

DESIGN PROJECT – SIE 2026

Study of the PM2.5 and PM10 concentrations and in depth analysis of existing Google API data compared to reference concentrations values and spatial and temporal study of the distribution of their difference.

Student
Lucas Massot

Company
Aerospec

Supervisors
Andrea Baccarini, Satoshi Takahama

CONTEXT

Aerosol pollution and fine particles represent a double threat, both for our health and our climate. Studies have linked important concentrations to premature deaths and have revealed their important role in climate change.

In this context we want to be able to estimate these concentrations in any given area. To answer this, Google launched an API providing these values around the world, relying on satellites, diffusion models and machine learning approaches

OBJECTIVE

We study the correctness of the Google API by comparing it to reference values, in particular the distribution of their difference and try to correct the API data in a deterministic approach to better fit the reference data.

The project's final deliverable is an automatic pipeline that outputs a report on the distribution of the concentrations, the quantification of the difference between the two, and the process of correcting Google data to fit the reference

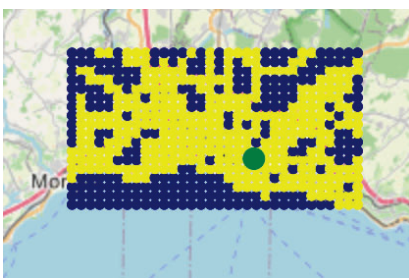
METHODOLOGY

- Definition of the anomaly : $\delta := \text{Google} - \text{Ref}$
- Polynomial fitting of the expression of the Google API using the Reference : $\text{Google} = \text{slope} \times \text{Ref} + \text{intercept}$
- Correction of the anomaly using the clustering and the polynomial fitting : $\text{Corrected} = \frac{\text{Google} - \text{intercept}}{\text{slope}}$

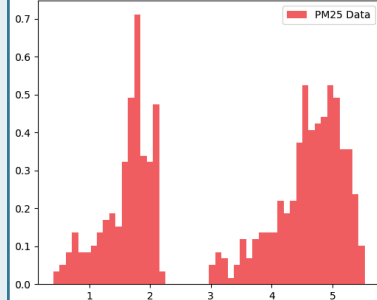
INITIAL RESULTS

- Multimodal distribution of the anomaly
- Clusters on the anomaly Map

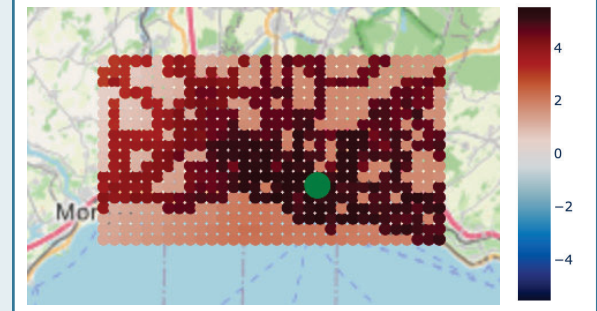
Clustering Map of the PM2.5 Anomaly in Lausanne



Statistical Distribution of the Anomaly



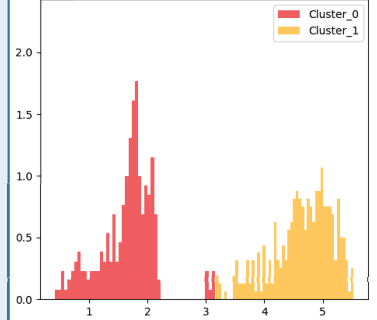
Initial PM2.5 Anomaly Map in Lausanne



CLUSTERING RESULTS

- Distinction of two anomaly clusters
 - Separation of the modes of the distribution achieved by clustering
 - Linear relationship between Google and Reference concentrations
- possibility to fit a degree 1 polynomial

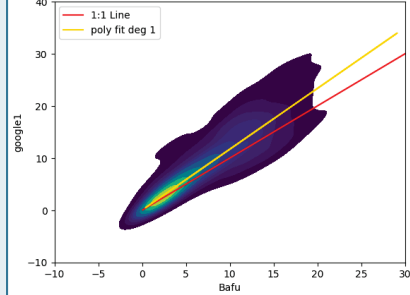
Anomaly PM2.5 Clusters



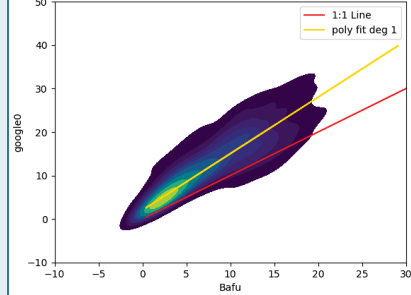
CORRECTION RESULTS

- Overall reduction in the anomaly
 - Disappearance of the Clusters
- Spatial differences only based on distance from reference

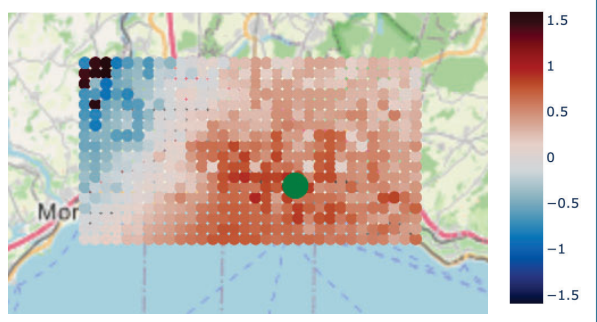
Cluster 1



Cluster 0



Corrected PM2.5 Anomaly Map in Lausanne



CONCLUSION

- Pipeline shows separation of anomaly into spatial clusters, possibly linked to land use
- Scatter plots reveal somewhat linear relationship between the reference and the Google
- Optimisation of polynomial modelling allows correction of the anomaly