Auteur principal : Alburquerque Otero, Cristina – cristinaalburquerque@porstdebalears.com – Balearic Port Authority – Moll Vell, 3-5. Palma – Spain

Co-auteurs :
- Martín Jiménez, Jorge – jmartin@portsdebalears.com – Balearic Port Authority
- Molina, Benjamín - benmomo@upvnet.upv.es - Universitat Politecnica de Valencia - Camino de Vera s/n 46022 Valencia, Spain
- del Rey Pérez, Amelia - adelrey@prodevelop.es - PRO Develop - C/ Cronista Carreres, 13. 46003 Valencia
- Clemente Pérez, José Antonio – jclemente@prodevelop.es - PRO Develop
- Centella Fuster, Víctor - vcentella@prodevelop.es –PRO Develop
- Vicent Camisón, Juan – jcamison@prodevelop.es – PRO Develop

Mots clés/Key words: Artificial Intelligence, in-situ data, satellite data, Copernicus, CAMS, AQI, pollution forecast, GEOSS

Proactive detection of pollution episodes with AI by exploiting in-situ data, satellite data and port activity

ABSTRACT:

Port of the Future is a recent concept referring to a next generation port able to cope with market challenges by incorporating latest technologies in an effective and efficient way, balancing business revenues and environmental impact. Such a complex task requires real engagement from port operators and stakeholders, especially for big ports and nearby cities where joint strategies are crucial for climate change adaptation and mitigation.
Nowadays, considering the availability of historical data and advances in machine learning techniques, it is possible to analyse and exploit them using artificial intelligence algorithms, including explainability (XAI).

Since 2013, Air Quality has been leading the list of the 10 top environmental priorities for ESPO (European Sea Ports Organization), seconded in 2020 by Climate Change, year when ESPO joined the PIANC leaded *Think Climate* coalition.

Here we present a Transport Management scenario supporting the Balearic Port Authority in environmental monitoring tasks and operations, under the on-going European H2020 research project EIFFEL[1] (GA 101003518). Three use cases where defined to assess the role of GEOSS as the default digital portal for building climate change mitigation applications.

Using in-situ and satellite data the web application can: (1) gather in-situ data from air quality monitoring (NO_2, SO_2, O_3, PM, etc.), (2) predict with ML techniques values for 24 - 48 - 72 upcoming hours, (3) compare predictions with Copernicus Atmosphere Monitoring Service (CAMS), even with lower resolution (10x10 km), (4)Integrate Sentinel 5P imagery (5.5x3.5km) for offshore areas, (5) provide a map-based user interface (UI) where stakeholders can access and manage data and predictions.

Challenges of on-going work are (c1) integration of different data sources, especially port activity and road traffic to establish correlations and design mitigation policies, (c2) design a mechanism for spatial super-resolution of satellite imagery to support use cases and (c3) reduce the bias of satellite readings vs. ground data

Main results that can be presented so far: (r1) tool is able to predict pollutants concentration and compare it to CAMS forecasts, (r2) Super Resolution (SR) mechanism produces results with 5 days delay due to the ERA5 tool availability, (r3) SR can train the model, useful for use case 2 aims of assessing the pollution in a regional offshore area without available in situ data (r4) SR accuracy up to 63,6% vs. AQ ground stations, lower this high BIAS could exceed the timeframe of the EIFEEL research project and could mean another dedicated research.

RÉSUMÉ:

Le port du futur est un concept récent qui fait référence à un port de nouvelle génération capable de faire face aux défis du marché en intégrant les dernières technologies de manière efficace et efficiente, tout en équilibrant les recettes commerciales et l'impact environnemental. Une tâche aussi complexe nécessite un réel engagement de la part des opérateurs portuaires et des parties prenantes, en particulier pour les grands ports et les villes voisines où des stratégies communes sont cruciales pour l'adaptation au changement climatique et l'atténuation de ses effets. Aujourd'hui, compte tenu de la disponibilité des données historiques et des progrès des techniques d'apprentissage automatique, il est possible de les analyser et de les exploiter à l'aide d'algorithmes d'intelligence artificielle, y compris l'explicabilité (XAI).

Depuis 2013, la qualité de l'air est en tête de la liste des 10 priorités environnementales de l'ESPO (Organisation européenne des ports maritimes), suivie en 2020 par le changement climatique, année où l'ESPO a rejoint la coalition *Think Climate* dirigée par l'AIPCN.

Nous présentons ici un scénario de gestion du transport soutenant l'autorité portuaire des Baléares dans ses tâches et opérations de surveillance environnementale, dans le cadre du projet de recherche européen H2020 en cours EIFFEL[1] (GA 101003518). Trois cas d'utilisation ont été définis pour évaluer le rôle du GEOSS en tant que portail numérique par défaut pour la création d'applications d'atténuation du changement climatique.

En utilisant des données in-situ et satellitaires, l'application web peut : (1) recueillir des données in situ à partir de la surveillance de la qualité de l'air (NO2, SO2, O3, PM, etc.), (2) prédire avec des techniques ML les valeurs pour les 24 - 48 - 72 heures à venir, (3) comparer les prédictions avec le Copernicus Atmosphere Monitoring Service (CAMS), même avec une résolution plus faible (10x10 km), (4) intégrer l'imagerie Sentinel 5P (5.5x3.5km) pour les

zones offshore, (5) fournir une interface utilisateur basée sur une carte où les parties prenantes peuvent accéder et gérer les données et les prédictions.

Les défis du travail en cours sont (c1) l'intégration de différentes sources de données, en particulier l'activité portuaire et le trafic routier pour établir des corrélations et concevoir des politiques d'atténuation, (c2) la conception d'un mécanisme de super-résolution spatiale de l'imagerie satellitaire pour soutenir les cas d'utilisation et (c3) la réduction du biais des relevés satellitaires par rapport aux données au sol.

Principaux résultats pouvant être présentés à ce jour: (r1) l'outil est capable de prédire la concentration des polluants et de la comparer aux prévisions CAMS, (r2) le mécanisme de super résolution (SR) produit des résultats avec un retard de 5 jours dû à la disponibilité de l'outil ERA5, (r3) la SR peut entraîner le modèle, ce qui est utile pour le cas d'utilisation 2 qui vise à évaluer la pollution dans une zone offshore régionale sans données in situ disponibles (r4) la précision de la SR atteint 63,6% par rapport aux stations terrestres AQ, une BIAS moins élevée pourrait dépasser le calendrier du projet de recherche EIFEEL et pourrait nécessiter une autre recherche dédiée.

Article :

1. Introduction and objectives

Since 2013, Air Quality has been leading the list of the 10 top environmental priorities for ESPO (European Sea Ports organization), seconded in 2020 (when ESPO joined *Think Climate*) by Climate Change and Energy Efficiency.
In 2019 Balearic Port Authority (BPA) deployed a network of air quality sensors consisting of twenty five stations between the 5 ports it manages (Alcúdia, Eivissa, Palma, Maó and la Savina), recording readings every ten minutes of parameters such as CO, NO_x, SO_2, PM_1, $PM_{2.5}$, PM_{10}, O_3, as well as climate conditions like wind speed and direction, relative humidity and atmospheric pressure.
Nevertheless, this amount of data is restricted to the port area and is insufficient for building pollution patterns or abnormal predictions outside the portal area or the nearby city, whereas local and regional variability must be considered.
The aim of the EIFFEL pilot 3 is to develop a web application to monitor and predict the impact of port activity on climate change and air quality at a regional scale, using AI, machine learning (ML), data mining and geospatial techniques. Real-time in situ measurements will be combined with historical data from GEOSS, as well as other external sources.
The web application aims to support the BPA in its daily environmental monitoring tasks to minimize or reduce its environmental impact by providing insights on emissions of the main air pollutants (NO_2, SO_2, O_3, PM_{10} and $PM_{2.5}$), and its impact on the nearby city and protected areas in maritime routes, as well as optimizing the berthing allocation
The climate change mitigation objectives for this pilot are:

- The monitoring and prediction of pollutions episodes at BPA generated by port activities at the sea-port area and the city of Palma de Mallorca.
- Determination of whether the origin of pollution episodes in port / city is caused by port activity.
- Improvement in the planning and optimization of port activities (cargo, loading/unloading operations, vessels and land traffic) considering pollution episodes forecasts, becoming BPA more efficient ports.
- Improved decision making and minimization of the carbon footprint of the port calls activity and the impact on the city.
- The app will also offer a map-based UI with temporal AQI (following OGC, INSPIRE standards), customised dashboards with pollution KPIs, notifications and data reports for BPA, vessel companies, local authorities and citizens among other stakeholders. This pilot will take advantage of data sets, results and collaboration with other projects sharing complementary objectives, such as PIXEL (H2020) and I2panema (ITEA3).
- The app will be integrated with the cognitive search tool, one of the transversal tools developed in aid of all project pilots.

2. Description of the work and tool development

2.1 Case study area

The pilot 3 study area covers the five ports managed by BPA.
With more than 50,000 annual port calls in 2018, BPA is the leading Spanish Port Authority in number of port calls, and the 2nd in the European ranking. With 2,400,000+ cruise passengers and 6,000,000 regular line passengers the BPA is also the leader in passenger traffic in Spain.
As introduced the BPA is already gathering real-time measurements of ground pollution, with valid historical data since September 2019.

2.2 Use cases and scenario

Three use cases have been defined:

- Use case 1: Analysis of atmospheric pollution in Palma. This use case focuses on the study of the correlation between port activity and pollution episodes in the city/port, analysing the following air quality parameters: NO_2, SO_2, O_3 and PM_{10} and $PM_{2.5}$, and other external parameters such as wind and traffic in the port/city and nearby areas using the XAI models developed.
- Use case 2: Atmospheric Emissions study in The Freus. The objective of this use case is the monitoring of air quality and emissions from vessels in the Freus area (route between Eivissa and la Savina ports).
- Use case 3: Berths allocation optimisation. This use case aims to: (i) provide advice on berth allocation based on lessons learned (monitoring and forecasting), for both vessel and vehicle emissions on board, (ii) provide insights for optimizing berth allocations and traffic routes from/to port are affecting the least number of people and (iii) identify which vessels and/or berths could change to a less impacting propulsion system.

2.3 Pilot App development

The work for Pilot 3 App development has been focussed both on data gathering and processing, AI pollution forecast models building as well as on the web UI development that supports the decision-making by data and forecast visualization.

2.4 Input datasets and models

2.4.1 Datasets : identified input datasets (sometimes aggregated):

Private data (PR)	Port Operations Data from the Port Management System: real time and historical port calls with AIS positions since 2016.
	Air quality in-situ sensors data from BPA: measures atmospheric pollutant parameters every 10 minutes. Historical data from 2019.
Open data (OD)	AEMET: Weather information from the Spanish Meteorological Agency.
	Regional Government Air Quality reference stations (SO_2, NO, NO_2, PM_x, DD, VV, TMP)
	Traffic Intensity around Palma's port: from Palma city council
GEOSS related data (GEO)	Sentinel 5P data: TROPOMI sensor through the Copernicus Sentinel Access Hub.Sentinel- 5P TROPOMI has available within 3 hours NRT (near real time) after reading and a spatial resolution of 7 x3.5 km to gather NO_2, SO_2, O_3 air quality parameters in NetCDF format. The interpretation of Sentinel- 5P data e.g. is complex and pixel data with q-value > 0.75 has been used for areas with no sensor in-situ data as "informative values" but cannot be compared with in-situ data.
	Copernicus Atmosphere Monitoring System (CAMS) data to assess and compare forecasts. The product Regional Forecasts, Model Ensemble is beign used for concentration predictions of parameters NO_2, SO_2, $PM_{2.5}$$O_3$,and PM_{10} with a spatial resolution of 10x10 km by accessing to https://ads.atmosphere.copernicus.eu/api-how-to

Table 1: Input datasets

2.4.2 Models applied for data processing

Predictions will be made on the basis of in-situ data from the stations deployed in BPA's ports.

For the predictions, XAI models are applied aiming is to develop predictions that not only show forecasts but also what if scenarios, to understand at a glance what would happen with the predictions if the value of such a variable increases or decreases.

In-situ data is received in temporary series, sets of data points collected or recorded at regular time intervals. These sequences are often used to analyse trends and make future predictions. In time series analysis, the temporal ordering of data points is crucial, requiring specialized methods for effective prediction.

An example of the time series used for a Palma station is shown here:

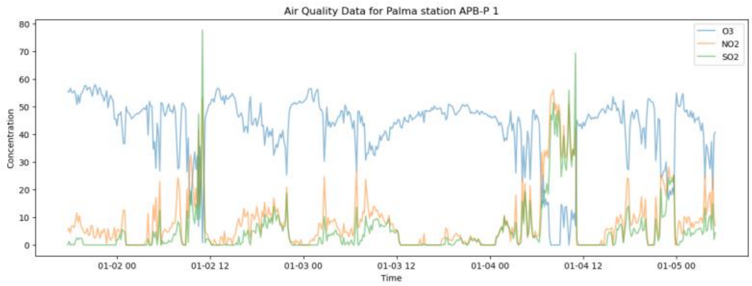

Figure 1: Air Quality data for Palma station APB-P1

For time series prediction, LSTM is used, a special kind of neural network designed for data sequences. LSTM works like a chain of memory cells processing each data point one by one. These cells have special 'gates' that decide what information to keep or forget, making LSTMs great at learning from the history of the series. Essentially, these cells remember important past information and use it to help make better predictions for future data points. LSTM is like a skilled note-taker: it remembers crucial details from earlier in the series to make smart guesses about next outcomes. This ability makes it well-suited for predicting future values in a time series, where the order and history matter.

To evaluate the performance of the predictive models, three key metrics were selected:

- **R^2 (R-squared)**: quantifies how well the model explains the variability in the data. Ranging from 0 to 1, a higher R^2 indicates a better fit, helping us understand the model's effectiveness across the entire range of the dataset.
- **RMSE (Root Mean Square Error)**: offers a tangible way to measure the model's average prediction error, translating it into the same unit as the original data. This is especially useful for interpreting the model's absolute accuracy.
- **RPIQ (Ratio of Prediction to Inter-Quartile Range):** focused on the model's ability to accurately capture outliers, RPIQ is calculated by dividing the inter-quartile range by the RMSE. A higher RPIQ value means the model is better at accounting for extreme values, which is crucial for datasets where outliers carry significant information, such as this one.

These metrics provide a balanced evaluation, with R^2 assessing overall model fit, RMSE quantifying the average error in understandable terms, and RPIQ evaluating the model's capability to capture outliers effectively.

Regardless of the metrics used to measure the accuracy of the models developed, the results obtained can be compared in two ways:

- Compare with CAMS outcome. The CAMS images are predictions that make use of the [2] ensemble model technique. That is, these predictions are based on the application of different models. With this comparison, the similarity or discrepancy between the model developed and the results given by CAMS could be observed.
- Compare with in-situ data. In this way it would be possible to compare what happened on 'X' day in the future with the predictions developed for the same day. Taking into account that in both cases the origin is the same: the in-situ data. By representing both data sets on the same axes, one could see if the model correctly predicted what happened.

ROAD TRAFFIC DATA ANALYSIS

Using city of Palma network of traffic intensity sensors an analysis of traffic surrounding the port area is being carried out with the aim to include the results in the XAI model developed. Producing traffic patterns and including them into the model as an external variable, it is expected to be able to deduct the road traffic contribution to the pollution models. This approach is due to the fact there are no real time data sources, instead an historical register is provided on demand. Steps:

1. Study traffic data independently to identify traffic profiles.
2. Include the data from the environmental stations and check for correlations between the congestion peaks and increases in the measured values for the pollutants. This is the current status of the study.
3. Include other additional parameters (previous correlation might be directly influenced by wind speed and direction). Next step.

EUROPEAN AQI EEA CALCULATION

The application offers AQI predictions made using satellite imagery (CAMS), at a number of points of interest (POI) to cover offshore areas with no in-situ, POI chosen are in use case 2 area.

Calculation of AQI predictions follow the [3] CAMS training hub notebook https://ecmwf-projects.github.io/copernicus-training-cams/proc-aq-index.html

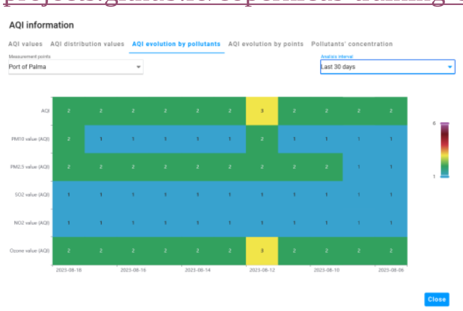

Figure 2: AQI evolution per pollutant and date

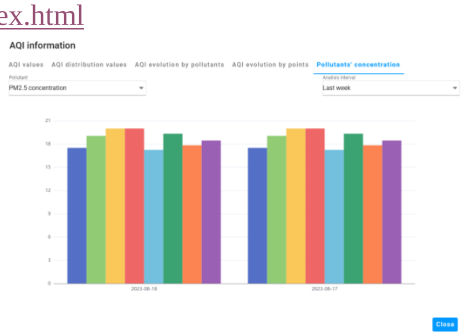

Figure 3: $PM_{2.5}$ concentration by date

2.5 Super Resolution mechanism

Spatial resolution augmentation of Sentinel 5p data is of particular interest for this pilot as the satellite imagery has higher resolution than required to implement the use cases.

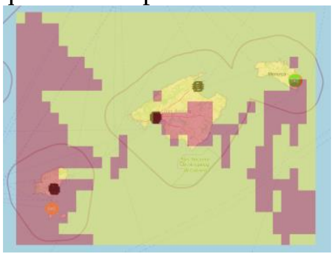

Figure 4: CAMS $PM_{2.5}$ levels in the study area during a Saharan dust intrusion episode

Sentinel 5p readings are gases concentrations of the whole tropospheric column, while in-situ sensors are measuring ground level concentrations. Both concentrations cannot be compared without further treatment. To estimate the ground level concentrations, it's been used Sentinel-5P satellite observations of NO_2, O_3, SO_2 and CO, meteorological data from ERA5 and GTOPO30 elevation data to train the XGBoost model with one year in situ ground measurements from over 3000 stations in Europe provided by the European Environmental Agency (EEA). Based on the first validation results, the NO_2 and O_3 models perform well and achieved mean absolute errors (MAE) close to the state-of-the-art methods. On the other hand, the validation results for SO_2 and CO ground level concentrations are not so well and with data gaps due to masking - exclusion of pixels with low quality values.

Therefore, the focus is on NO_2 products and results are being evaluated at the moment of writing this paper in order to assess if they can have value in current or future processes and algorithms.
The measurements from a dense network of air quality ground stations can be potentially acquired real-time and used to infer ground level concentrations of various air pollutants from Sentinel 5p.
Through ML-based methods combining in-situ ground measurements with regression models information about the spatial distribution of air pollutants could be derived. DL-based algorithms, trained with a large number of in-situ measurements collected from AirBASE (Air Quality e-Reporting data repositories maintained by the European Environmental Agency - EEA) can also be applied when in-situ measurements are not available, as required in use case 2.Additionally, model-based methods, will also be tested that can infer ground-level concentrations from satellite data by using a chemical transport model instead of in-situ measurements.

3. Conclusion and perspectives :

- The tool is able to predict pollutants concentration and compare it to CAMS forecasts
- Super Resolution (SR) mechanism produces results with 5 days delay due to the ERA5 tool availability. SR can train the model, useful for use case 2 aims of assessing the pollution in a regional offshore area without available in situ data
- SR accuracy up to 63,6% vs. AQ ground stations, lower this high BIAS could be another project
- New launching of Sentinel 4 and Sentinel 5 are promising in terms of improving the temporal resolution and detection accuracy

REFERENCES

[1] EIFFEL: https://www.eiffel4climate.eu/

[2] https://towardsdatascience.com/ensemble-models-5a62d4f4cb0c

[3] Copernicus Atmosphere Monitoring Service (CAMS) Data Tutorials: https://ecmwf-projects.github.io/copernicus-training-cams/intro.html