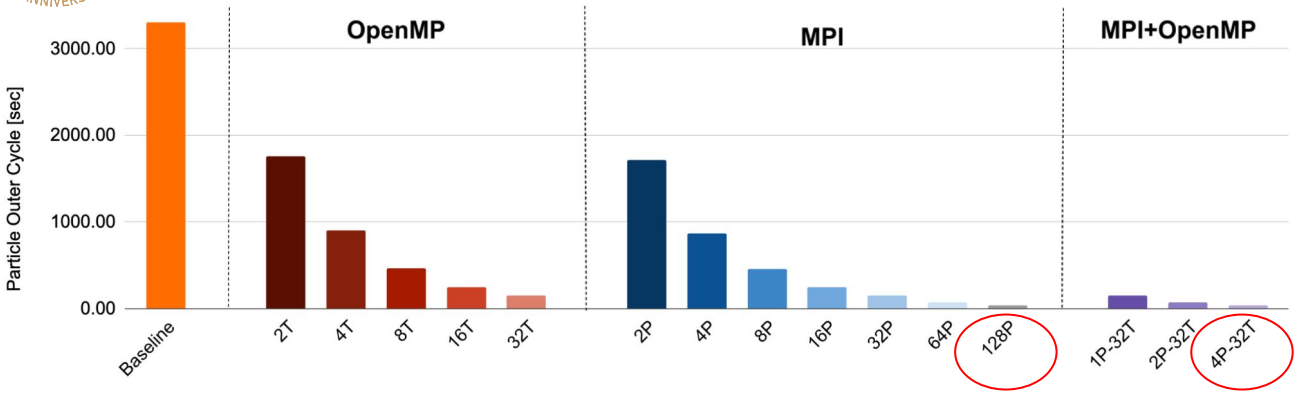
WaComM++ evaluation configuration

- **Baseline** → one process, only one thread, and no GPUs (sequential mode)

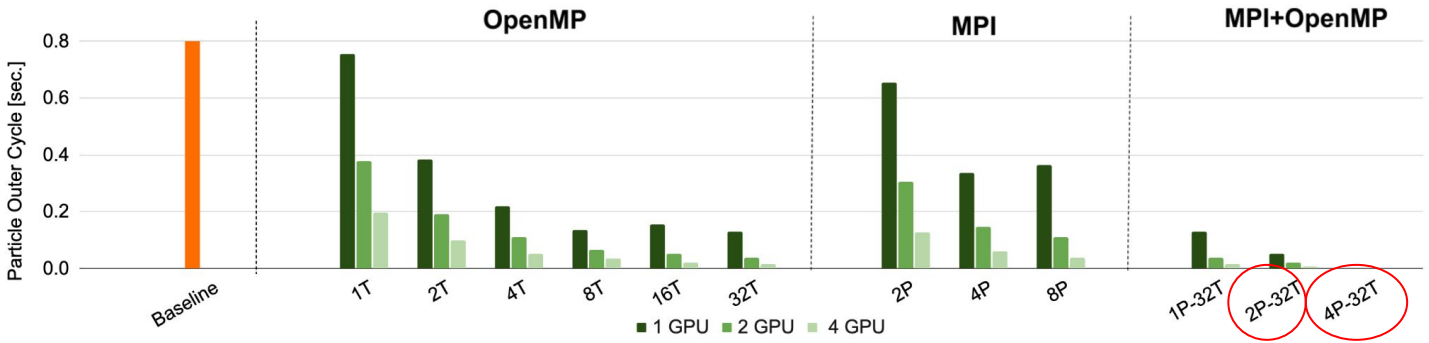
- **Distributed Memory (MPI approach)**: 1, 2, 4, 8, 16, 32, 64, and 128P on four computing nodes, considering only 1T.
- **Shared Memory (OpenMP approach)**: Considering only one MPI process, we used 1, 2, 4, 8, 16, and 32T on one computing node with 1P.
- **Shared Memory and CUDA (OpenMP-CUDA approach)**: We consider both the single GPU and the multi-GPU cases. A single process is tested from shared memory threads varying from 1 to 32T, sharing 1, 2, and 4G.
- **Distributed Memory and CUDA (MPI-CUDA approach)**: We consider both the single GPU and the multi-GPU cases. Multiple processes are tested, varying from 1 to 8P, using only one thread sharing 1, 2, and 4G.

- **Distributed Memory, Shared Memory, and CUDA (MPI-OpenMP-CUDA approach):** We consider both the single GPU and the multi-GPU cases. A multiple 1 to 4P is tested using shared memory fixed on 32T sharing 1, 2, and 4G.

WaComM++: evaluation



Particle Outer-Cycle execution times results without CUDA approach

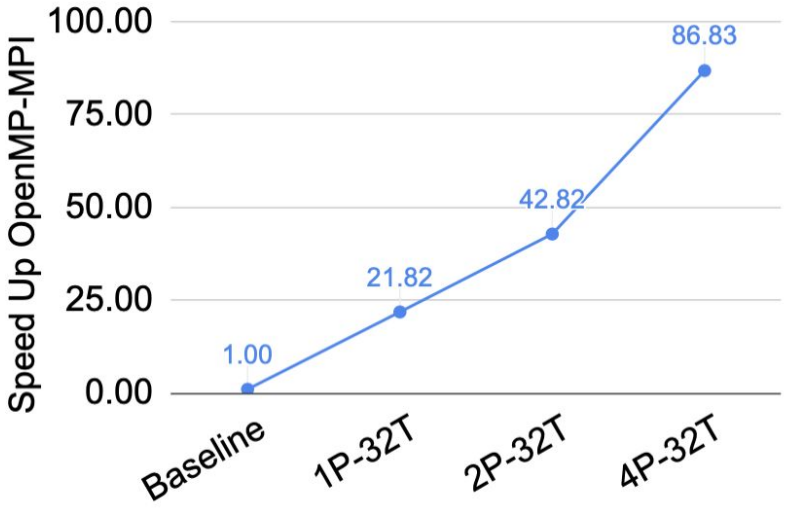


Particle Outer-Cycle execution times results using CUDA approach.

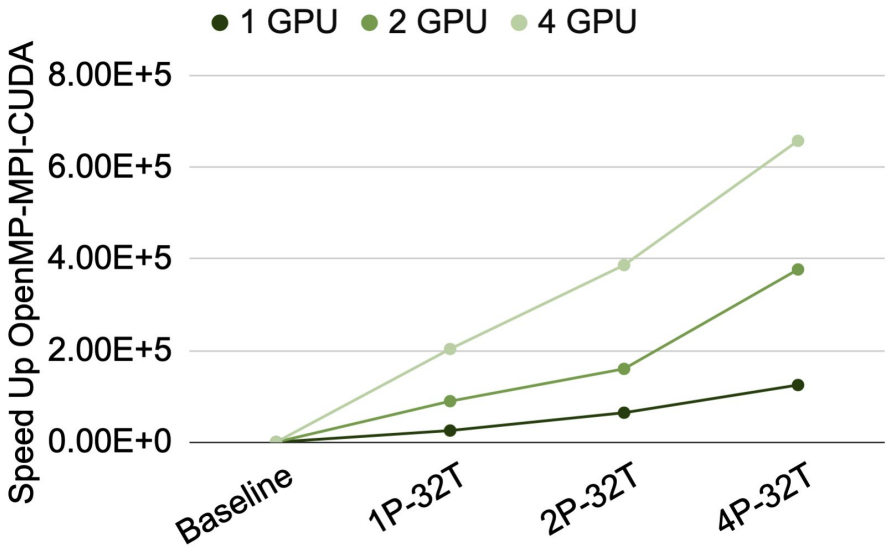
"P" is the number of MPI processes, "T" is the number of OpenMP processes, and "G" identifies the number of GPU devices considered for the computation.

EuroHPC Joint Undertaking
ADMIRE
malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

WaComM++: evaluation



Speed up of the OpenMP-MPI approach

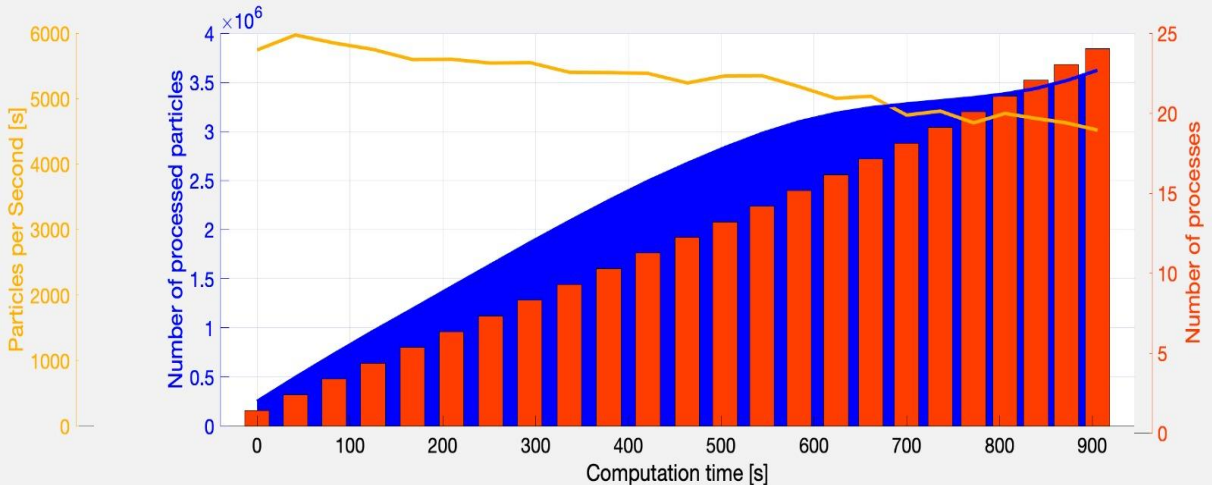


Speedup of the OpenMP-MPI-CUDA approach

Computation Malleability

- Within the ADMIRE project, WaComM++ has been enhanced using Flex-MPI for computational malleability.

- Computational resources vary during the computation to:



  - increase the number of processor as the problem size grows
  - sustain a constant performance in terms of computed particles per second

- Simulation of 24 hours of 250K particles emitted per hour from a single emission point.
- The problem size is distributed on a new spawn process each simulated hour.

WaComM++

- Is an open-source Lagrangian for modelling pollutants transport and diffusion at the sea.

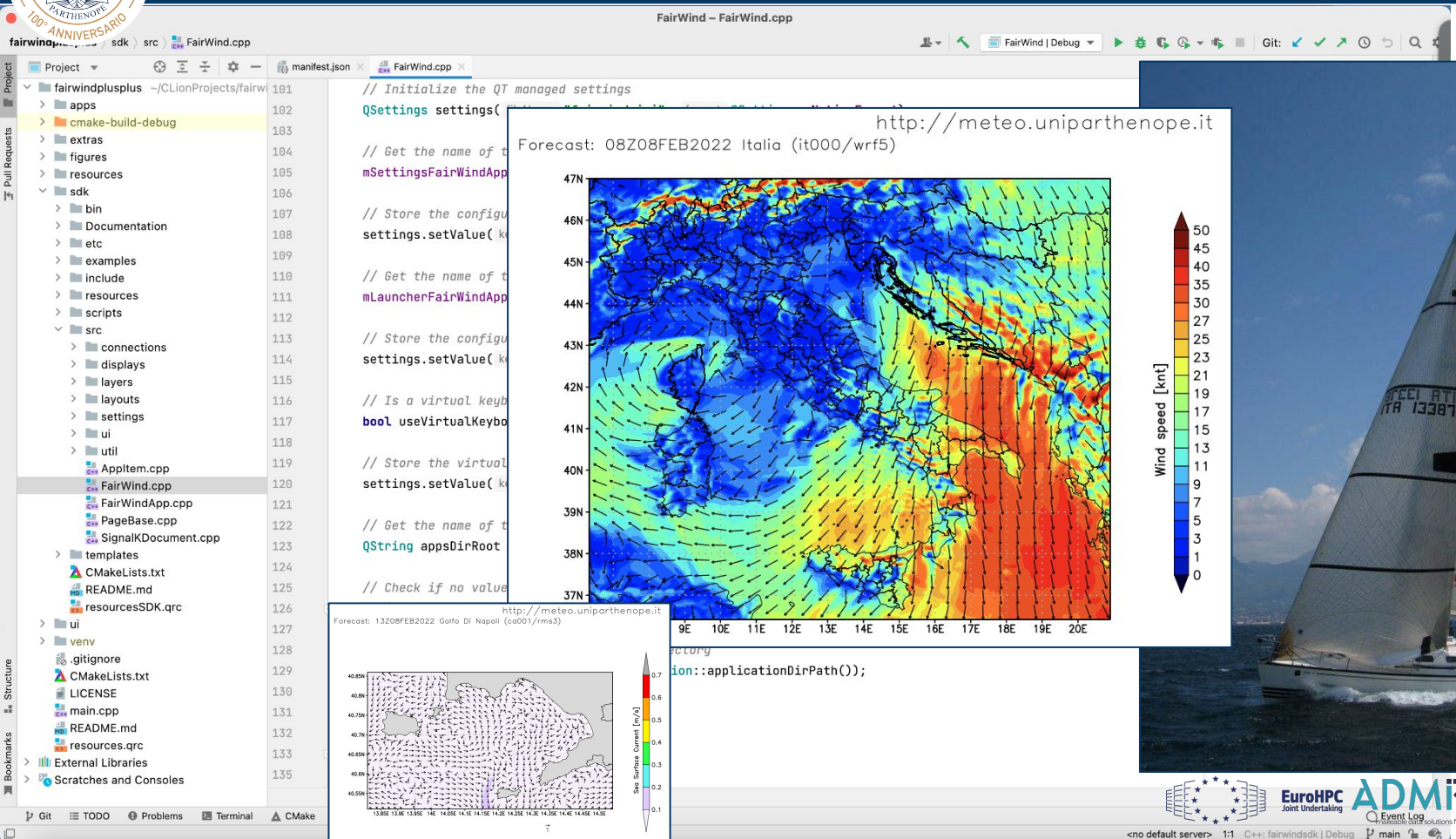
- It uses a parallelization schema enabling the users to run it on heterogeneous parallel architectures.

- Using distributed memory, shared memory, and multi GPUs parallelism, WaComM++ performed more than 650.000 times faster than the sequential case.

Scientists do love mix things

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
EuroHPC Joint Undertaking
ADMIRE
malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

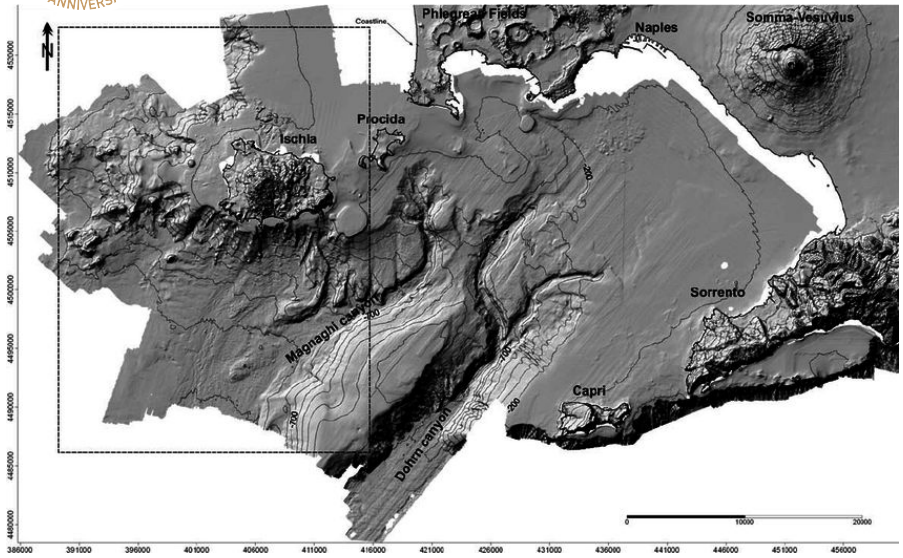
Scientists do love mix things

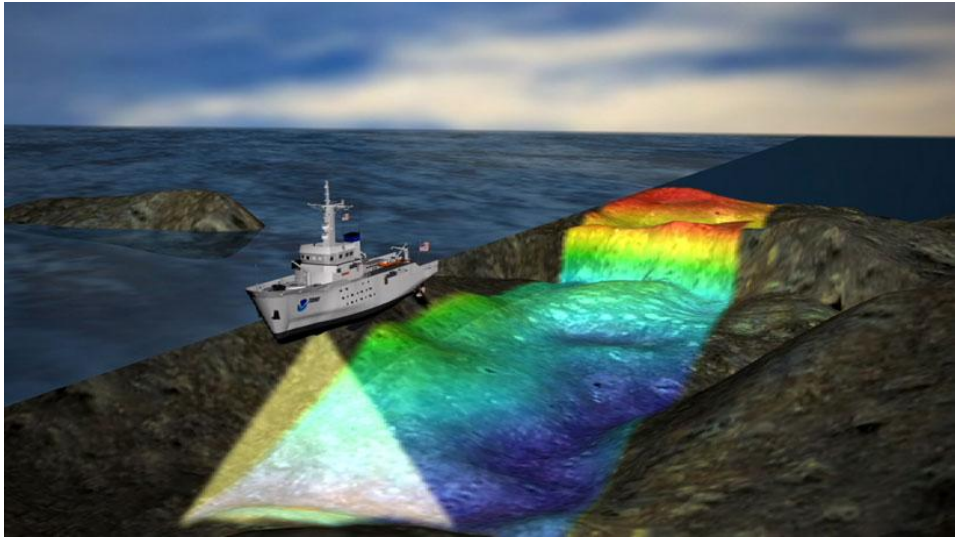


EuroHPC Joint Undertaking

ADMIRE — ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

Introduction, contextualization and motivations





Bathymetry: "*the study of underwater depth of lake or ocean floors. In other words, bathymetry is the underwater equivalent to hypsometry or topography.*"

Echosounder: a device using echo sounds from the sea bottom to evaluate the depth.

Multibeam: an echo sounder with steroids.

EuroHPC Joint Undertaking
ADMIRE malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

Hard to obtain, lack of public high resolution open data.

Citizen Science

ADMIRE
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE
malleable data solutions for HPC

Introduction, contextualization and motivations

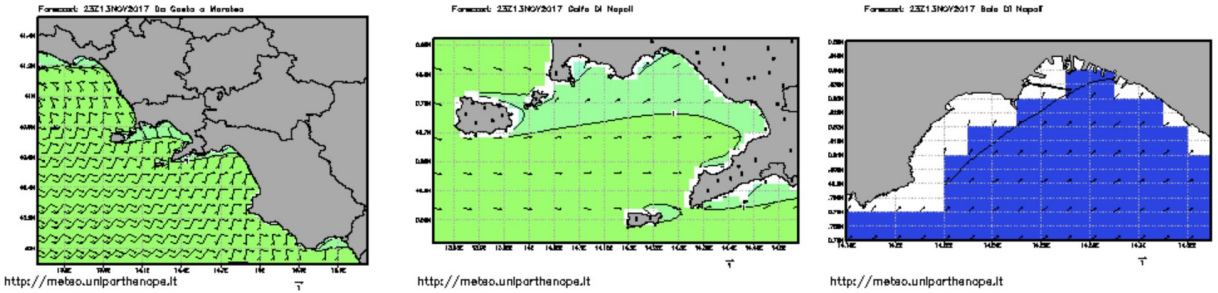
- Huge amount of surface data.
- Problem size characterized by a remarkable variability.
- Data have to be continuously updated.

- Internet of things based crowdsourcing tools.
- Using an ad-hoc IoT Data Transfer Protocol.
- Poor and intermittent data connection availability.

Introduction, contextualization and motivations

Wind driven wave model

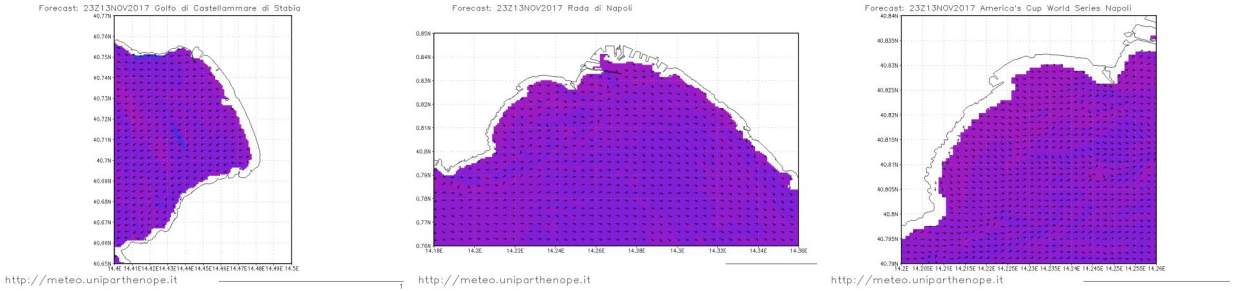
Wind-driven sea waves: Wave Watch III (WW3) max ground resolution 1Km, outlook 144h, updated each 24 hours



WaveWatch III

Surface currents: Regional Ocean Model System (ROMS) max ground resolution 80m, outlook 120h, updated each 24 hours

Current model



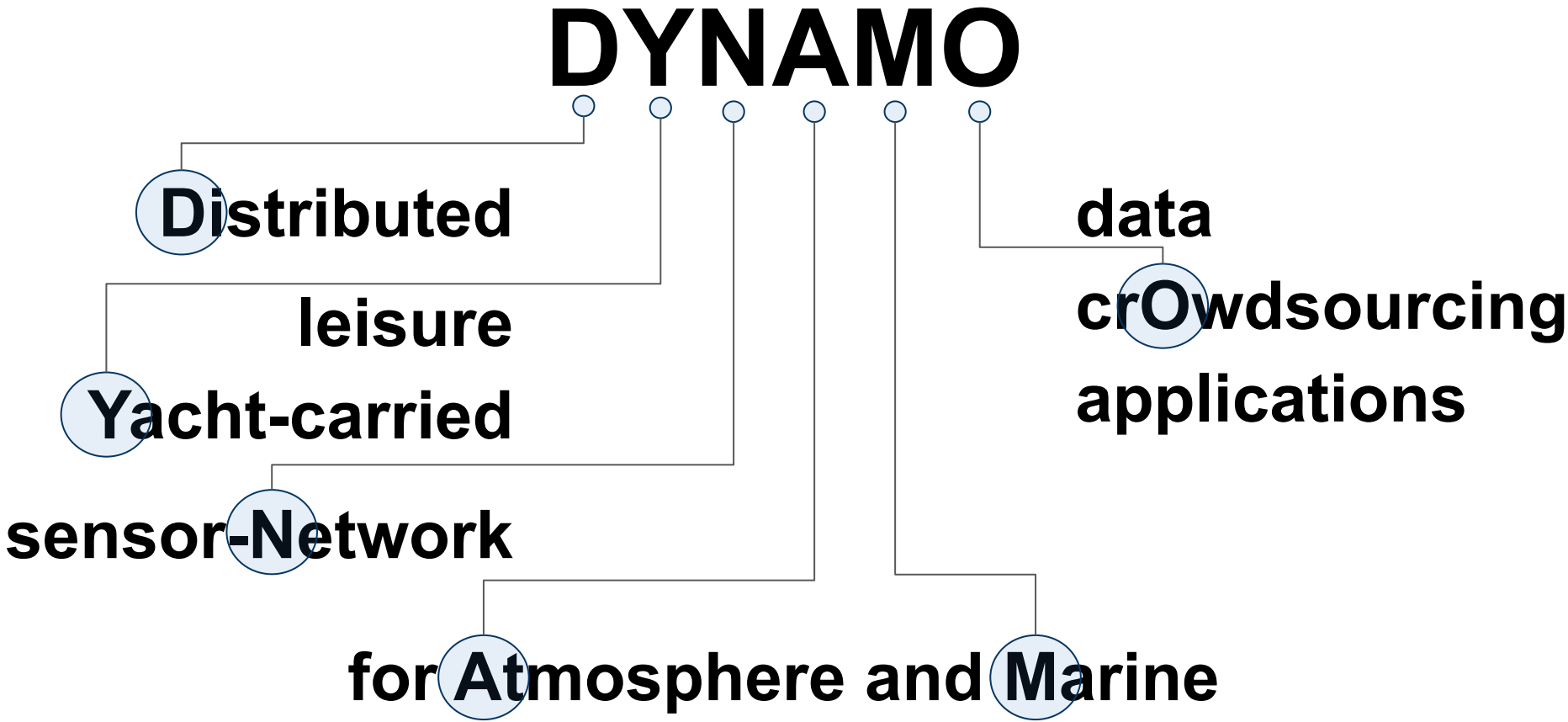
Regional Ocean Model System

We need for more accurate models for environment management.

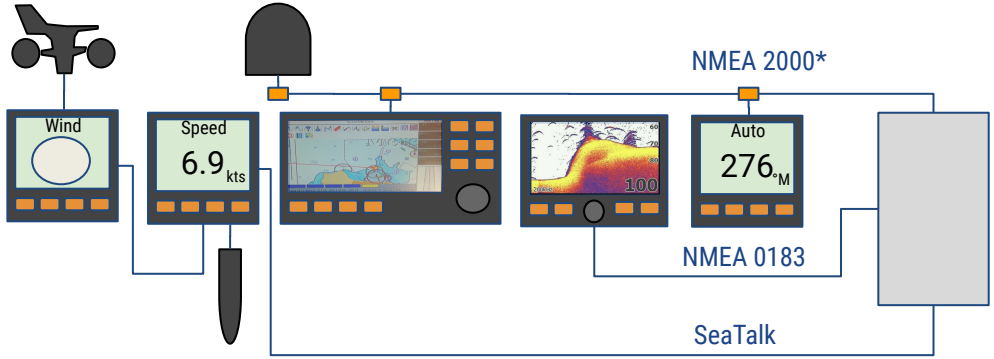
Introducing the DYNAMO ecosystem

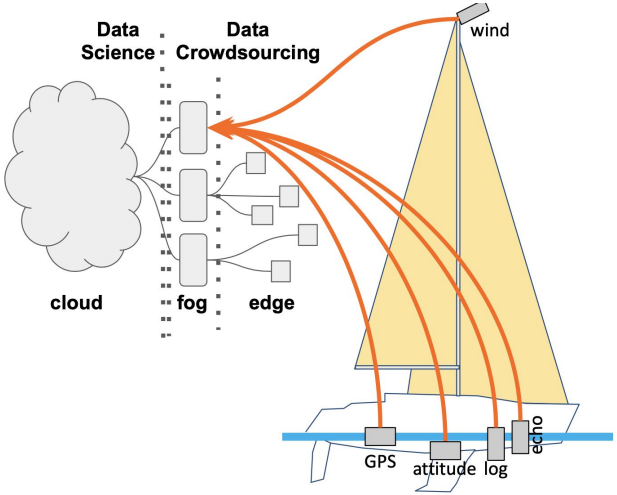
GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
EuroHPC Joint Undertaking
ADMIRE
malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

What's in a name



Leisure Yachts-carried sensor Network





- Today, almost any kind of leisure boat is equipped by sensor connected by a local network.
- Sensors collect various data (environment, navigation).
- Data are shown by displays and gauges… and then dropped away…

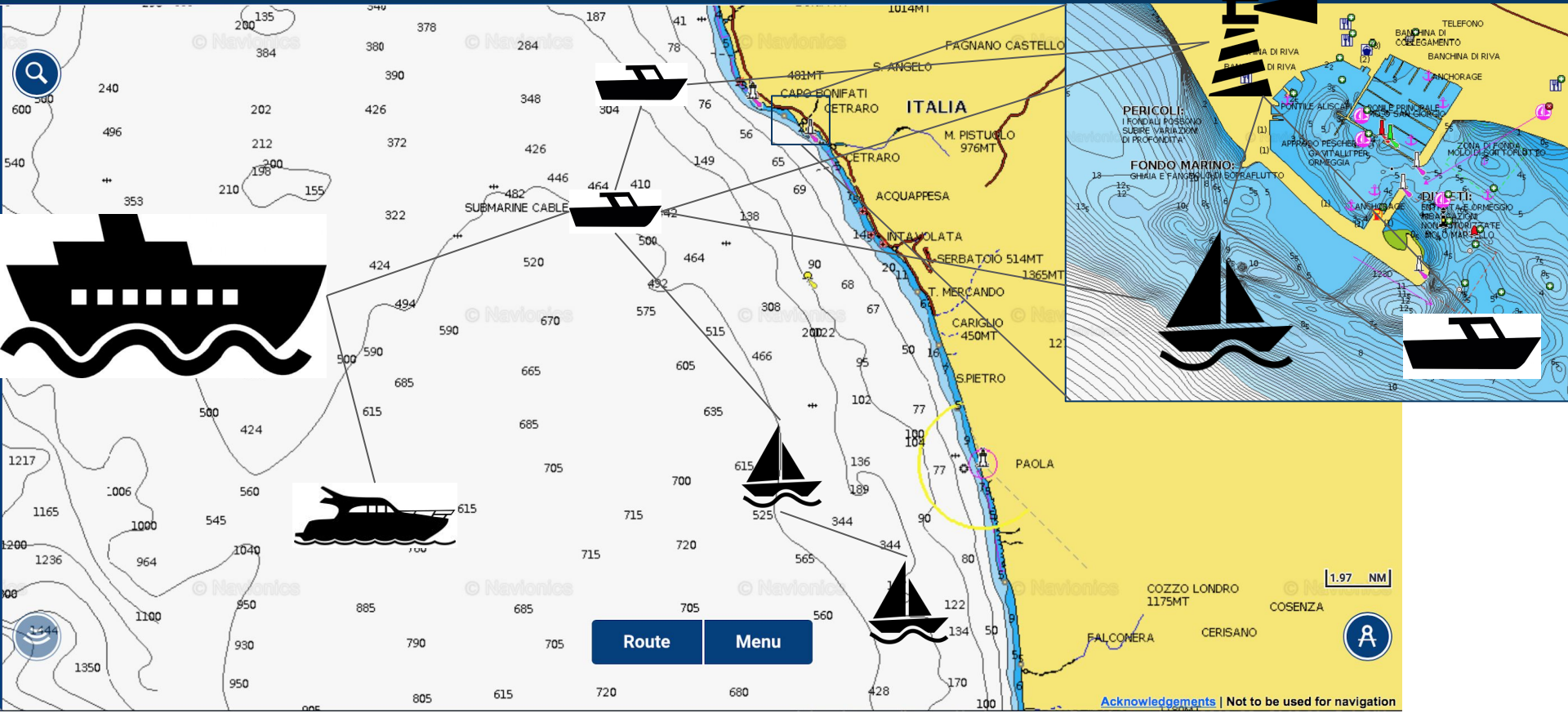
Data are gold coins in the realm of the scientists.

Our Vision

Using leisure boats as “weather stations” connecting all the sensors in a distributed instrument.

EuroHPC Joint Undertaking ADMIRE malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

The Internet of Floating Things



Coastal and open water data **crowdsourcing**

EuroHPC Joint Undertaking ADMIRE malleable data solutions for HPC

ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

Data crowdsourcing at the sea: big picture

Leisure boats and yachts are equipped with marine data sensor networks.
Position, Attitude, Speed, Wind, **Depth**, ...

Acquisition / Consuming final products

Marine electronics data protocols:

NMEA0183, SeaTalk, NMEA2000, SignalK

High Performance [Cloud] Computing

Weather / Ocean Forecasts

A.I. improved Weather / Ocean predictions

Already available dataset

1. **Sensor Calibration**
2. **Tide Correction**
3. **Anonymization**
4. **Harmonization**
5. **A.I. training**
6. **Forecast improvement**
7. **Publishing & sharing**

New dataset

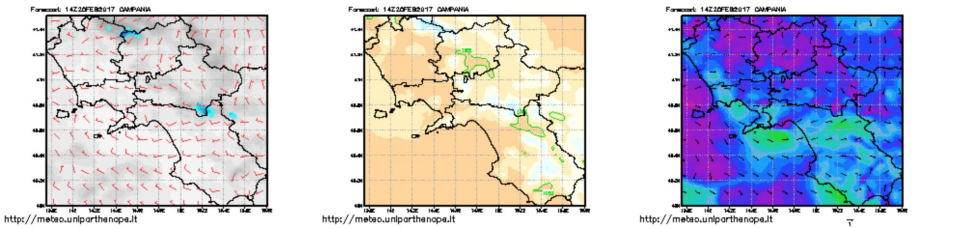
EuroHPC Joint Undertaking

ADMIRE
malleable data solutions for HPC

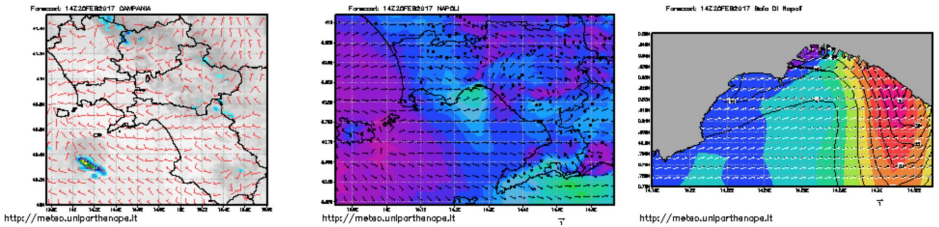
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

Science SDK

Weather forecast: Weather Research and Forecast (WRF) max ground resolution 3Km, outlook 144h, updated each 6 hours



Weather forecast: Weather Research and Forecast (WRF) max ground resolution 1Km, outlook: 24h, updated each 6h



Wind forecast: CALMET, max ground resolution 250m, outlook: 24h, updated each 6h

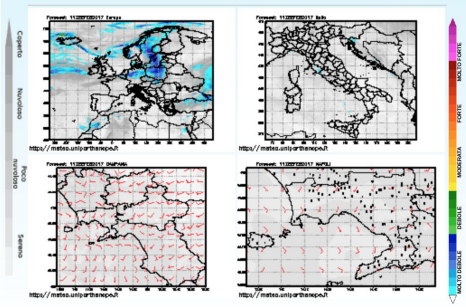
Using Artificial Intelligence trained with crowdsourced data to improve forecasting models' results.

meteo uniparthenope

ANIMATION | +3H | +6H | +9H | +12H | +18H | +24H | +36H | +48H

Weather forecast with mesh of 3 Km - 10m sea surface wind - cloudiness, hourly cumulative rainfall and temperature each 3 hours

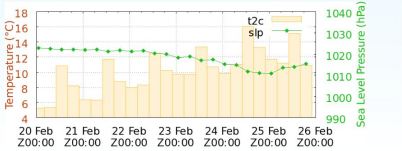
the grey scale on the left is the cloudiness, the blu scale is the rainfall, and the red vectors are the winds.



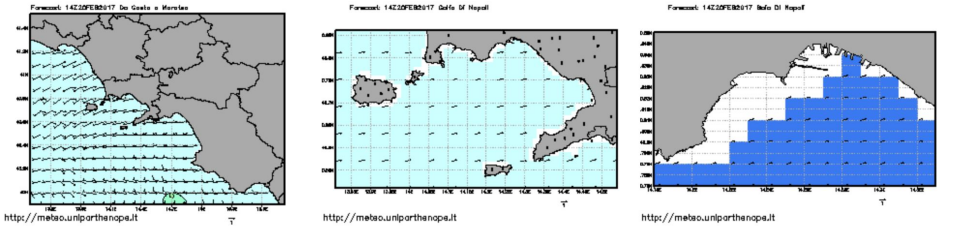
Meteo Comune di Napoli

Previsione	T min	T max	Vento	Pioggia
Lunedì 20	5°C	12°C	NNW 2 knt	-
Martedì 21	6°C	12°C	WNW 1 knt	-
Mercoledì 22	9°C	13°C	SSE 4 knt	-
Giovedì 23	9°C	14°C	SSW 4 knt	0.2 mm
Venerdì 24	11°C	16°C	SSE 9 knt	0.2 mm
Sabato 25	9°C	14°C	SSW 4 knt	-

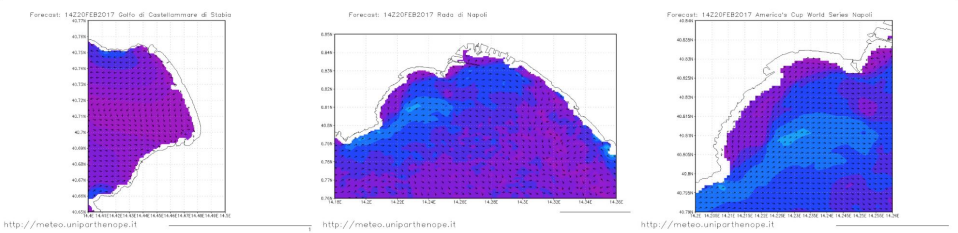
Trend of temperature and pressure for next 6 days



Wind-driven sea waves: Wave Watch III (WW3) max ground resolution 1Km, outlook 144h, updated each 24 hours



Surface currents: Regional Ocean Model System (ROMS) max ground resolution 80m, outlook 120h, updated each 24 hours

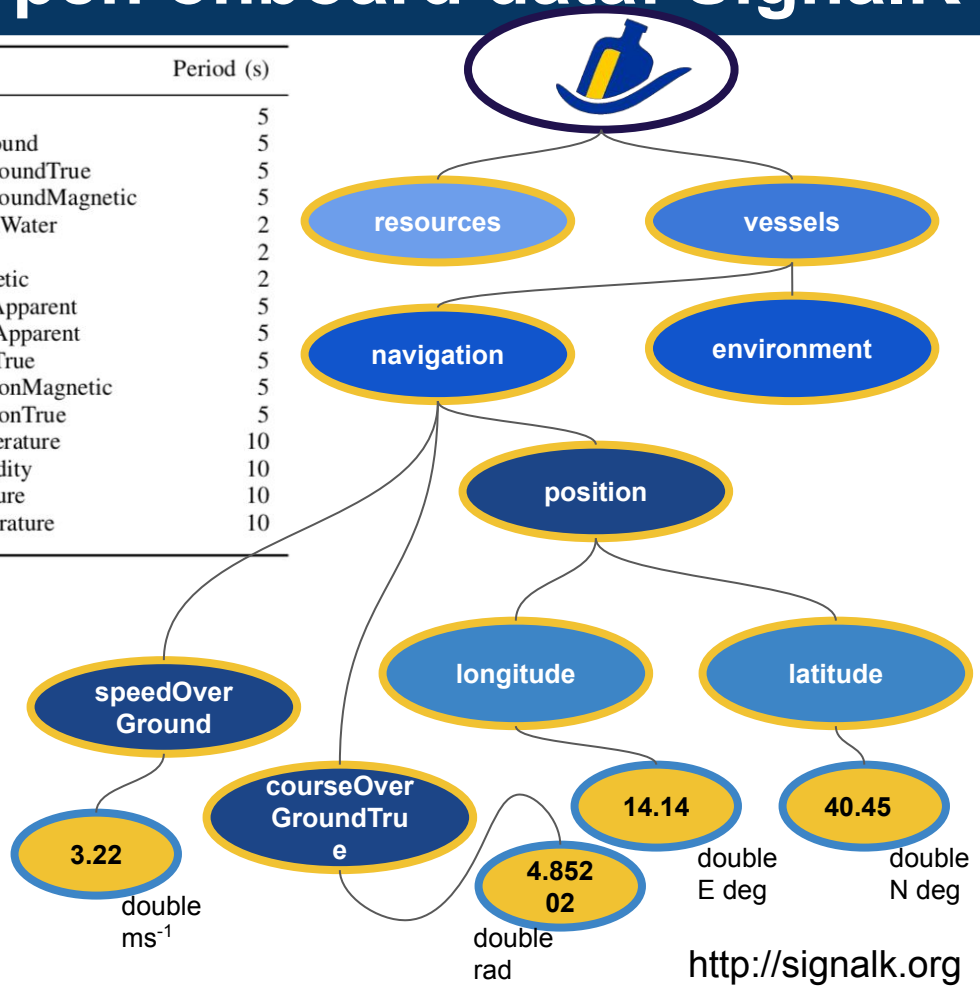


Open onboard data: SignalK


```
{
  "version": "v1.0.0",
  "self": "urn:mrn:signalk:uuid:705f5f1a-efaf-44aa-9cb8-a0fd6305567c",
  "vessels": {
    "urn:mrn:signalk:uuid:705f5f1a-efaf-44aa-9cb8-a0fd6305567c": {
      "navigation": {
        "speedOverGround": {
          "value": 4.32693662,
          "$source": "ttyUSB0.GP",
          "sentence": "RMC",
          "timestamp": "2017-05-16T05:15:50.007Z"
        },
        "position": {
          "value": {
            "altitude": 0.0,
            "latitude": 37.81479,
            "longitude": -122.44880152
          },
```


Full Document

SignalK	Period (s)
navigation.position	5
navigation.speedOverGround	5
navigation.courseOverGroundTrue	5
navigation.courseOverGroundMagnetic	5
navigation.speedThroughWater	2
navigation.headingTrue	2
navigation.headingMagnetic	2
environment.wind.angleApparent	5
environment.wind.speedApparent	5
environment.wind.speedTrue	5
environment.wind.directionMagnetic	5
environment.wind.directionTrue	5
environment.inside.temperature	10
environment.inside.humidity	10
environment.inside.pressure	10
environment.water.temperature	10



- Modern and open data format for marine use.
- Internet friendly standard built on common web technologies, such as JSON and WebSockets.

http://signalk.org

Open onboard data: SignalK


```
{
  "version": "v1.0.0",
  "self": "urn:mrn:signalk:uuid:705f5f1a-efaf-44aa-9cb8-a0fd6305567c",
  "vessels": {
    "urn:mrn:signalk:uuid:705f5f1a-efaf-44aa-9cb8-a0fd6305567c": {
      "navigation": {
        "speedOverGround": {
          "value": 4.32693662,
          "$source": "ttyUSB0.GP",
          "sentence": "RMC",
          "timestamp": "2017-05-16T05:15:50.007Z"
        },
        "position": {
          "value": {
            "altitude": 0.0,
            "latitude": 37.81479,
            "longitude": -122.44880152
          },
```


Full Document

SignalK	Period (s)
navigation.position	5
navigation.speedOverGround	5
navigation.courseOverGroundTrue	5
navigation.courseOverGroundMagnetic	5
navigation.speedThroughWater	2
navigation.headingTrue	2
navigation.headingMagnetic	2
environment.wind.angleApparent	5
environment.wind.speedApparent	5
environment.wind.speedTrue	5
environment.wind.directionMagnetic	5
environment.wind.directionTrue	5
environment.inside.temperature	10
environment.inside.humidity	10
environment.inside.pressure	10
environment.water.temperature	10


```
{
  "context": "vessels.urn:mrn:imo:mmsi:234567890",
  "updates": [{
    "source": {
      "label": "N2000-01",
      "type": "NMEA2000",
      "src": "017",
      "pgn": 127488
    },
    "timestamp": "2010-01-07T07:18:44Z",
    "values": [{
      "path": "propulsion.0.revolutions",
      "value": 16.341667
    }, {
      "path": "propulsion.0.boostPressure",
      "value": 45500.0
    }]
  }]
}
```


Delta Document

- Signal K is developed in the open with help from the marine community.
- Free and Open Source software.
- Apache License, Version 2.0.
- Accepted as a standard for the marine electronics community

https://github.com/signalk

DYNAMO & FairWind: science and business

If you are a boat builder

If you are a cruiser

If you are a powerboat lover

If you are a scientist

If you are a yacht club

If you are an sporting/amateur fisherman

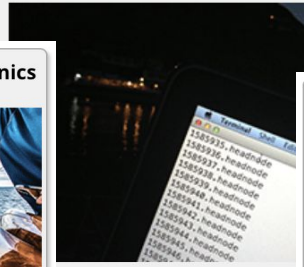
It you are a charter

If you are a mooring or docking company

If you are a racer!

If you build marine electronics

If you are software developer



If you work in education

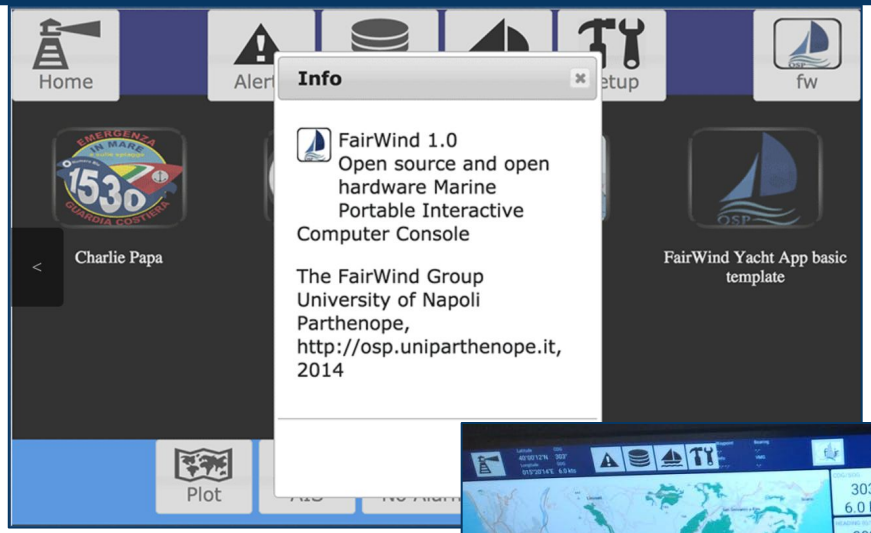
- Designed to be attractive for boat lovers, business companies and, of course, to scientists.
- Open-source with Apache 2.0 “by design”.

EuroHPC Joint Undertaking
ADMIRE
malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

DYNAMO & FairWind: Evolution and Trends

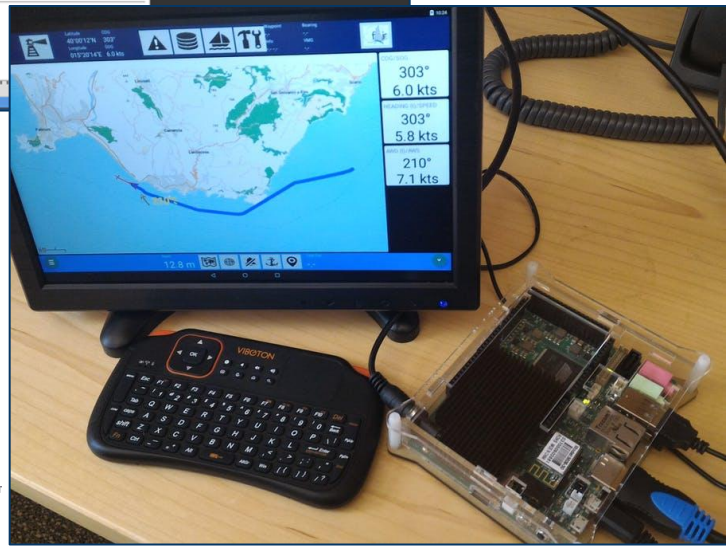
The origins

- 2010: open-source was almost not already used in marine electronics.
- 2012: Open Sailing Processor (OSP): fully HTML, CSS, Javascript, and Python.



The startup era

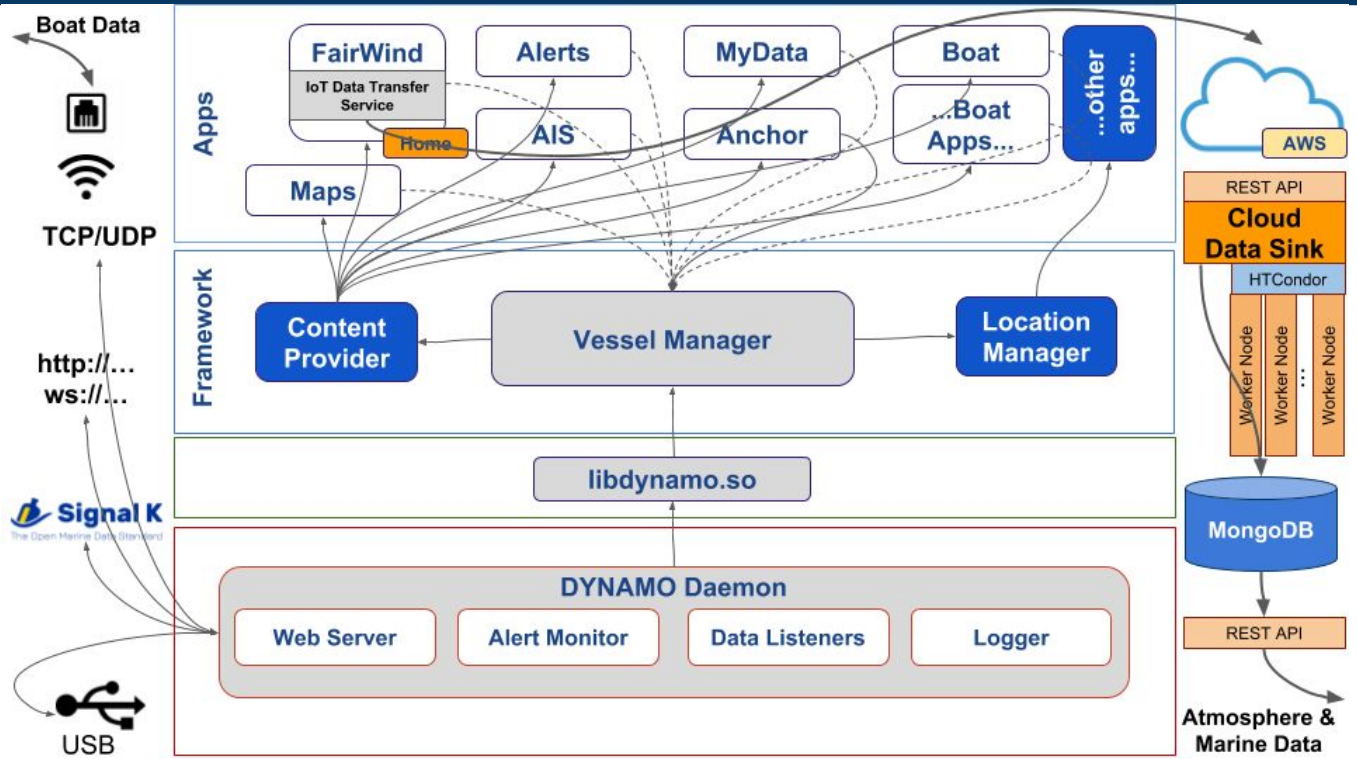
- 2014: OSP is rebranded FairWind.
- 2015: investors suggests to switch to closed-source, developed as an Android Launcher.



EuroHPC Joint Undertaking ADMIRE malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

DYNAMO & FairWind: Evolution and Trends

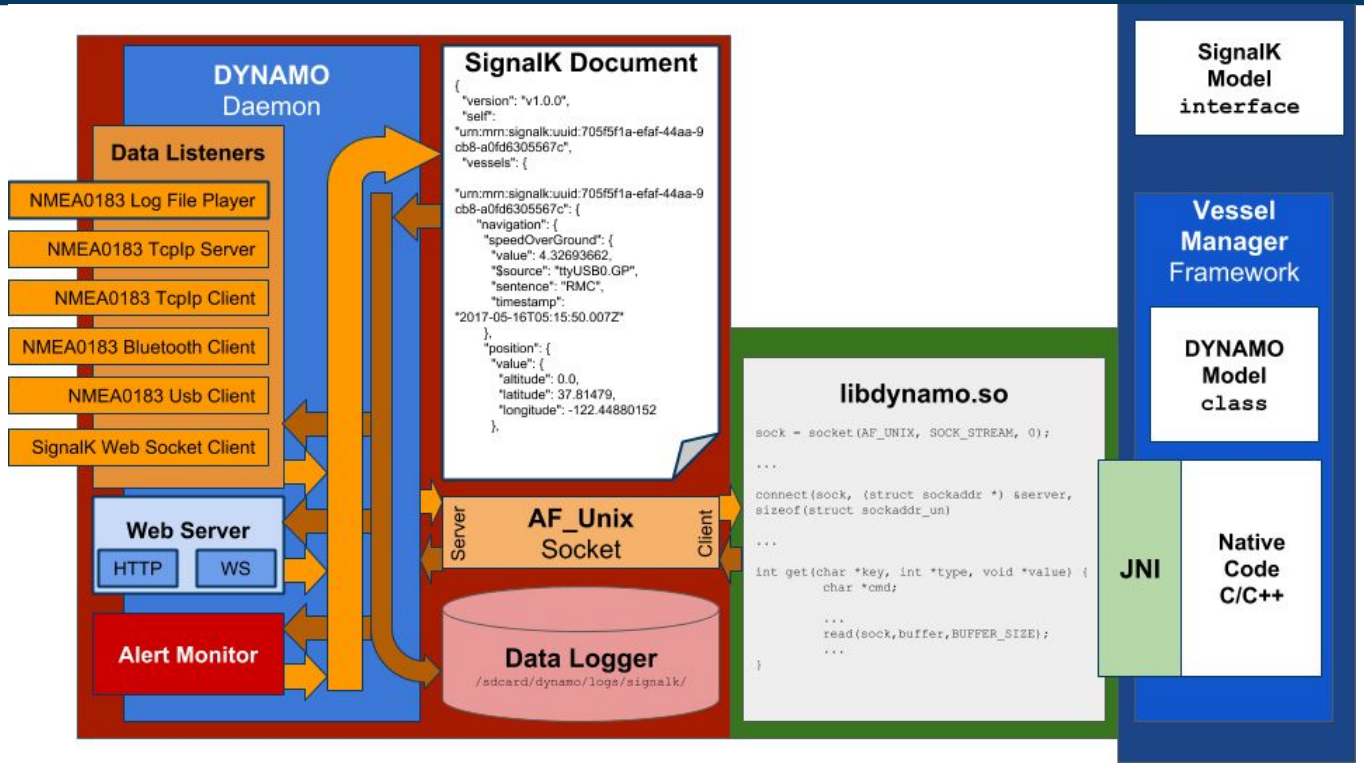
Remember the open source? It is back.



- 2018: UNP assigns a grant to the DYNAMO project to support the development.
- FairWind moves to a customized open-source Android.
- FairWind is open-source.

DYNAMO & FairWind: Evolution and Trends

Remember the open source? It is back.



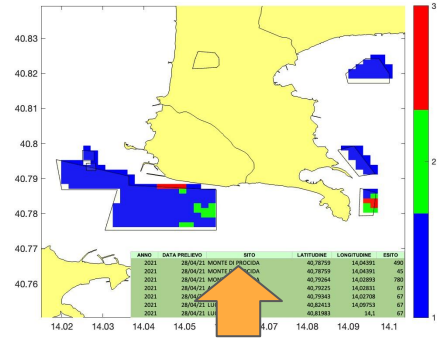
- 2018: UNP assigns a grant to the DYNAMO project to support the development.
- FairWind moves to a customized open-source Android.
- FairWind is open-source.

DYNAMO & Fairwind: Fundings

REGIONE CAMPANIA

Research Grants funded by the Regione Campania - Veterinary Sector

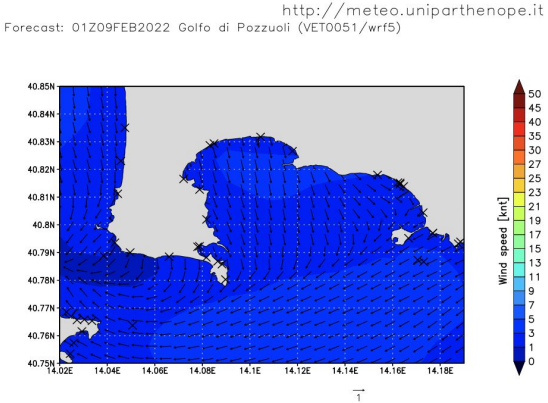
- **MytiluSE** - **M**odelling m**ytilus** farming **S**ystem with **E**nhanced web technologies
- **MytilAI** - **M**odelling **mytil**us farming with **A**rtificial **I**ntelligence technologies



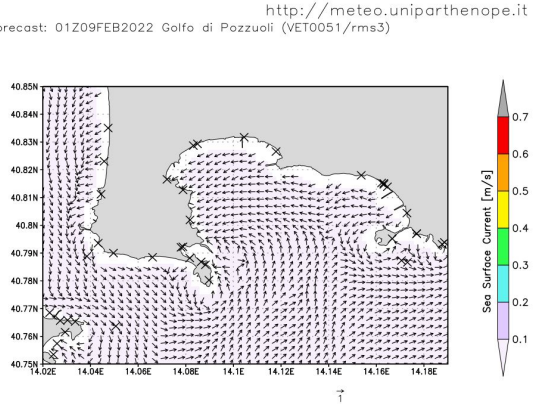
Mussel Contamination Prediction

A.I. - D.N.N.

High Resolution Seafloor DTM

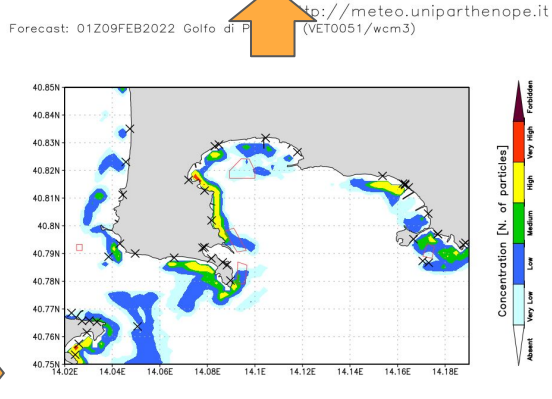


Weather Forecasts



Sea Currents Forecasts

WaComM++



Transport and Diffusion Forecasts

Moving data from vessels to shore

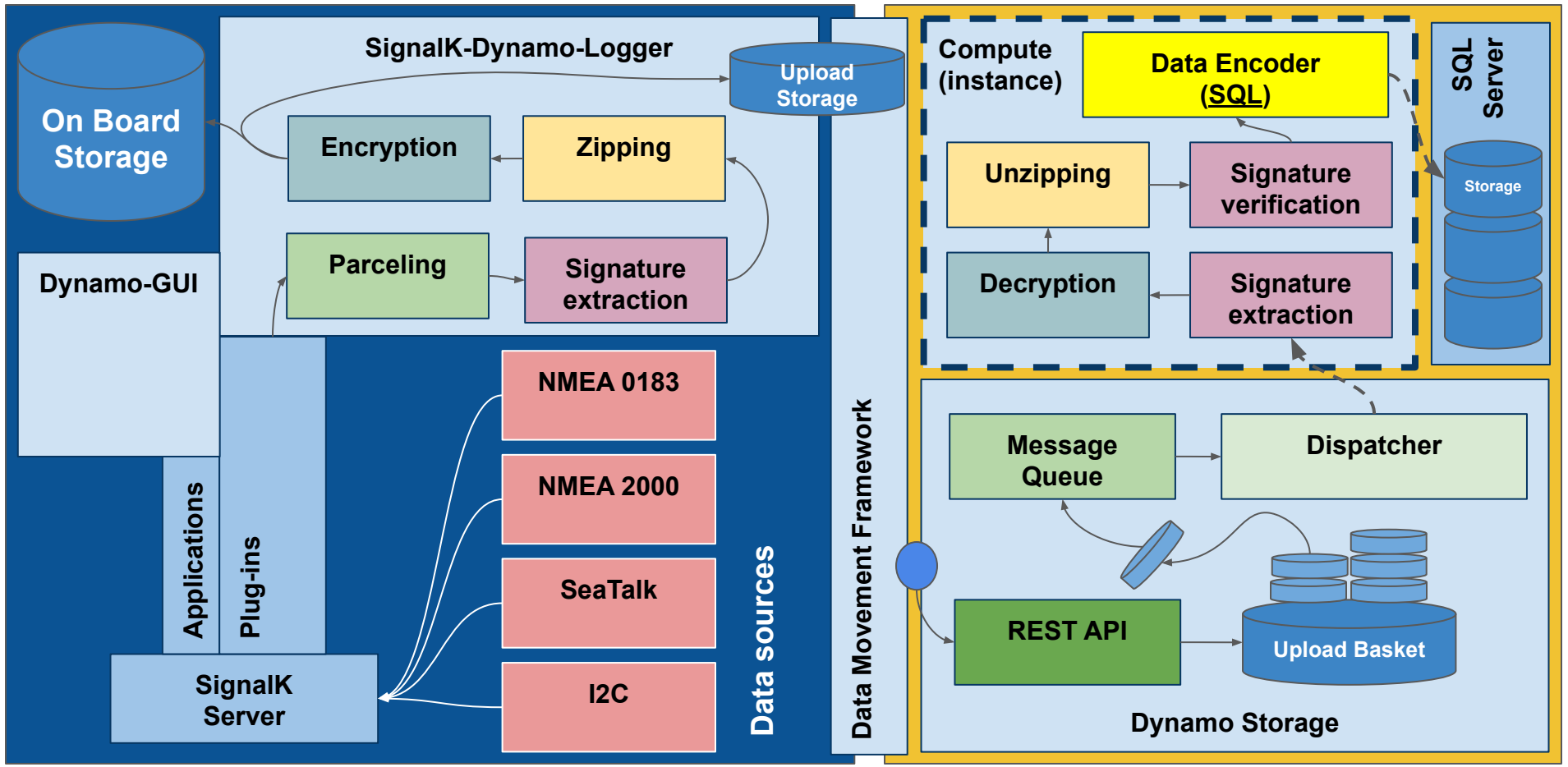
GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
EuroHPC Joint Undertaking
ADMIRE
malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

DYNAMO: Data Movement Framework

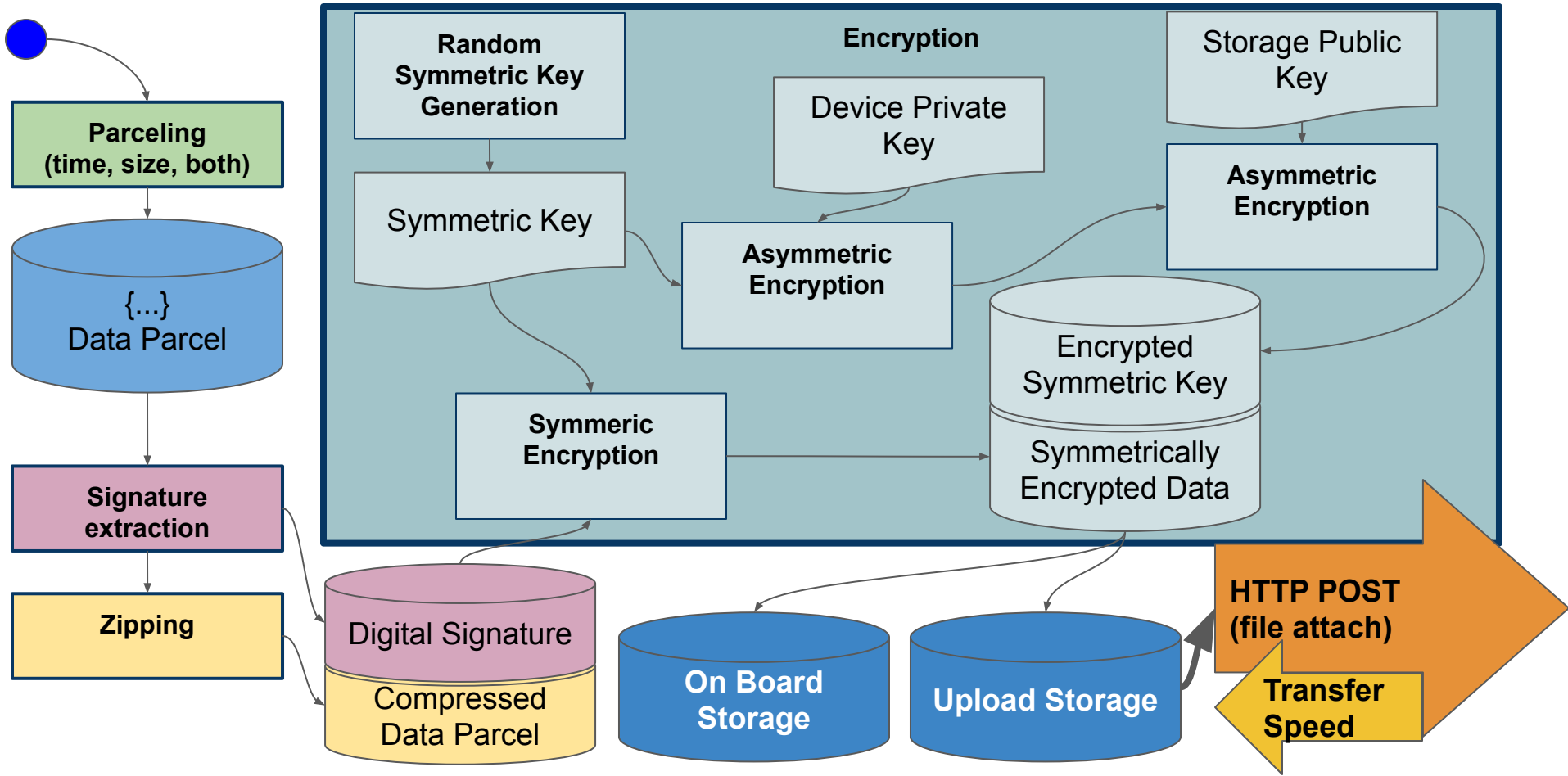
- **Resilient** because it takes advantage of short connection times using a dynamic thread generation algorithm;
- **Lightweight**, because data parcels could be compressed;
- **HTTP-based**, mainly using GET and POST with file attach verbs;
- **Bidirectional**, but loosely coupled;
- **Firewall** and **proxy friendly**, enforcing the security without a mandatory use of HTTPS;
- **Application independent,** could be used in different contexts;
- **Fully customizable**, because each feature could be deactivated, depending on the application design-

Secure thanks to encryption and data signature

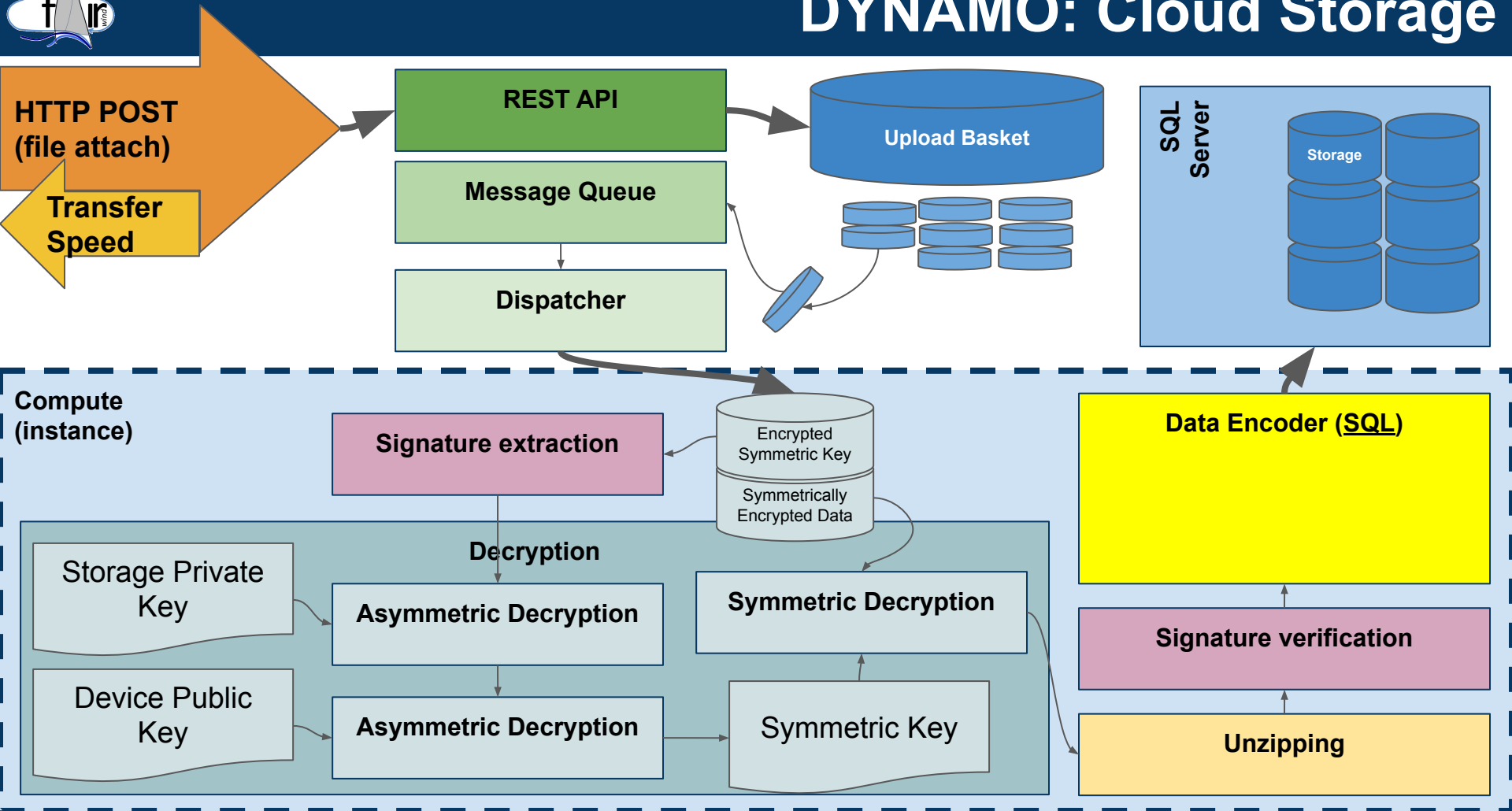
DYNAMO: the data movement prospective



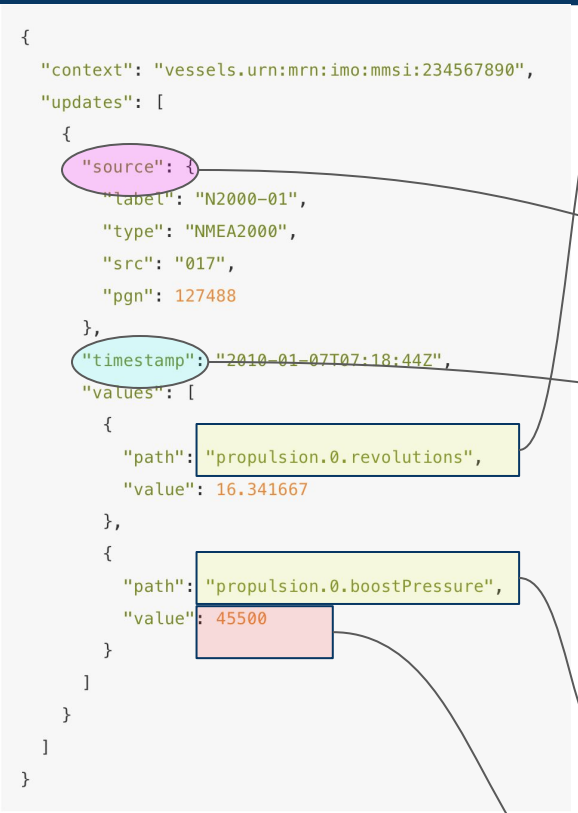
DYNAMO: SignalK Logger



DYNAMO: Cloud Storage



SQL Data Encoder



- SQL schema

propulsion_0_revolutions
context
timestamp
source
value

navigation_position
context
timestamp
source
value

name
context
timestamp
source
value

navigation.speedOverGround
context
timestamp
source
value

contexts
value

environment.depthBelowTransducer
context
timestamp
source
value

propulsion_0_boostPressure
context
timestamp
source
value

DYNAMO (small production, ~20 pcs)

Headless

- Raspberry Pi 3B+
- Power management
- 2nd Wifi Adapter
- 4G Cellular Dongle

- GPS/Glonass 5Hz
- Barometer/Hygrometer/Thermometer
- Pitch/Yaw/Roll

- NMEA 0183
- NMEA 2000
- SeaTalk
- I2C

7” Touch

Publications and Awards

- Montella, Raffaele, Sokol Kosta, and Ian Foster. "**DYNAMO: Distributed leisure yacht-carried sensor-network for atmosphere and marine data crowdsourcing applications.**" In *2018 IEEE International Conference on Cloud Engineering (IC2E)*, pp. 333-339. IEEE, 2018.

- Montella, Raffaele, Mario Ruggieri, and Sokol Kosta. "**A fast, secure, reliable, and resilient data transfer framework for pervasive IoT applications.**" In IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 710-715. IEEE, 2018.

- Montella, Raffaele, Diana Di Luccio, Sokol Kosta, Giulio Giunta, and Ian Foster. "**Performance, resilience, and security in moving data from the fog to the cloud: the DYNAMO transfer framework approach.**" In International Conference on Internet and Distributed Computing Systems, pp. 197-208. Springer, Cham, 2018.

Best Paper Award

- Di Luccio, Diana, Sokol Kosta, Aniello Castiglione, Antonio Maratea, and Raffaele Montella. "**Vessel to shore data movement through the internet of floating things: A microservice platform at the edge.**" Concurrency and Computation: Practice and Experience 33, no. 4 (2021): e5988.

DYNAMO & FairWind: Evolution and Trends

- Using a customized version of Android in the DYNAMO ecosystem is challenging, but demanding.
- The original idea of providing regular Android applications with data from boats works in the lab but not for production.
- Laws about privacy and data management become more restrictive (GDPR).
- Spending human time in hacking and modifying an industrial software component is not productive.
- Open source in marine electronics is real: SignalK servers are available on both DIY or off the shelf marine electronics.

2021, the world has changed

FairWind++ an execution environment for the citizen science at the sea

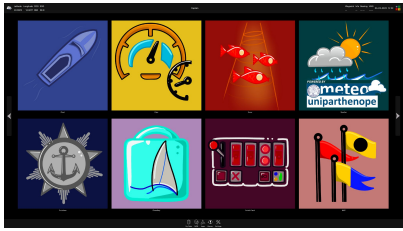
EuroHPC
Joint Undertaking

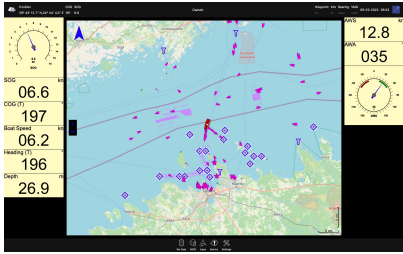
ADMIRE
malleable data solutions for HPC

ADAPTIVE MULTI-TIER INTELLIGENT
DATA MANAGER FOR EXASCALE

Presenting FairWind++

- FairWind has completely redesigned for the 3rd time.
- FairWind++ is an **execution environment for the FairWind Apps** running on a regular operating system.
- Any hardware architecture supporting QT5 can run FairWind++.
- System tested at the date:
  - MacOS
  - Ubuntu Linux (64 bit)
  - Raspberry Pi OS - on Raspberry Pi 4 (4GB) and Raspberry Pi 3b+ (1GB), 32 bit.
  - Microsoft Windows 10



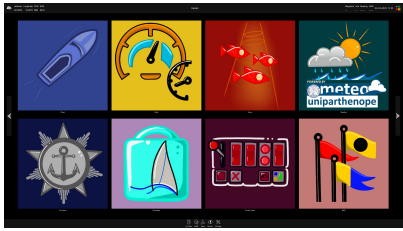


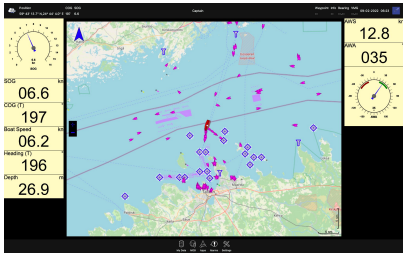
2021, the world has changed

Presenting FairWind++

Technical information

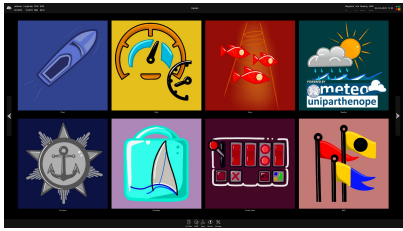
- Developed in C++ 17 and QT5 framework.
- “*Qt, pronounced as the word **cute**, is a widget toolkit for creating graphical user interfaces as well as cross-platform applications that run on various software and hardware platforms such as Linux, Windows, macOS, Android or embedded systems with little or no change in the underlying codebase while still being a native application with native capabilities and speed.*”
- Leveraging on external projects (we contributed to):
  - QtJsonSchema
  - Qt Custom Gauge Widget
  - QGeoView

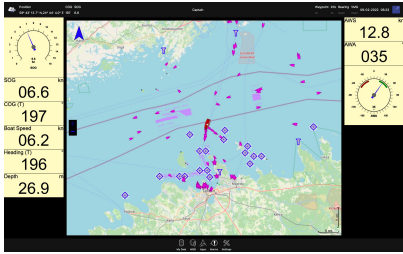




Presenting FairWind++

- Built using the FairWind SDK.
- The FairWindSDK is used to develop FairWind Apps.
- FairWind Apps are technically QT plugins honouring a standard interface.
- FairWind++ is highly modular: the most part of the UI is implemented by a FairWind App.
- FairWind Apps are shared libraries.
- FairWind Apps will be distributed using the app store metafore.

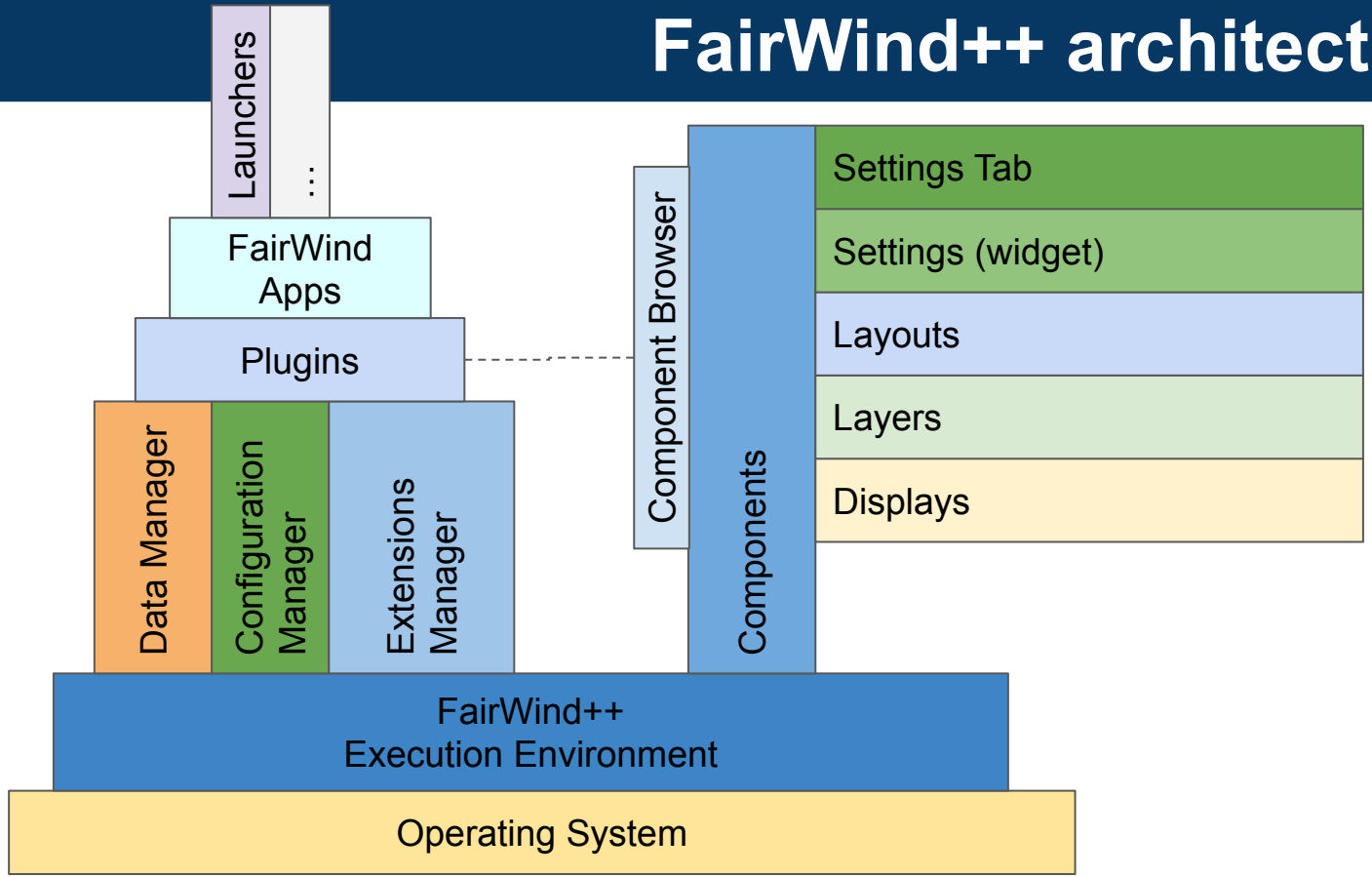




Technical information

FairWind++ architecture

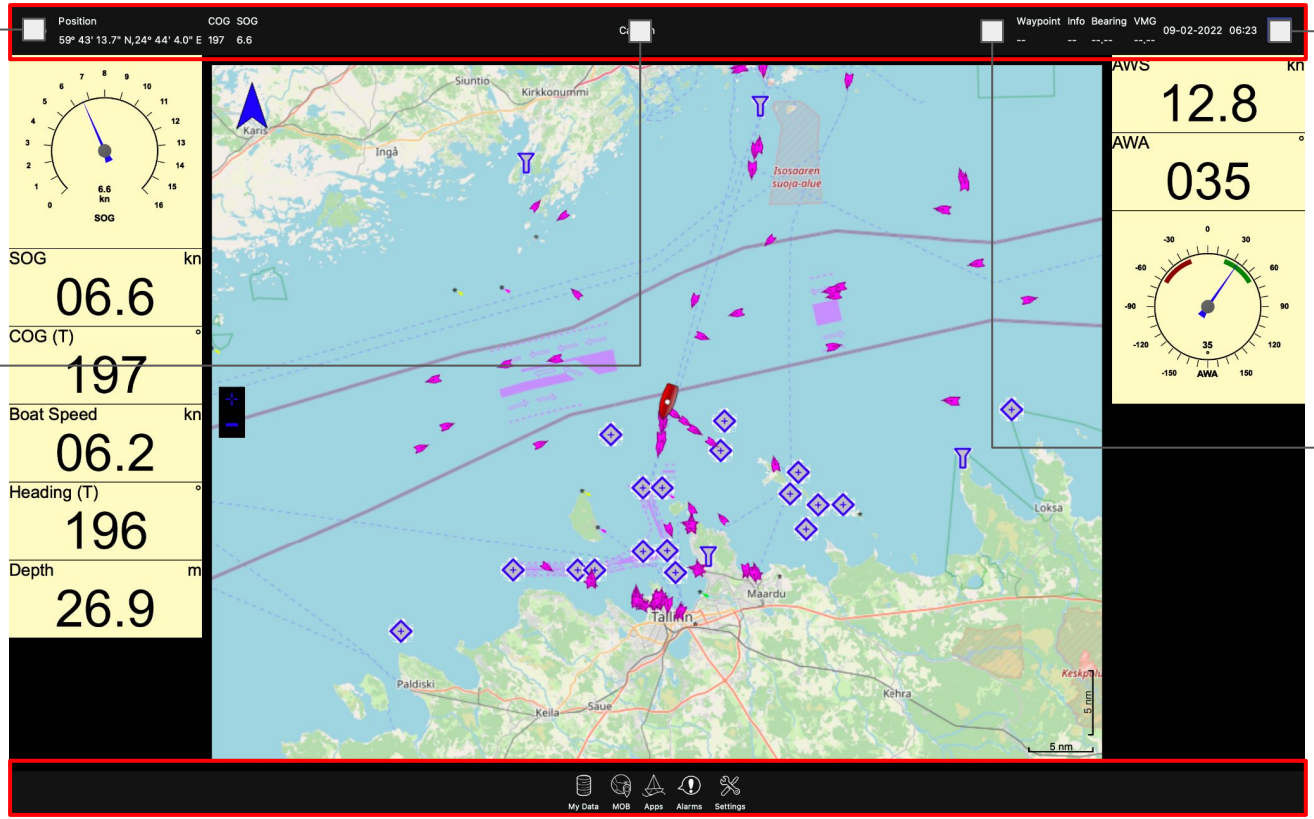
The block diagram



- FairWind++ is heavily modular leveraging on replaceable Plugins
- If a plugin registers a component, the component is made available to all other plugins

FairWind++ UI

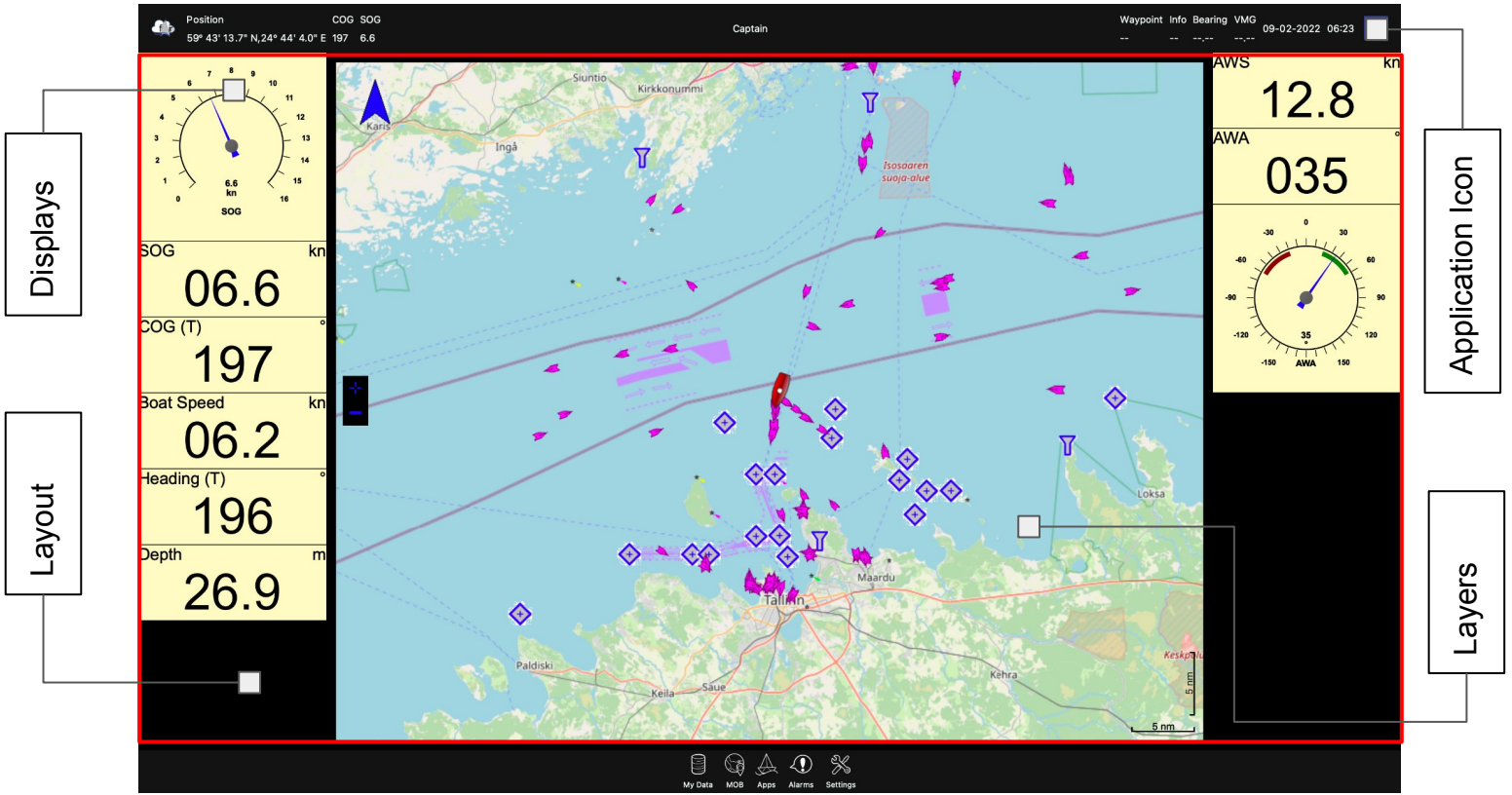
Mocking a Multi-functional Display



- The execution environment manages the top bar and the bottom bar.

FairWind App anatomy

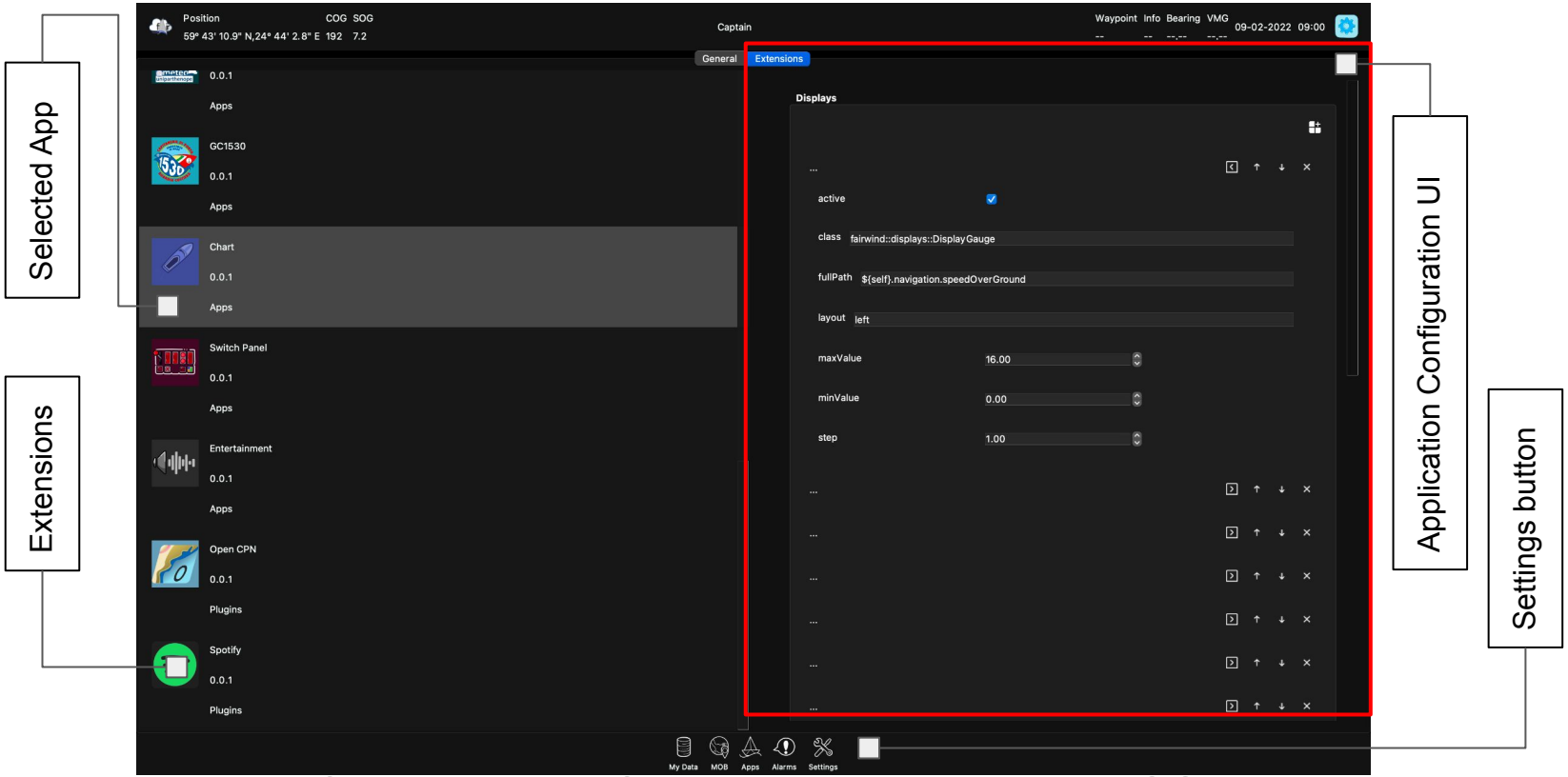
FairWind Apps live in the “App canvas”



- FairWind Apps implement the `IFairWind` interface

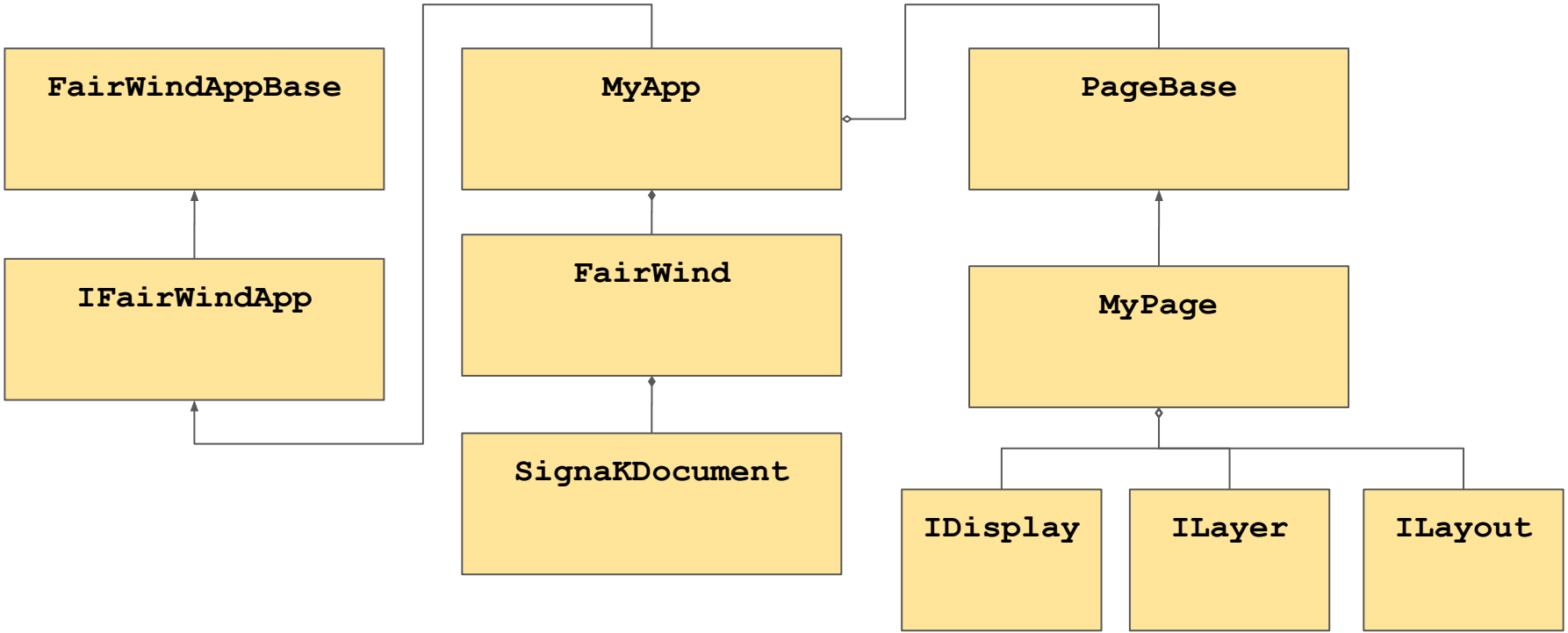
FairWind App anatomy

Configuration manager



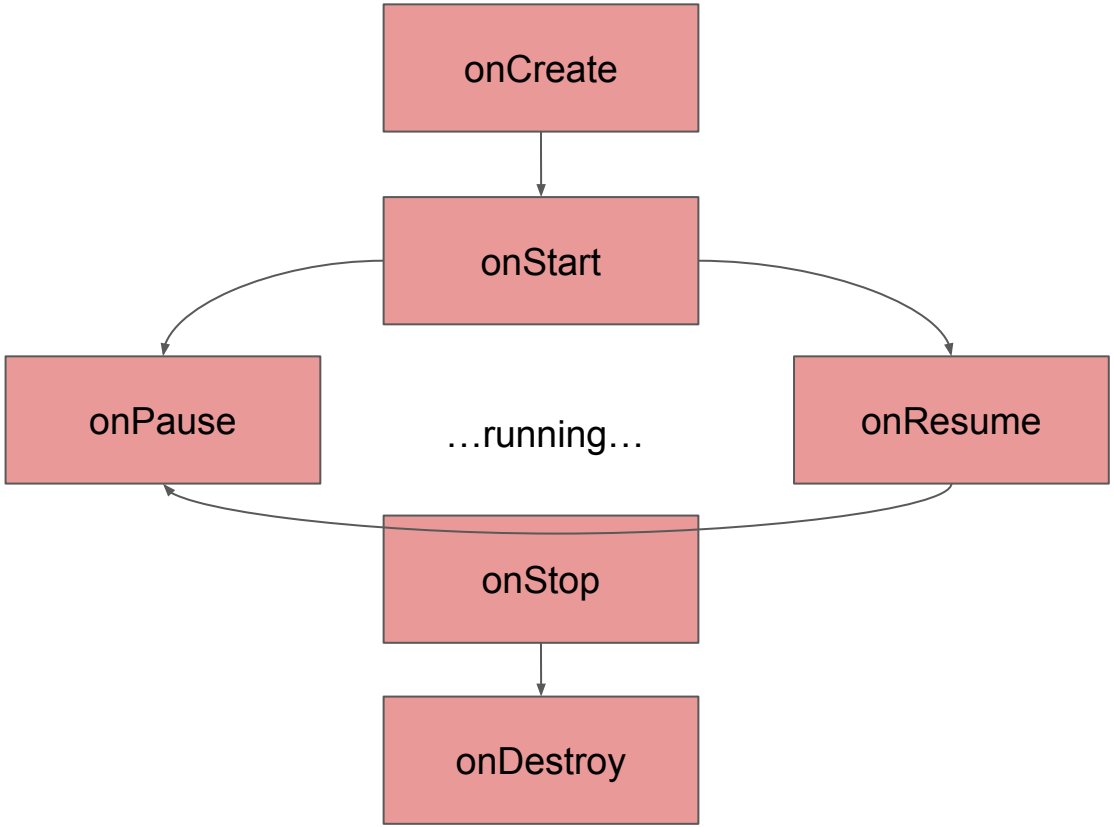
- The App configuration is defined using an extended JSON schema

FairWind App class diagram



- All the apps start as a “service”. The UI is provided by pages.

FairWind App lifecycle



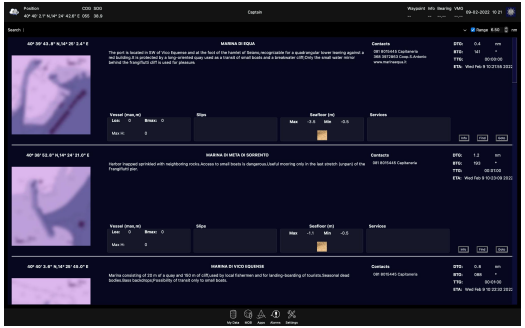
- …yes, it remembers a well known lifecycle.

So many kind of apps in FairWind++

- **Native FairWind Apps.**
  Developed with the FairWind SDK. Can be downloaded from an app store (when it will be available).

- **Mobile Apps.**
  Are apps for mobile devices connecting to the Signal K server and interacting with FairWind++.

- **Web Apps.**
  Are web applications running in a customized browser and hosted on a local or remote server. The Signal K application is an example.



The big picture: one step behind the A.I. for previsions augmented by predictions

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
EuroHPC Joint Undertaking
ADMIRE
malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

DYNAMO Architecture

EuroHPC Joint Undertaking ADMIRE — ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE — malleable data solutions for HPC

Vessel side

Data Movement Framework

Shore side

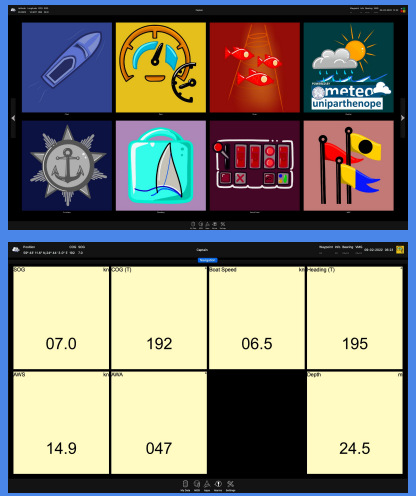
DYNAMO Signal K logger plugin

Signal K Server

Instruments

Actuators

FairWind++ Execution Environment



FairWind Apps

DYNAMO Storage

Labels

Features

Forecasting Models

Training

Trained Model

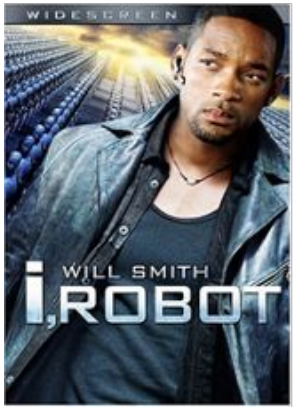
Predicting

Forecasts improved by predictions

Conclusion

- One of our core business is the weather and marine precision forecasts.
- In order to improve forecasting models' results we designed the DYNAMO ecosystem to implement the marine data crowdsourcing as the citizen science paradigm.
- Plan when the idea was born: data assimilation.
- The current approach: forecasts corrected by crowdsourced data using A.I.

Artificial Intelligence



…in the movies!

Artificial Intelligence

Face detection

Formal verification

A young science…
(about 50 years old)

- Exciting and dynamic field, lots of uncharted territory left.
- Impressive success stories.
- “Intelligent” in specialized domains.
- Many application areas.

Artificial Intelligence

- There is no clear consensus on the definition of AI
- John McCarthy coined the phrase AI in 1956

"*It is the science and engineering of making intelligent machines, especially intelligent computer programs.*

It is related to the similar task of using computers to understand human or other intelligence, but AI does not have to confine itself to methods that are biologically observable."

Artificial Intelligence

- AI is a collection of hard problems which can be solved by humans and other living things, but for which we don't have good algorithms for solving.

  Examples: understanding spoken natural language, medical diagnosis, circuit design, learning, self-adaptation, reasoning, chess playing, proving math theorems, etc.

- Russsell & Norvig: a program that
  - Acts like human (Turing test)
  - Thinks like human (human-like patterns of thinking steps)
  - Acts or thinks rationally (logically, correctly)

- Some problems used to be thought of as AI but are now considered not anymore.

Magic mirror on the wall, who is the fairest one of all?

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
EuroHPC Joint Undertaking
ADMIRE
malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

Data, big data & data science

Definition:

"Big data is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software."

Technologies aimed at extracting knowledge and value from data

Analysis of large amounts of information

Data, big data & data science

Big data: “Big, Fat, and Ugly”

Structured and unstructured

Economy, geolocation, time series, social networks

Petabyte, Zettabyte (billions of Terabytes)

Parallel Computing, GPGPU

The Storm Seeker

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
EuroHPC Joint Undertaking
ADMIRE
malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

Motivation

The Mediterranean area is subject to a range of destructive weather events, including middle-latitudes storms, Mediterranean sub-tropical hurricane-like storms (“medicanes"), and small-scale but violent local storms.

Although predicting large-scale atmosphere disturbances is a common activity in numerical weather prediction, the tasks of recognizing, identifying, and tracing trajectories of such extreme weather events within weather model outputs remains challenging.

A new approach to this problem → **StormSeeker**

Uses machine learning techniques to recognize, classify, and trace the trajectories of severe storms in atmospheric model data.

The StormSeeker approach

STEP 1

UNSUPERVISED

Identify clusters

The points with similar characteristics are identified through an autoencoder that produces a reduced dimensionality latent space representation of the meteorological forecasting model output.
To this new weather model output we then apply a NEgentropy Clustering (NEC) to assign a cluster label to each point.

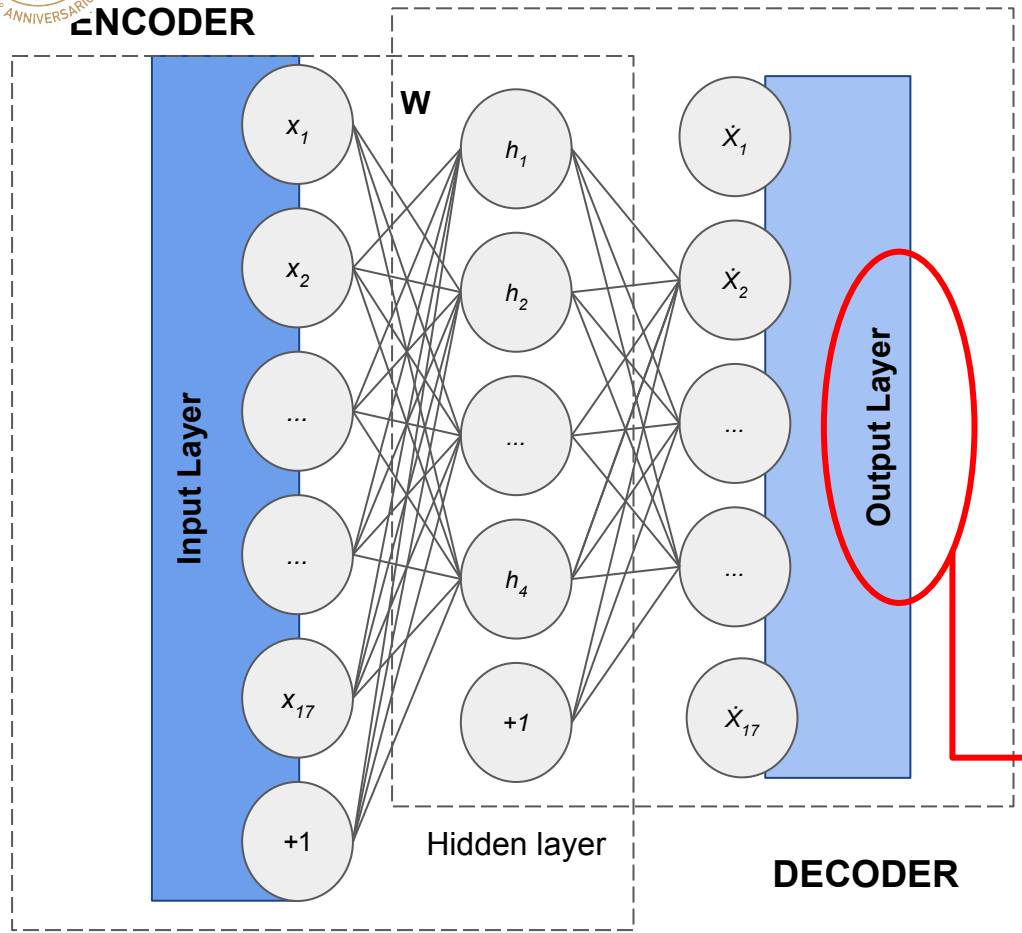
STEP 2

SUPERVISED

Training and applying a storm tracker

In order to identify the clusters that represent the actual storm we use a neural network trained on data labeled by experts.

Step 1: the autoencoder



Autoencoder → an unsupervised neural network (NN) that learns a non-linear representation (encoding) of a set of data.

Non-linear Principal Component Analysis

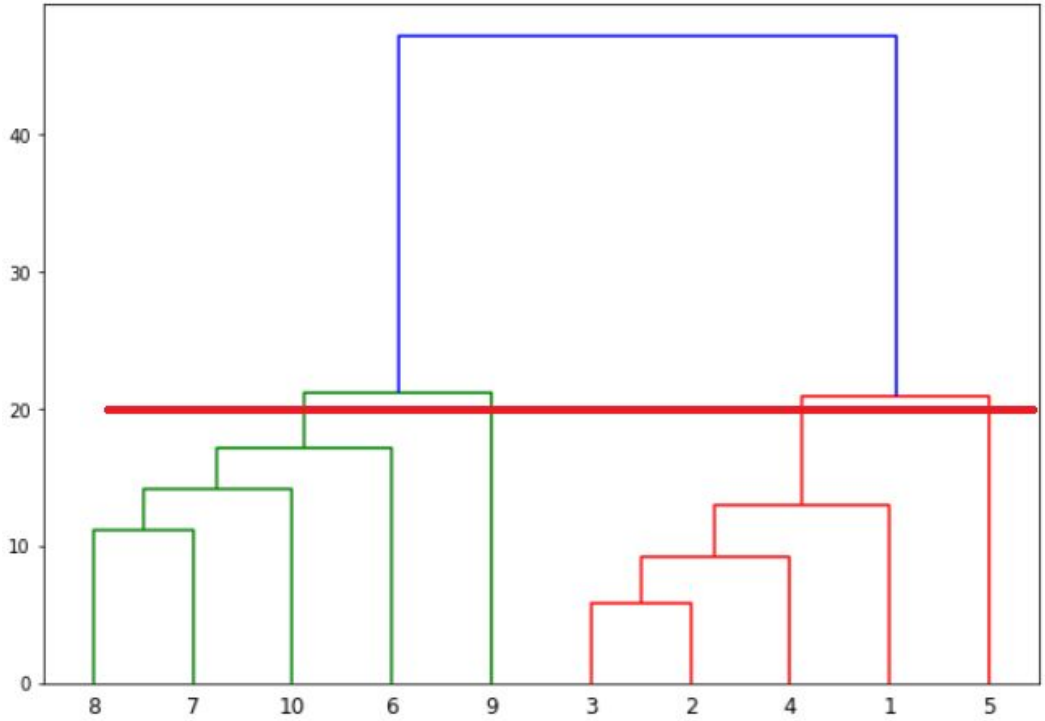
ENCODER → 17 inputs and 4 outputs (neural network hidden nodes)

DECODER → 4 inputs and 17 outputs

NEgentropy Clustering

EuroHPC Joint Undertaking ADMIRE malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

Step 1: NEgentropy Clustering



NEC → a hierarchical clustering method based K-means.

K-means clustering algorithm

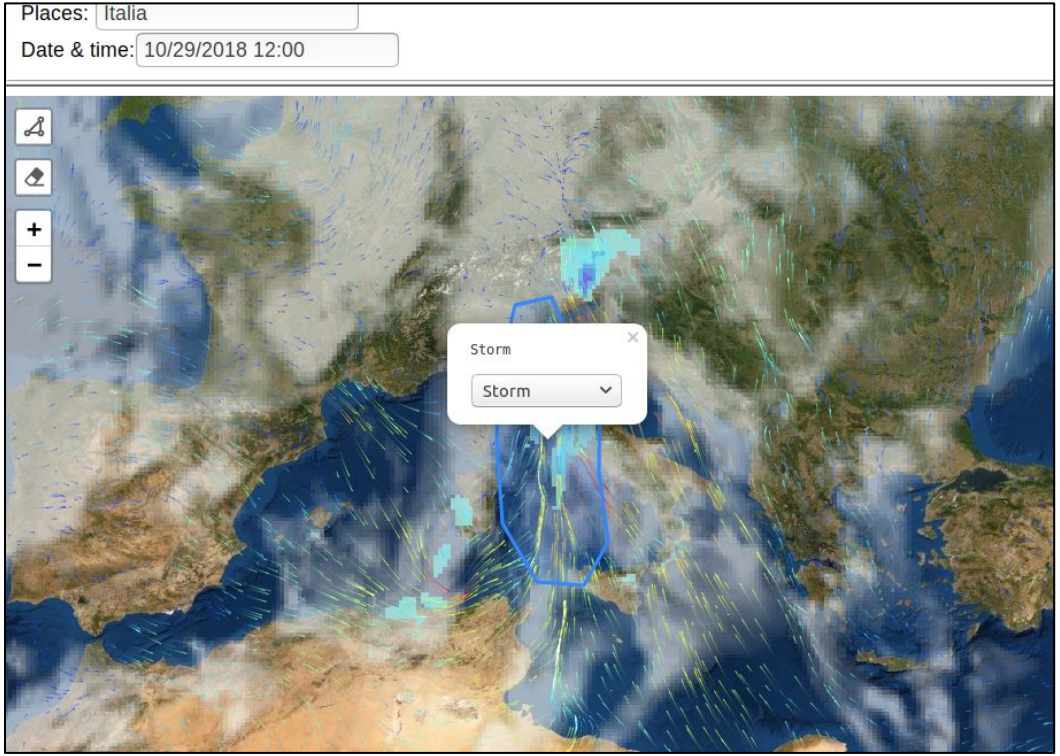
Target: groups the elements of a dataset into non-overlapping sets, such that all elements belonging to the same set are as similar as possible.

EuroHPC Joint Undertaking ADMIRE malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

Step 2: Labeling

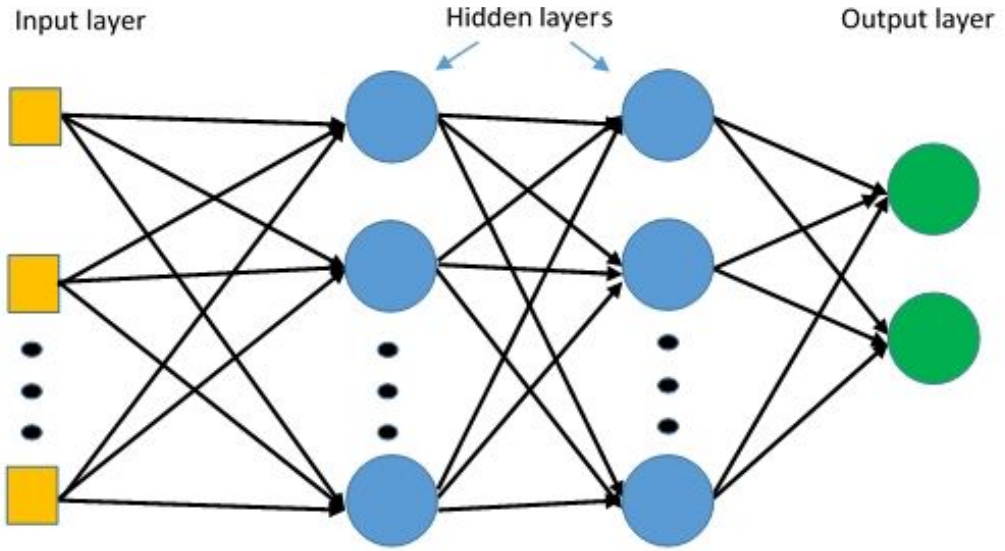
Target: allows the experts to mark and label the found clusters by polygons in weather model output. The polygons correspond to severe weather events.

Weather Labeling Web Application



EuroHPC Joint Undertaking
ADMIRE
malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

Step 2: Training and applying a storm tracker

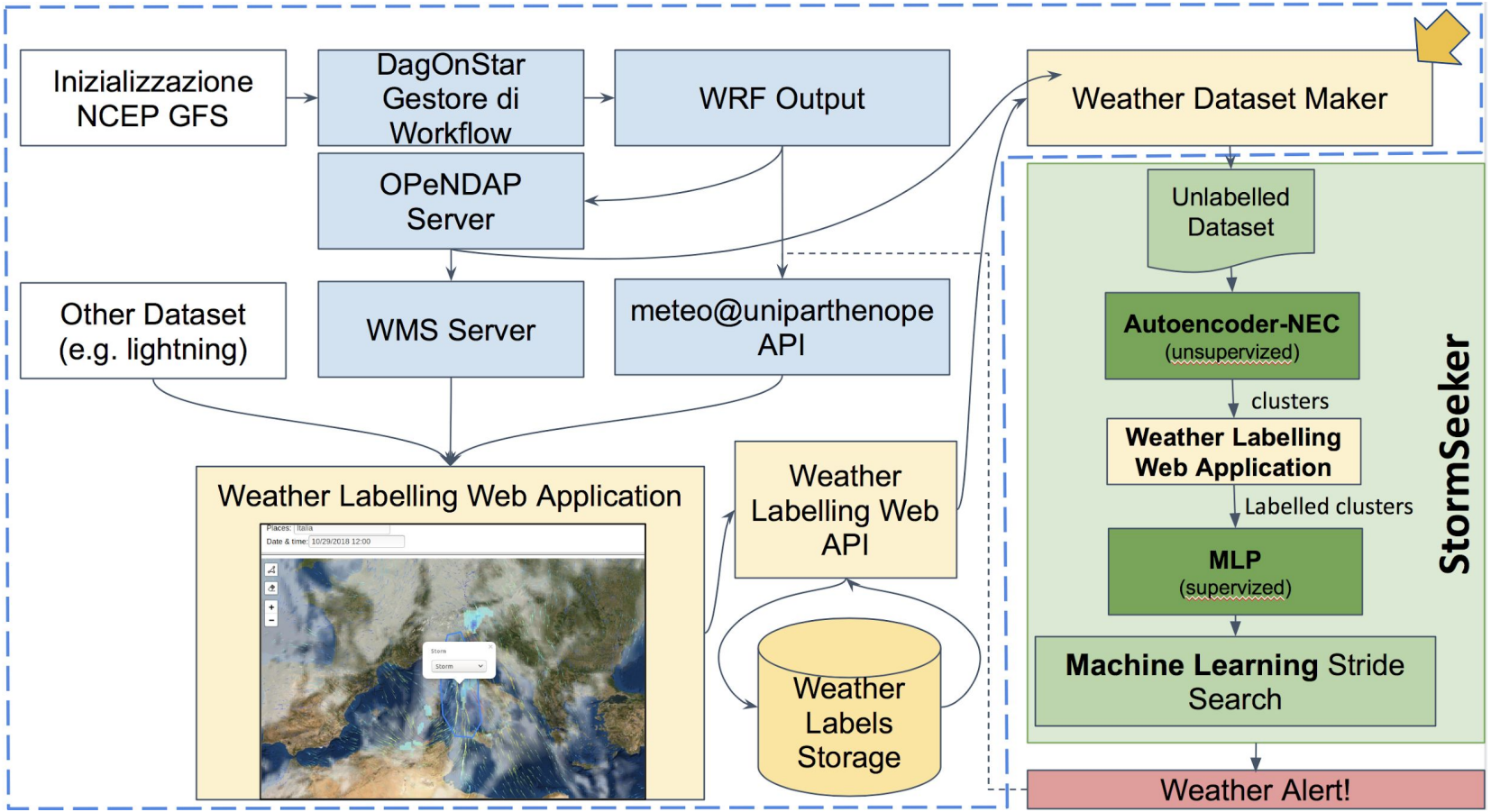


Target: use label data to train the NN that will subsequently be used for predicting new phenomena.

MultiLayer Perceptron approach (MLP) → uses a supervised learning technique called **backpropagation** for training.

We use a cross-validation methodology, a MLP with a single hidden layer on the labeled data to obtain a predictive model.

StormSeeker and Meteo@Uniparthenope

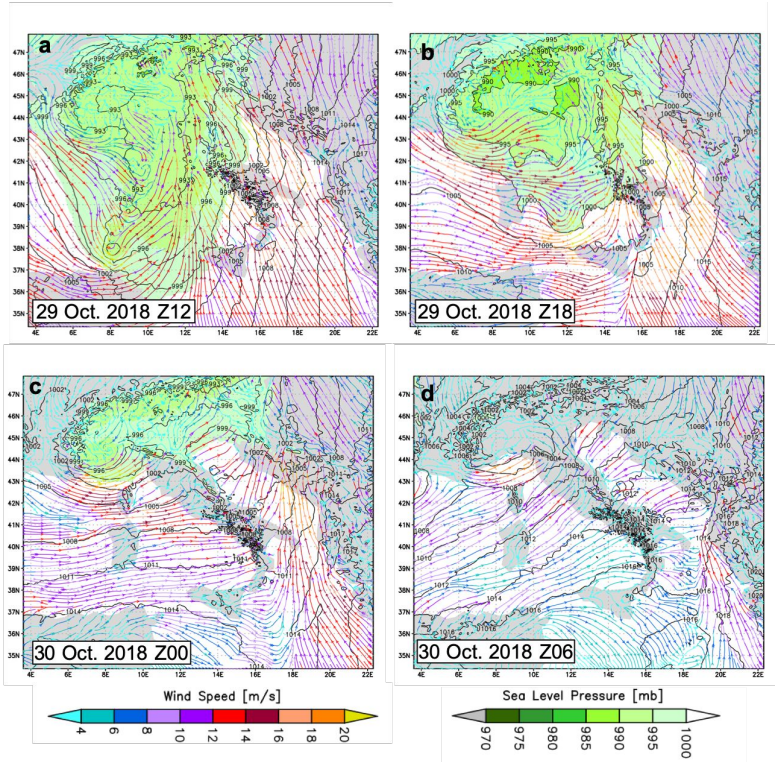


Case study: VAIA storm event

On 29-30 October 2019, the VAIA storm hits the North-Eastern regions of Italy by wind gusts exceeding 200 km/h.

Destructive effects of the storm connected to the storm VAIA in the Rapallo harbor (Genoa, Liguria).

WRF model weather forecast.

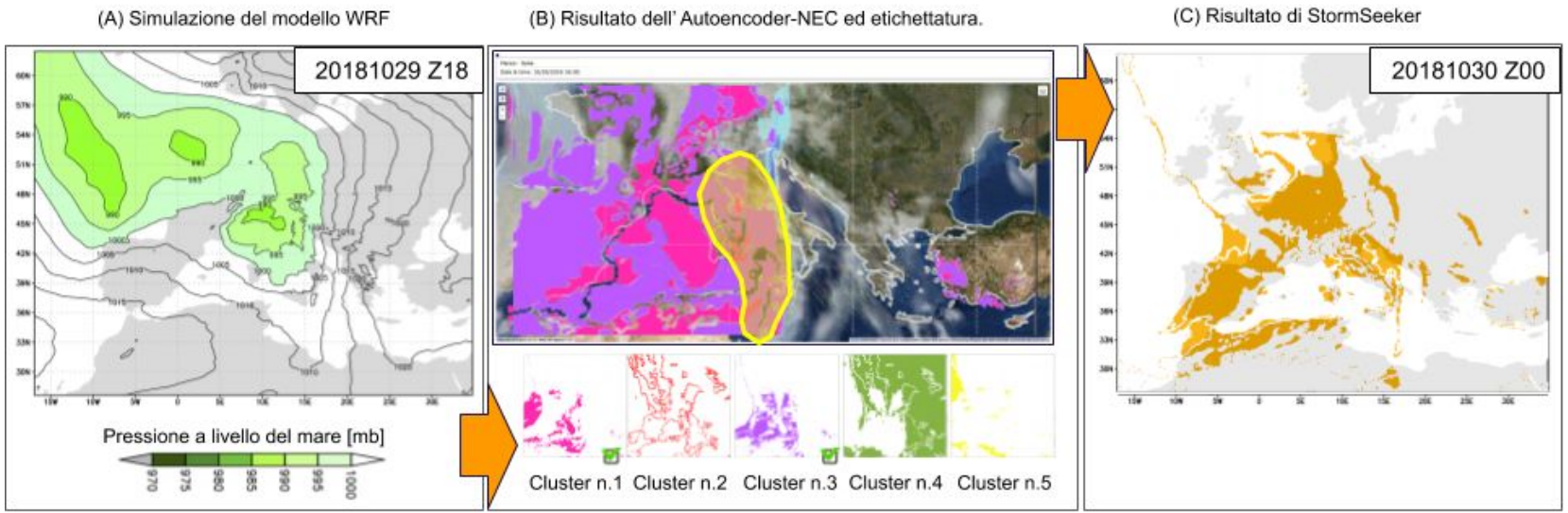


EuroHPC Joint Undertaking
ADMIRE
malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

Features used

Name	Description	Unit	S	U
T2C	Temperature at 2m	°C	✓	✓
SLP	Sea level pressure	hPa	✓	✓
WSPD10	Wind speed at 10m	ms^{-1}	✓	✓
WDIR10	Wind direction at 10m	°	✓	✓
HR2	Relative humidity at 2m	%	✓	✓
UH	Updraft helicity	m^{-2}/s^{-2}	✓	✓
MCAPE	Max. convective available potential energy	JKg^{-1}		✓
TC500	Temperature at 500 hPa	°C		✓
TC850	Temperature at 850 hPa	°C		✓
GPH500	Geopotential height at 500 hPa	m		✓
GPH850	Geopotential height at 850 hPa	m		✓
CLOUD	Cloud fraction	%	✓	✓
U10M	Wind at 10m (u-component)	ms^{-1}		✓
V10M	Wind at 10m (v-component)	ms^{-1}		✓
WSPD10$_\Delta$	Wind speed change from previous time step	ms^{-1}		✓
WDIR10$_\Delta$	Wind direction change from previous time step	°N		✓
RAIN	Hourly cumulative rain	mm	✓	✓

StormSeeker application during VAIA storm event



An example of StormSeeker:
A) The WRF output is reduced in its dimensions using the Autoencoder-NEC;
B) the clusters 1 and 3 are overlaid on the WRF output in order to assist the field expert in severe weather events labelling;
C) Final result.

AUGMENTING MARINE POLLUTANTS TRANSPORT AND DIFFUSION MODEL RESULTS USING AI PREDICTION

ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

malleable data solutions for HPC

Introduction

- The most of the fish commerce is constituted by the fish and mussels (Mytilus galloprovincialis) farming
- Continuous monitoring to enforce food security and quality

Motivation and novel contributions

Motivations

- Relationship between **exposure time of the mussel to pollutants** and the **actual absorption of pollutants**
- Numerical model of mussel assimilation is unknown
- Algorithm for the modeling of the processes of accumulation of pollutants in mussel based on artificial intelligence technologies

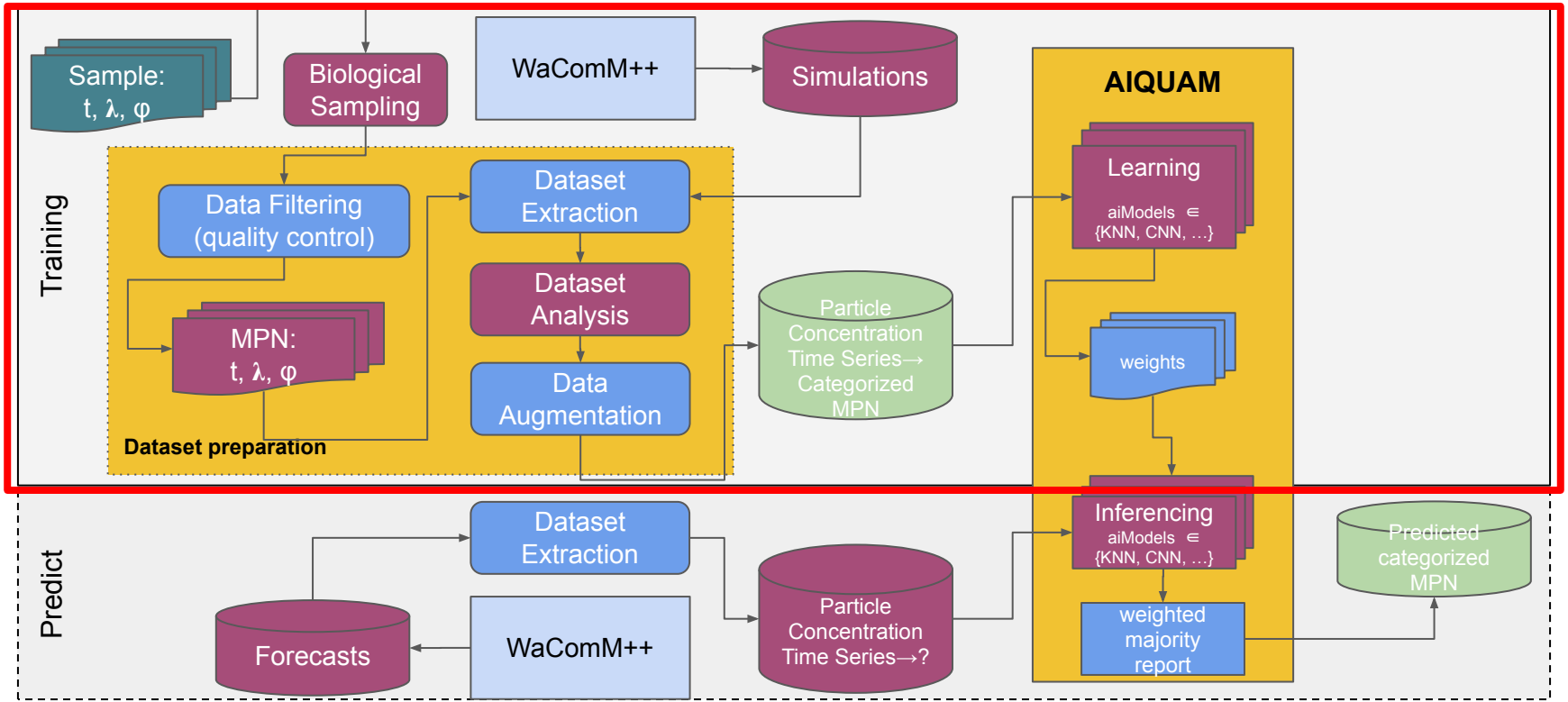
Novel contributions

- **Dataset time series creation**
- **Machine Learning Models**

Artificial Intelligence-based water QUAlity Model

- Software component enabling the user to predict, with a reasonable probability, the contaminant levels in shellfish farms
- **High-resolution numerical models** + **Machine Learning**
  - multiple AI models
  - selecting the best result with a weighted majority report methodology

Architecture - Training phase



Data filtering

- Microbiological sampling is the label of the training dataset
- max, **mean** or median

MPN (Most Probable Number): analytical methodology for counting the microbial burden of an organic sample.

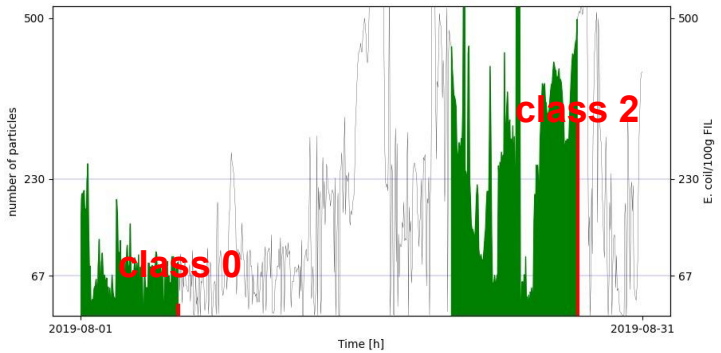
	A	B	C	D	E	F	G	H	I	J	K
1	NUMERO SCHEDA	ANNO AC	NUMERO ACCETTAZIONE	IZS ACCETTAZIONE	DATA PRELIEVO	DATA ARRIVO	DATA ESITO	REGIONE	Codice SITO	SITO	ESITO
1295	1043A-101608-B	2019	101608	ISTITUTO ZOOPROFILATTICO SPERIMENTALE DEL MEZZOGIORNO	30/09/19	30/09/19	04/10/19	CAMPANIA	1500016	VARCATURO GIACOBBE	230
1296	1043A-101610-B	2019	101610	ISTITUTO ZOOPROFILATTICO SPERIMENTALE DEL MEZZOGIORNO	30/09/19	30/09/19	04/10/19	CAMPANIA	1500009	MONTE DI PROCIDA	230
1297	1043A-101611-B	2019	101611	ISTITUTO ZOOPROFILATTICO SPERIMENTALE DEL MEZZOGIORNO	30/09/19	30/09/19	04/10/19	CAMPANIA	1500038	ACQUAMORTA	18
1298	1043A-101611-B	2019	101611	ISTITUTO ZOOPROFILATTICO SPERIMENTALE DEL MEZZOGIORNO	30/09/19	30/09/19	04/10/19	CAMPANIA	1500038	ACQUAMORTA	45
1299	1043A-101611-B	2019	101611	ISTITUTO ZOOPROFILATTICO SPERIMENTALE DEL MEZZOGIORNO	30/09/19	30/09/19	04/10/19	CAMPANIA	1500038	ACQUAMORTA	18
1300	1043A-102056-B	2019	102056	ISTITUTO ZOOPROFILATTICO SPERIMENTALE DEL MEZZOGIORNO	01/10/19	01/10/19	04/10/19	CAMPANIA	1500026	TORRE DI PESCOPAGANO	45
1301	1043A-102125-B	2019	102125	ISTITUTO ZOOPROFILATTICO SPERIMENTALE DEL MEZZOGIORNO	01/10/19	01/10/19	04/10/19	CAMPANIA	1500012	PUNTA CAVALLO, NISIDA	18
1302	1043A-102133-B	2019	102133	ISTITUTO ZOOPROFILATTICO SPERIMENTALE DEL MEZZOGIORNO	01/10/19	01/10/19	04/10/19	CAMPANIA	1500012	PUNTA CAVALLO, NISIDA	18
1303	1043A-102583-B	2019	102583	ISTITUTO ZOOPROFILATTICO SPERIMENTALE DEL MEZZOGIORNO	02/10/19	02/10/19	07/10/19	CAMPANIA	1500038	ACQUAMORTA	110

Data extraction

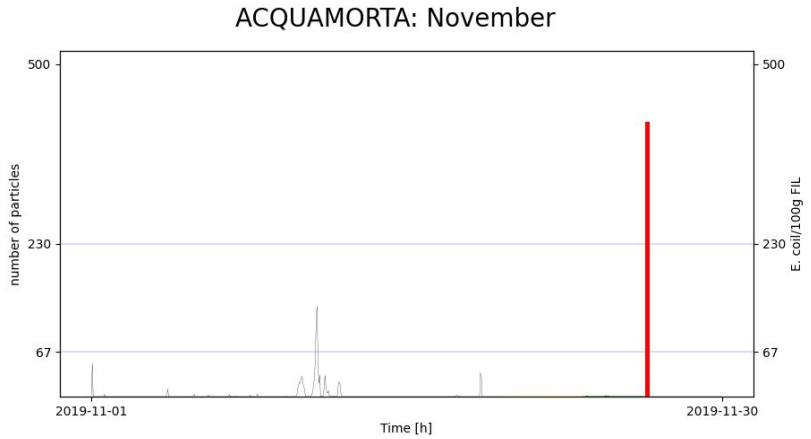
- The features are built by the concentration of particles per hour (72) produced by the **WaComM++ model**
- The time of assimilation of mussels can be expressed by

$$\int_{t_0}^{t_0-\Delta t} f\, dt$$

- Dataset is composed as follows:
  - **features**
  - **labels:**
    - **class 0 (0-67 MPN / 100 g)**
    - **class 1 (67-230 MPN / 100 g)**
    - **class 2 (230-4600 MPN / 100 g)**
    - **class 3 (> 4600 MPN / 100 g)**

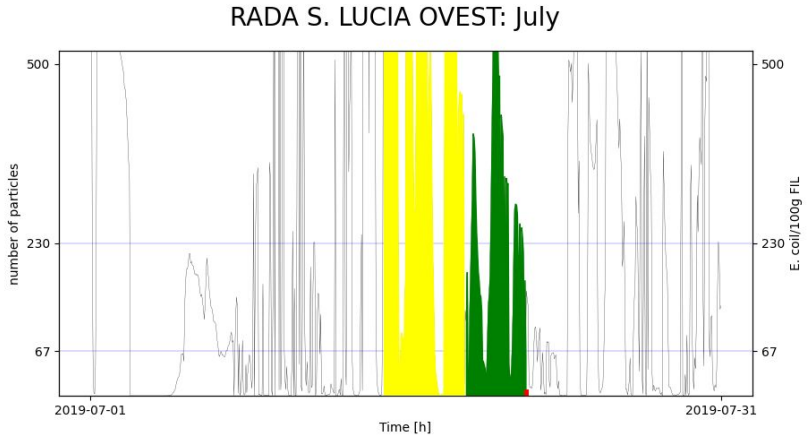


Data analysis



← Low concentration - High value

High concentration - Low value →



Data augmentation

- **Data augmentation** technique to amplify dataset
  - is used to increase amount of data. Two methods:
    - **duplicate sample** in the minority class
    - **synthesize new samples** from existing examples: **SMOTE** (Synthetic Minority Oversampling Technique

- After the “data augmentation” phase, our dataset now has 221 samples, of which:
  - **111** belonging to **class 0** (0-67 MPN/100g);
  - **55** belonging to **class 1** (67-230 MPN/100g);
  - **55** belonging to **class 2** (230-4600 MPN/100g)

- The result is a less unbalanced dataset.

Time-series classification

- **AIQUAM** uses multiple machine learning models to solve the *"time-series classification"* (TSC) problem

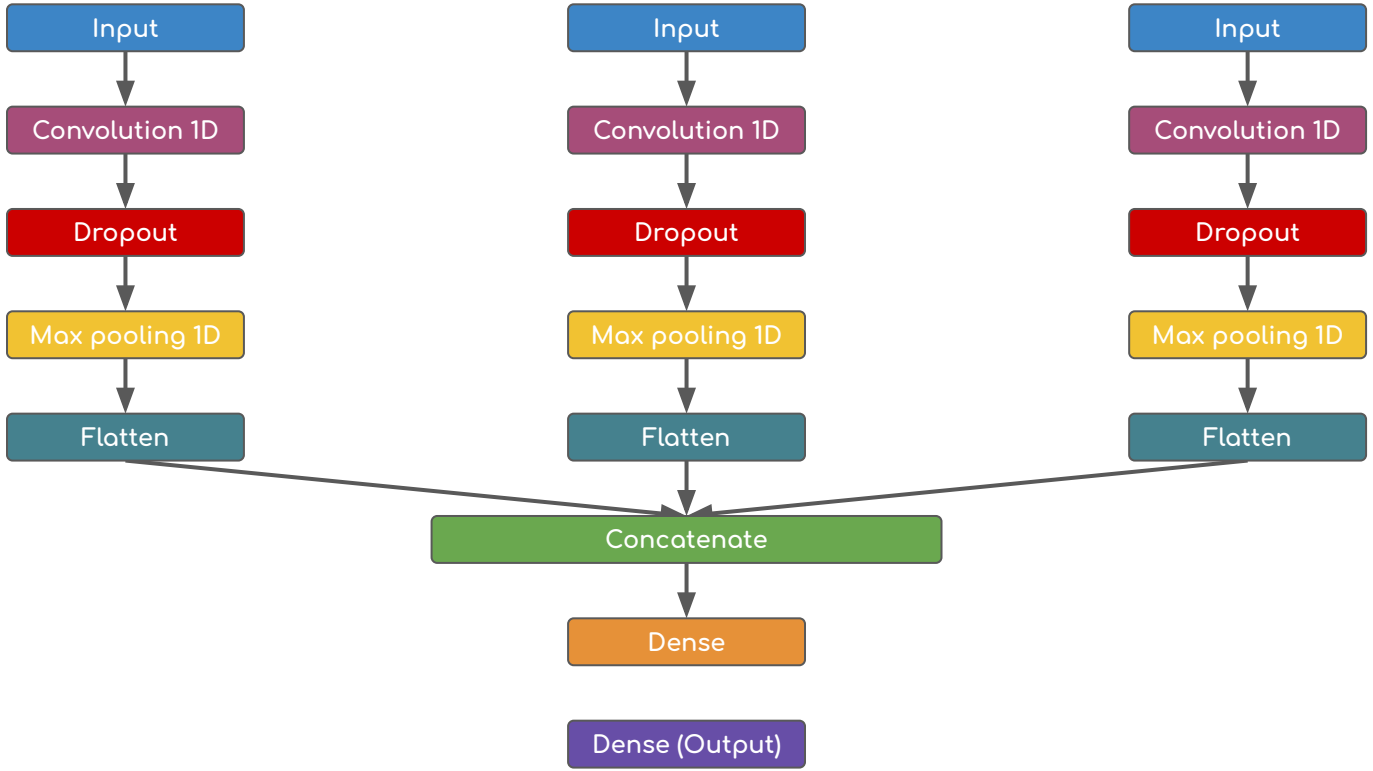
- The goal is to train a model that can accurately predict a time series class, given a dataset with labeled time sequences

- There are many approaches for time series classification, which can be summarized in three large categories according to the classification scheme:
  - **model based**
  - **distance based**
  - **feature based**

Machine Learning models

- We tested two machine learning models:
  - **Convolutional Neural Network** (features-based)
  - **K-Nearest Neighbour** (distance-based)
- During the prediction stage, **AIQUAM** selects the best results using a ***weighted majority report strategy***

Convolutional neural network



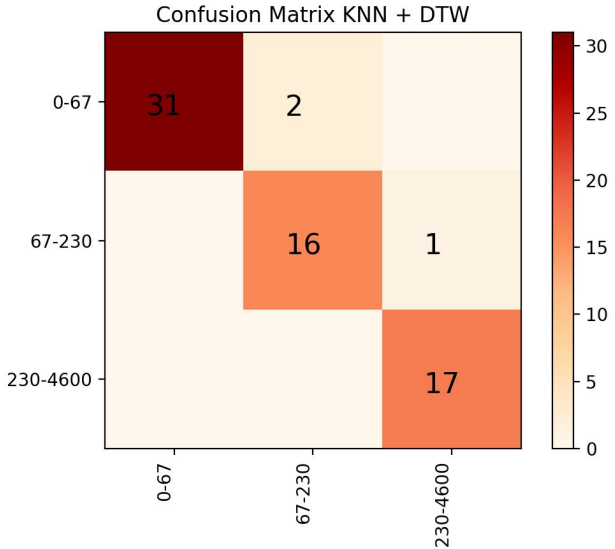
- These three CNNs differ in kernel size
- The kernel size in the first CNN is equal to 3, in the second, it is equal to 5, and in the third, it is equal to 11
- Small filter sizes capture great details of the input, while large ones leave minute details in the input
- These three CNN are concatenated to create the **Fully Connected** (FC) **Layer**

K-Nearest Neighbors

- K-Nearest Neighbors (KNN) algorithm stores all the available cases and classifies the new data or case based on a similarity measure
- In the classification setting, the KNN algorithm essentially boils down to forming a majority vote between the k most similar instances to a given ”unseen“ observation
- The similarity is defined according to a distance metric between two data points calculated using the following methods:
    - **Euclidean Distance**
    - **Minkowski Distance**
    - **Dynamic Time Warping (DTW)**

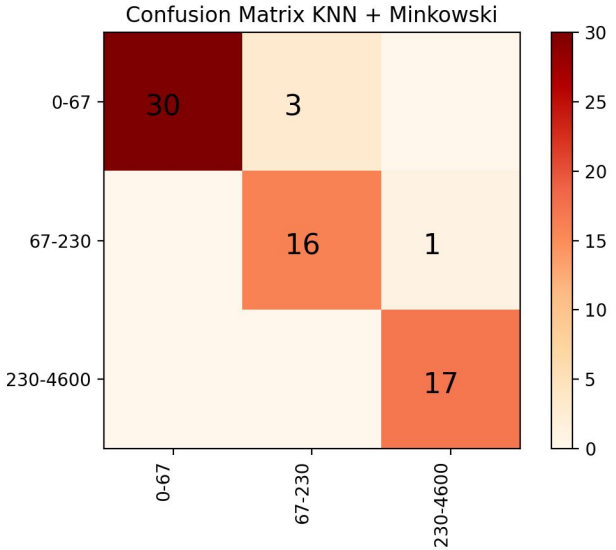
Evaluation: KNN+DTW

KNN + DTW				
	precision	recall	f1-score	support
0-67 MPN/100g	1.00	0.94	0.97	33
67-230 MPN/100g	0.89	0.94	0.91	17
230-4600 MPN/100g	0.94	1.00	0.97	17
accuracy			0.96	67



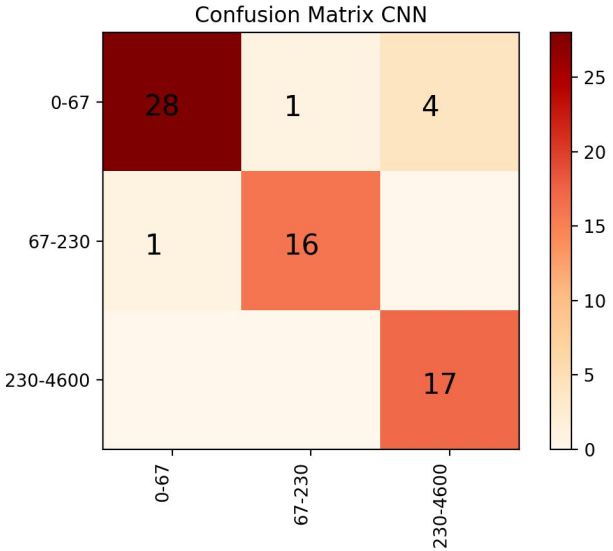
Evaluation: KNN + Minkowski

KNN + Minkowski				
	precision	recall	f1-score	support
0-67 MPN/100g	1.00	0.91	0.95	33
67-230 MPN/100g	0.84	0.94	0.89	17
230-4600 MPN/100g	0.94	1.00	0.97	17
accuracy			0.94	67



Evaluation: CNN

CNN				
	precision	recall	f1-score	support
0-67 MPN/100g	0.97	0.85	0.90	33
67-230 MPN/100g	0.94	0.94	0.94	17
230-4600 MPN/100g	0.81	1.00	0.89	17
accuracy			0.91	67



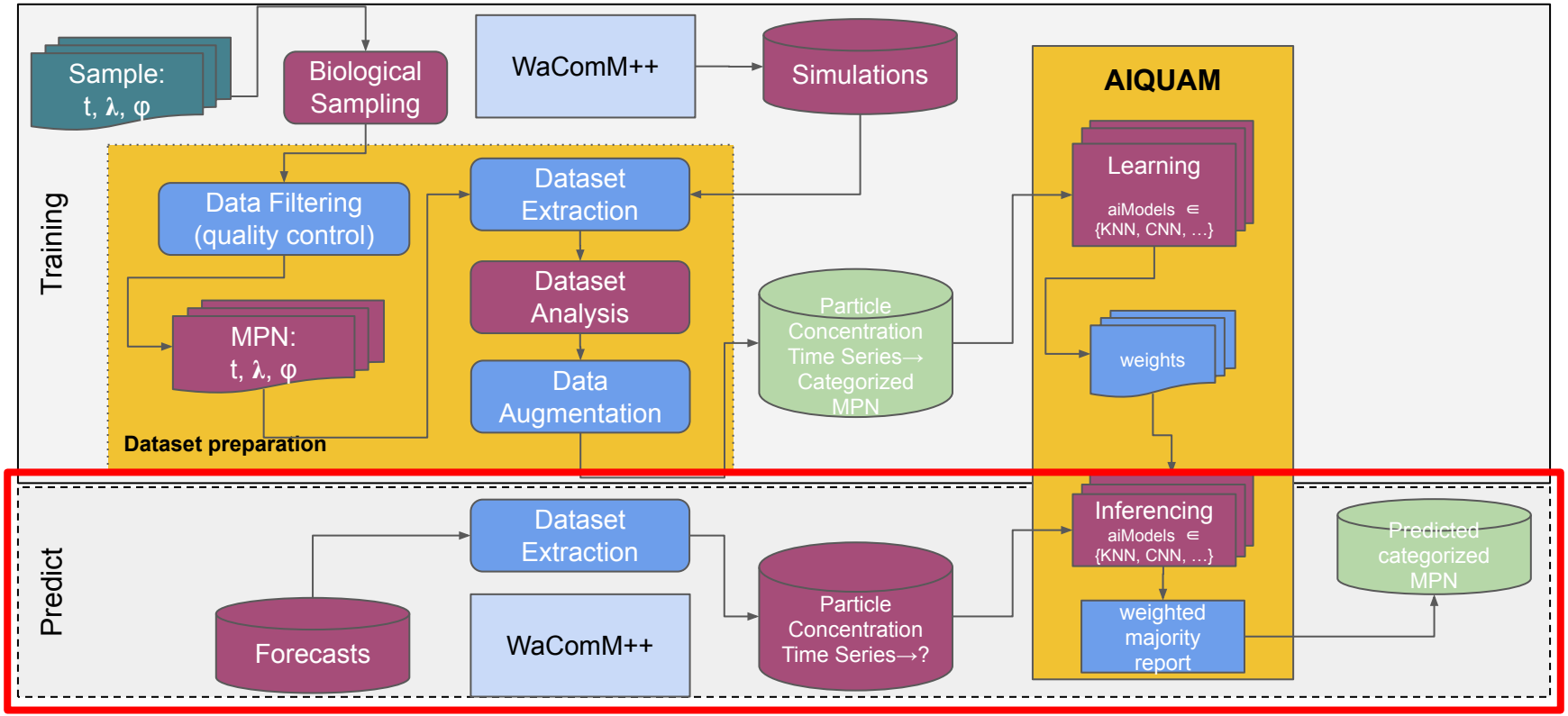
EuroHPC Joint Undertaking
ADMIRE
malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE

Results summary

- Cross Validation**: Leave-One-Out**
- **3 machine learning models** = **3** (probability) **different results**
- **Majority vote** to decide the **class to belong to**

Model	accuracy
KNN + DTW	**0.96**
KNN + Minkowski	**0.94**
CNN	**0.91**

Architecture - Prediction phase



Use case

- The Machine Learning models were trained with samples taken by competent authorities in the Bay of Naples in 2019 and 2020.
- We fed the machine learning models with unseen data taken in 2021 in the Bay of Naples to test our models.

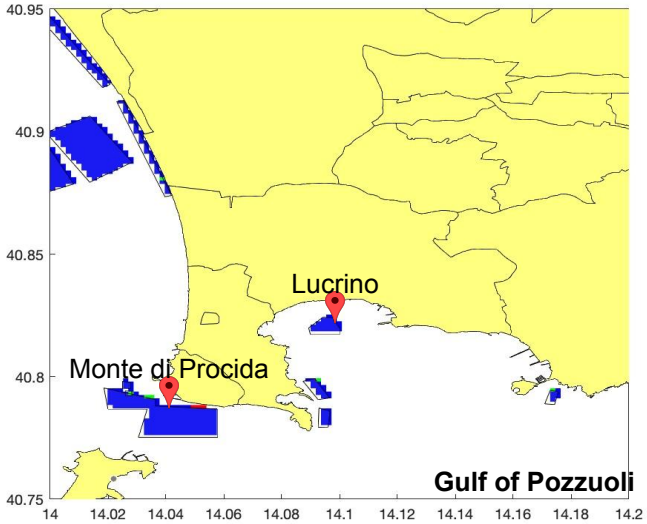
Date	Zone	Microbiological concentration of E. coli (class)	Predicted class
24/08/2021	Monte di Procida	78 MPN/100g (0)	0
24/08/2021	Monte di Procida	230 MPN/100g (1)	1
24/08/2021	Monte di Procida	4900 MPN/100g (2)	2
30/08/2021	Lucrino	67 MPN/100g (0)	0
30/08/2021	Lucrino	67 MPN/100g (0)	0
30/08/2021	Lucrino	67 MPN/100g (0)	0

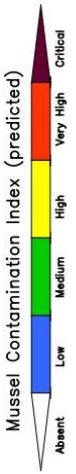
Use case

a)

Predicted results

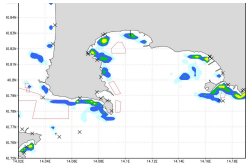
August 24, 2021 -
10:00 UTC



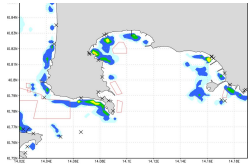


b)

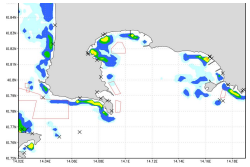
WaCom M++



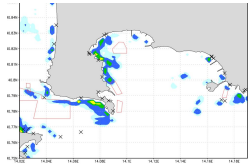
August 21 2021 -
10:00 UTC



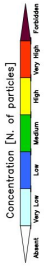
August 22 2021 -
10:00 UTC



August 23 2021 -
10:00 UTC



August 24 2021 -
10:00 UTC



Conclusion

- A new methodology has been proposed to predict the concentration of pollutants in mussels using **Artificial Intelligence**
- **KNN method** performs very well for this type of problem even though we had very few microbiological samples available
- The best results were obtained with the **KNN + DTW method** with an accuracy higher than **90%**.

Future work

- Implement a semi-automatic protocol to assimilate data obtained from microbiological analysis

- Use a **decision support system** based on **Explainable AI**, set of tools and frameworks to help to understand and interpret predictions made by machine learning models

The End

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
EuroHPC Joint Undertaking
ADMIRE
malleable data solutions for HPC
ADAPTIVE MULTI-TIER INTELLIGENT DATA MANAGER FOR EXASCALE