# IMPROVING AIR DISPERSION PREDICTION EXPLOITING COMPLEX INFRASTRUCTURE AND ENSEMBLE FORECAST

AUTHORS: FABIEN BROCHETON, DAVID POULET (NUMTECH), ANTONELLA GALIZIA, ANTONIO PARODI (CIMA FOUNDATION, IMATI-CNR), MARTIN GOLASOWSKI, JAN MARTINOVIC (IT4INNOVATIONS)

## Concept and approach

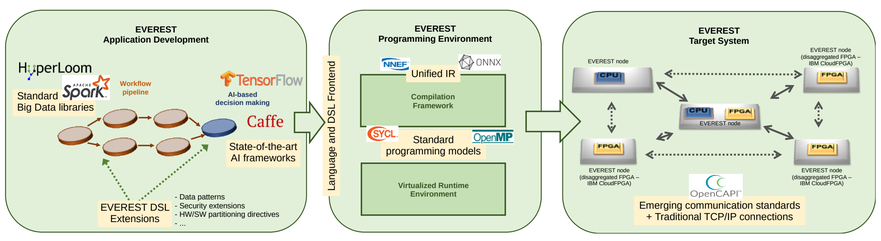
EVEREST focuses on **High Performance Big Data Analytics (HPDA) applications.**

- Future Big Data systems will be data-driven.
- Complex heterogeneous and reconfigurable architectures are difficult to program.

**The EVEREST project aims at developing a holistic approach for co-designing computation and communication in a heterogeneous, distributed, scalable, and secure system for HPDA.**

Main features:

- **data-driven design approach;**
- combination of **compiler transformations, high-level synthesis, and memory management;**
- efficient monitoring of the execution with a **virtualisation-based environment.**

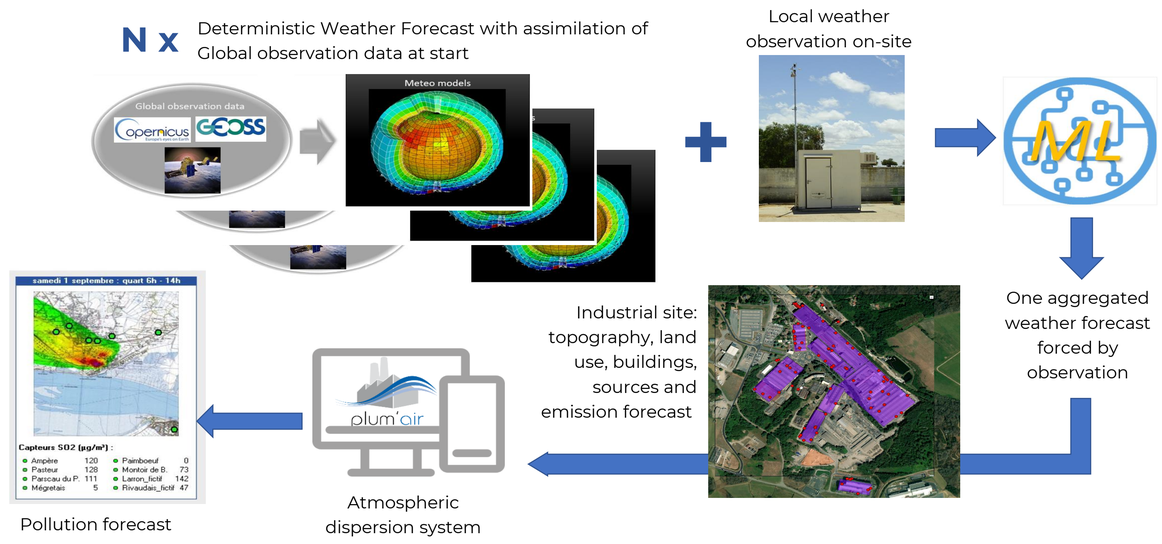



EVEREST proposes a **design environment** that combines state-of-the-art, stable programming models, and emerging communication standards with novel and **dedicated domain-specific extensions.** The EVEREST approach will be validated on three industrial use cases, one is related to **air quality dispersion.**

## Air quality dispersion prediction

The use case will combine deterministic **high resolution weather forecast** processed by WRF models and improved by **global data assimilation,** with an application based on **Machine Learning (ML)** algorithms analysing **historical site-specific datasets** to obtain **atmospheric impact forecast from industrial emission.**

**GOAL:** achieve **better accuracy of forecast** of atmospheric impact of a site in a context of daily environmental management (reduce pollution peaks due to weather situation and process operation forecast) and in consequence to reduce either non predicted pollution peaks or to reduce financial cost (activation of emission treatment, reduction of production, etc.) when false peaks are predicted.




## Meteorological model

The production (and the prediction) of weather-based scenarios is provided as an EVEREST service for downstream applications.

The WRF model can be described as a computational and memory intensive model highly demanding for ICT resources.

## WRF model configuration

WRF is a state-of-the-art numerical prediction model and encompasses physics schemes, numeric/dynamics options, initialisation routines, and a data assimilation package (WRFDA). The EVEREST design environment allows:

- to push forward data assimilation aspects to achieve augmented descriptions of the atmospheric state – used as initial conditions of the run,
- to improve model resolution in terms of spatial and temporal scale thus to better fit the geographical domain,
- to speed up performance figure of the WRF model by the means of the EVEREST FPGA-hardware to implement and test an ensemble prediction.

## Development of a ML approach

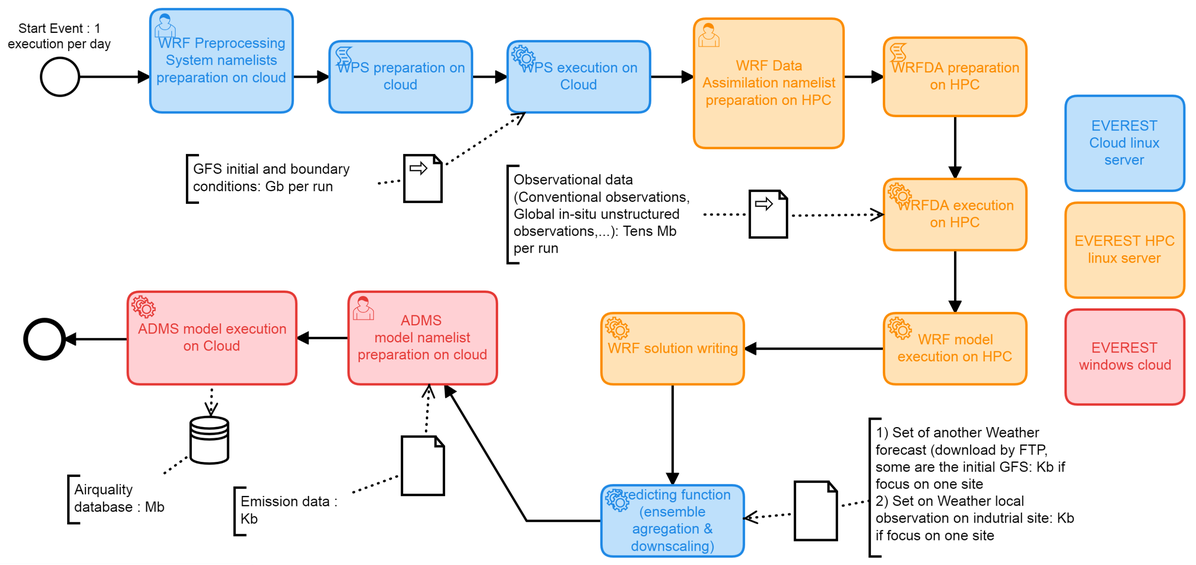
Selection of **Ridge regression with discounted loss method**, with objective to minimise the RMSE between the aggregate prediction and the observation over a period.

The application is written in **Python.**

**Training** need one year of historical data (hourly weather observation on site and hourly weather forecast from different sources (EVEREST forecast, NUMTECH forecast, global GFS forecast).

Ensemble aggregation will be first applied for **temperature, wind speed and wind direction**.

## Workflow implementation

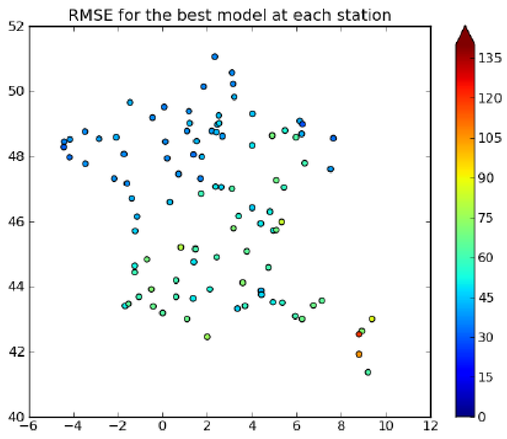



## Exploitation / Validation

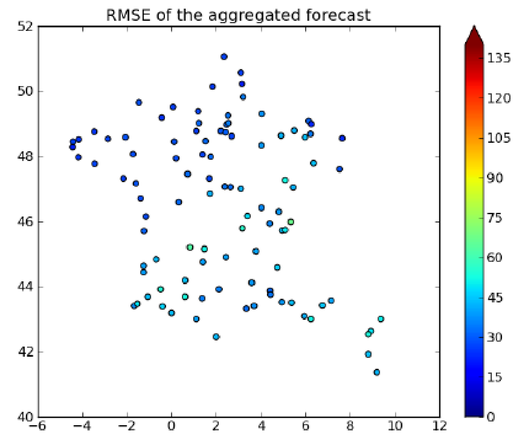
**To validate EVEREST implementation:**

- Execution of workflow will be compared to NUMTECH operational system (execution time, ....)
- Results will be compared to past forecast from NUMTECH operational system for at least **three different industrial sites** in France (different weather typology) and by comparing atmospheric results to air quality measurement network around industrial sites.

## First results of ML approach – wind direction for different locations in France (72 hours of forecast)




Deterministic forecast




Ensemble forecast forced by observation

This work was partially supported also by the LEXIS project – the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825532.

EVEREST
DESIGN ENVIRONMENT FOR EXTREME-SCALE BIG DATA ANALYTICS ON HETEROGENEOUS PLATFORMS

WWW.EVEREST-H2020.EU
EVEREST-INFO@A.ALARI.CH
WWW.LINKEDIN.COM/COMPANY/EVEREST-H2020
WWW.FACEBOOK.COM/EVERESTH2020

**PROJECT COORDINATOR:** CHRISTOPH HAGLEITNER (IBM ZURICH) | **SCIENTIFIC COORDINATOR:** CHRISTIAN PILATO (POLITECNICO DI MILANO)

**GRANT NUMBER: 957269 | TOPIC: ICT-51-2020 – BIG DATA TECHNOLOGIES AND EXTREME-SCALE ANALYTICS | FUNDING SCHEME: RIA**