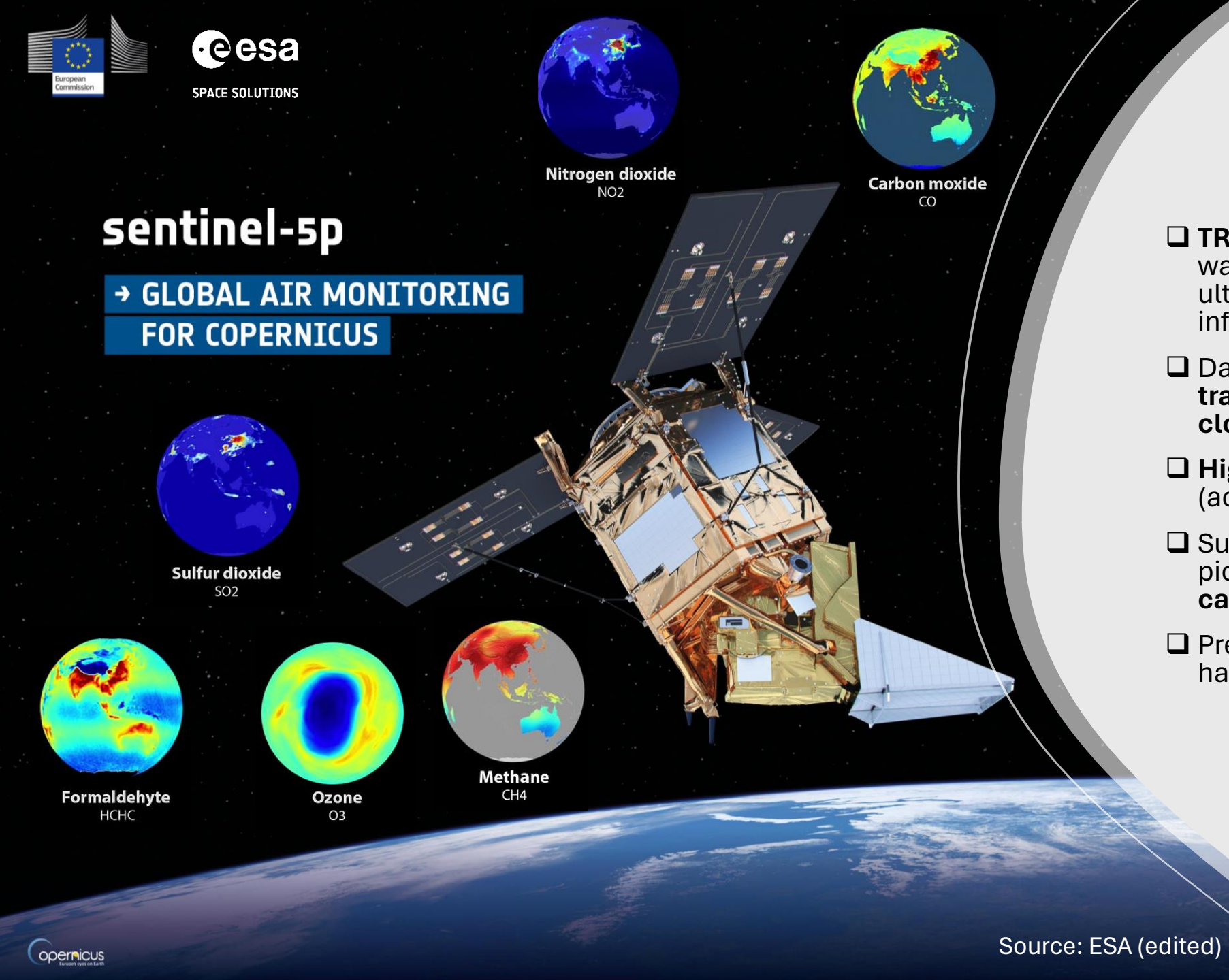
POLITECNICO
MILANO 1863

Gap-filling techniques for air quality earth observations

Giorgia De Moliner – POLIMI PhD candidate

ARIANET



- **TROPOMI** spectrometer covering wavelength bands between the ultraviolet and the shortwave infrared.
- Data products: several **atmospheric trace gases**, as well as **aerosol** and **cloud** properties
- **High spatial resolution**: 3.5 x 5.5 Km (across x along track)
- Sun-synchronous polar orbit → a picture of **the whole earth is captured daily**
- Precursor of Sentinel-5; Level 2 data has been available since April **2018**

Source: ESA (edited)

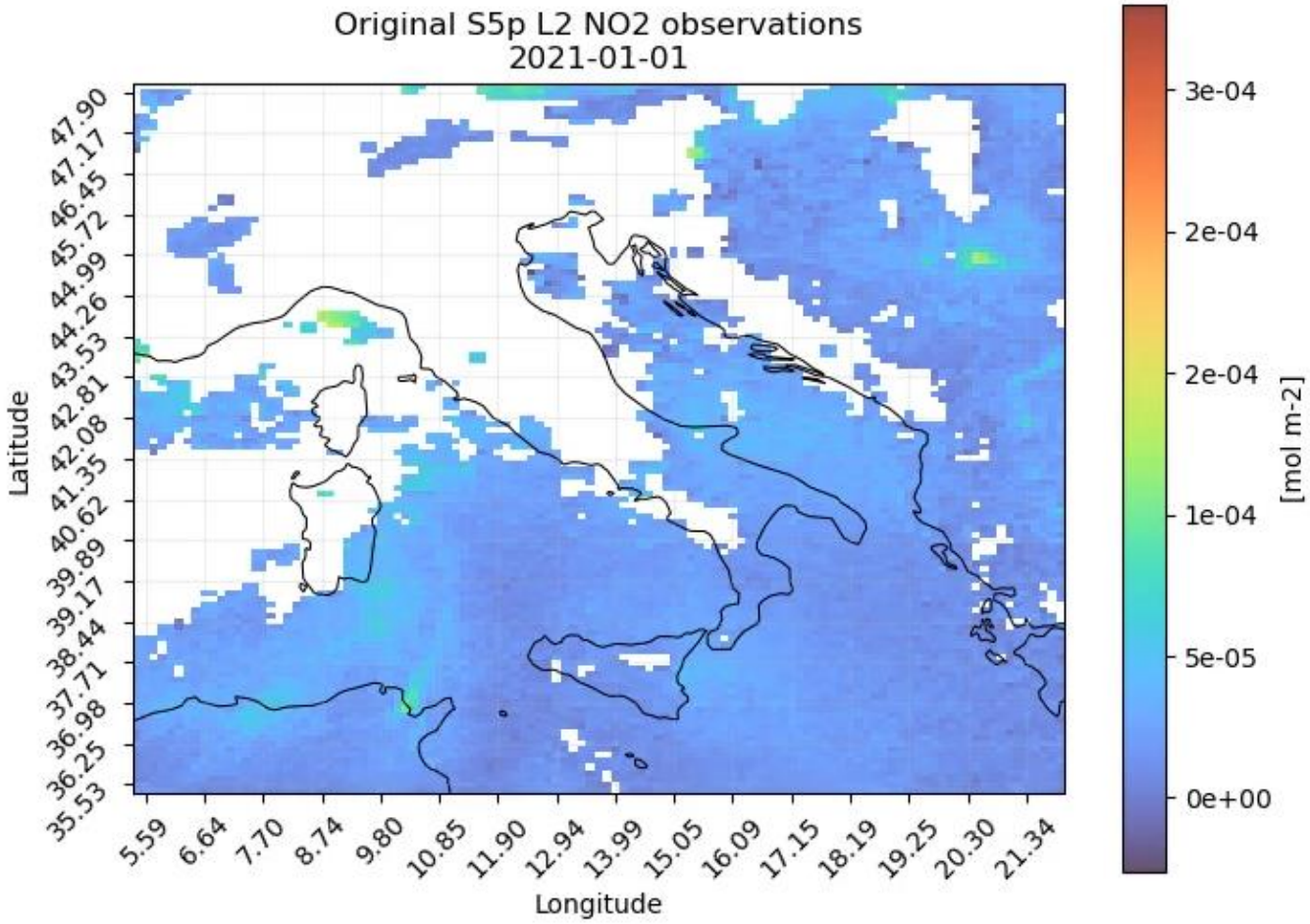
Missing values

Complex, large-scale, and unavoidable missing values that **differ from one species to another**

Due to **sensor malfunction** or sensor **limitations**, presence of **thick clouds** or **snow/ice**, **low-qaulity retraivals**

Missing values can **hinder** further, **trend analysis** and can **obscure physical dependencies** between observations

Among others, initial gap-filling is needed to be able **to feed EO to a ML model** and use it as a predictor

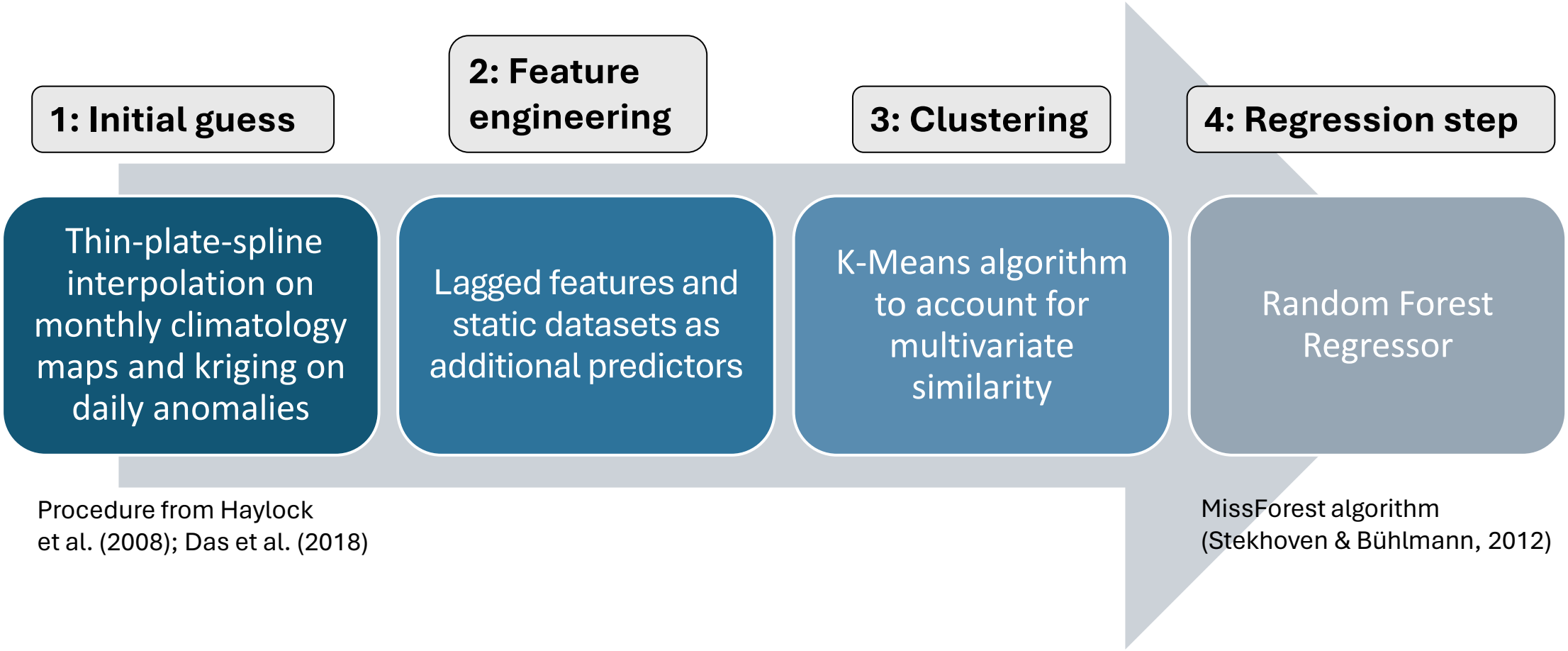


CLIMFILL Available on GitHub at: Climfill

[1] Bessenbacher, V., Seneviratne, S. I., and Gudmundsson, L.: CLIMFILL v0.9: a framework for intelligently gap filling Earth observations, Geosci. Model Dev., 15, 4569–4596, https://doi.org/10.5194/gmd-15-4569-2022, 2022.
[2] Bessenbacher, V., Schumacher, D. L., Hirschi, M., Seneviratne, S. I., & Gudmundsson, L. (2023). Gap-filled multivariate observations of global land–climate interactions. Journal of Geophysical Research: Atmospheres, 128, e2023JD039099

- ❑ Independent of physical models and **entirely data-driven**
- ❑ **Combines** state-of-the-art **spatial interpolation** with a **machine learing model**
- ❑ No 'donor' involved: all input maps have missing values and the algorithm can digest complex patterns of missingness in **multivariate EO → mutual gap filling**
- ❑ Computationally feasible for **large datasets**

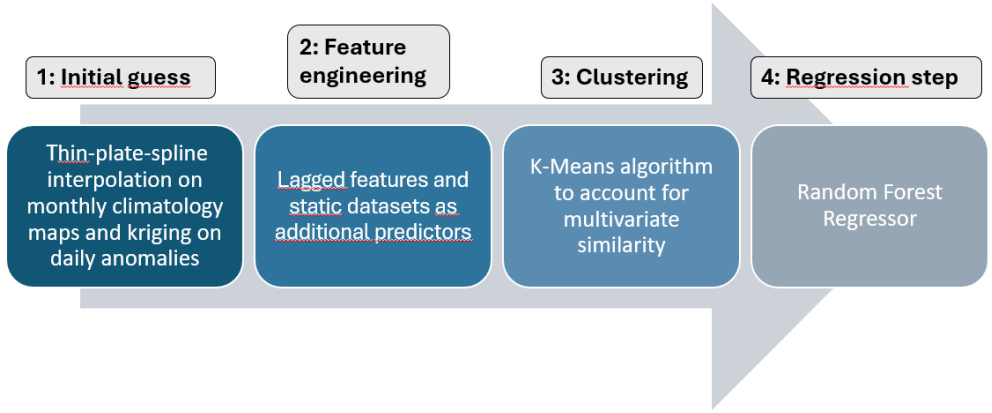
Methodology

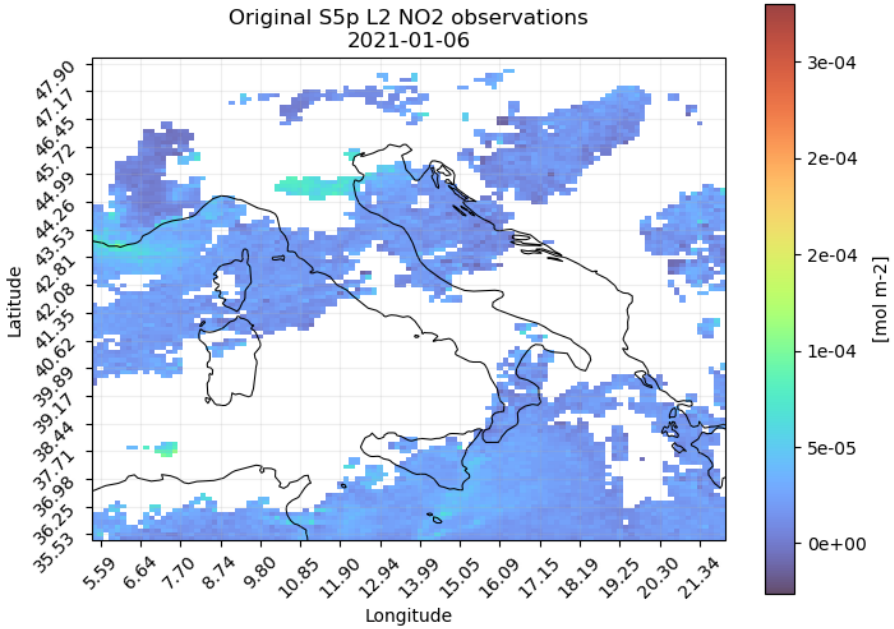


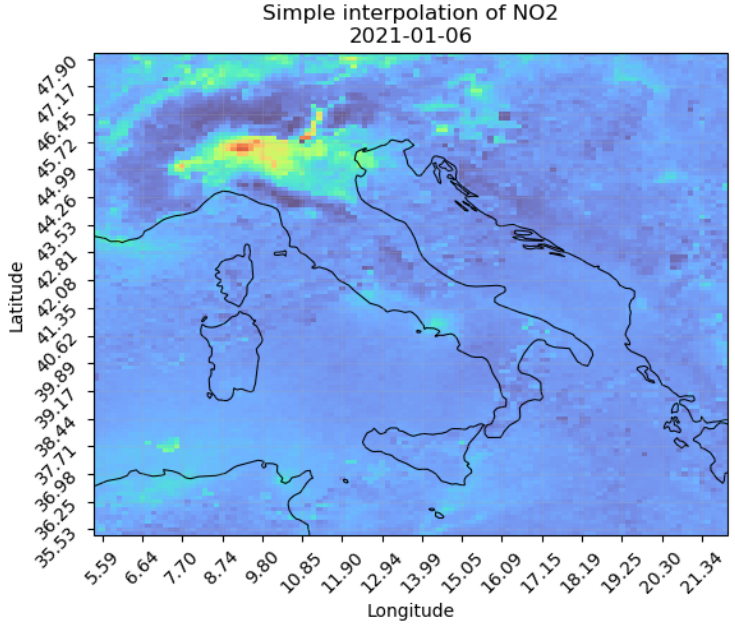
[1] Bessenbacher, V., Seneviratne, S. I., and Gudmundsson, L.: CLIMFILL v0.9: a framework for intelligently gap filling Earth observations, Geosci. Model Dev., 15, 4569–4596, https://doi.org/10.5194/gmd-15-4569-2022, 2022.

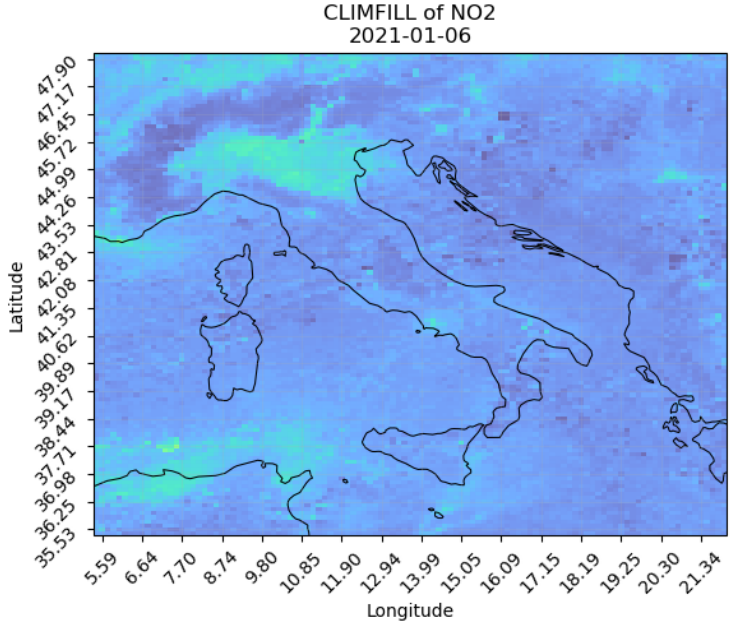
Preliminary results

One month of Sentinel-5p gap-filled NO2, SO2, CO, HCHO observations









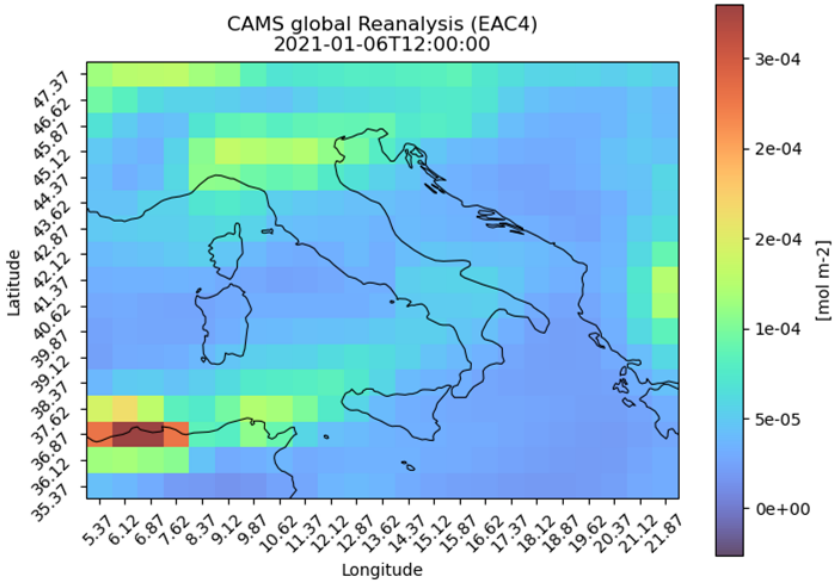
0: Gappy map → **1: Initial guess** → **4: Regression step**

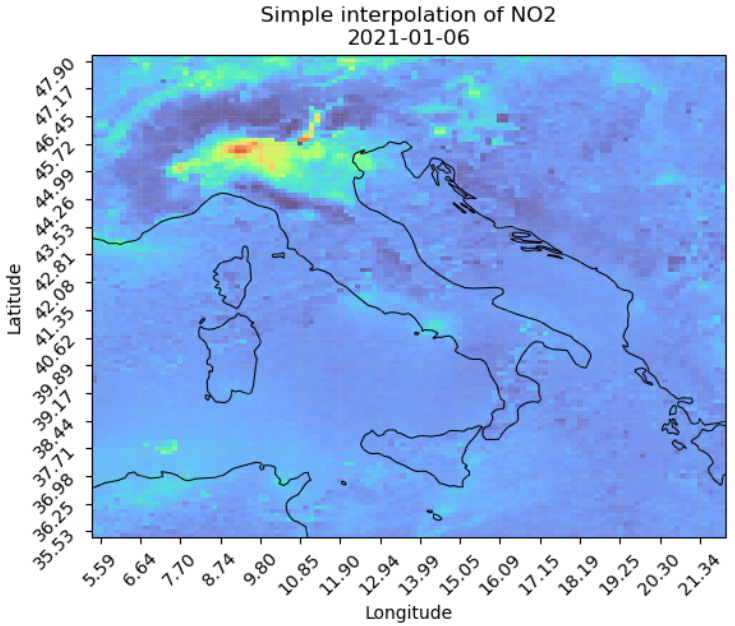
Preliminary results

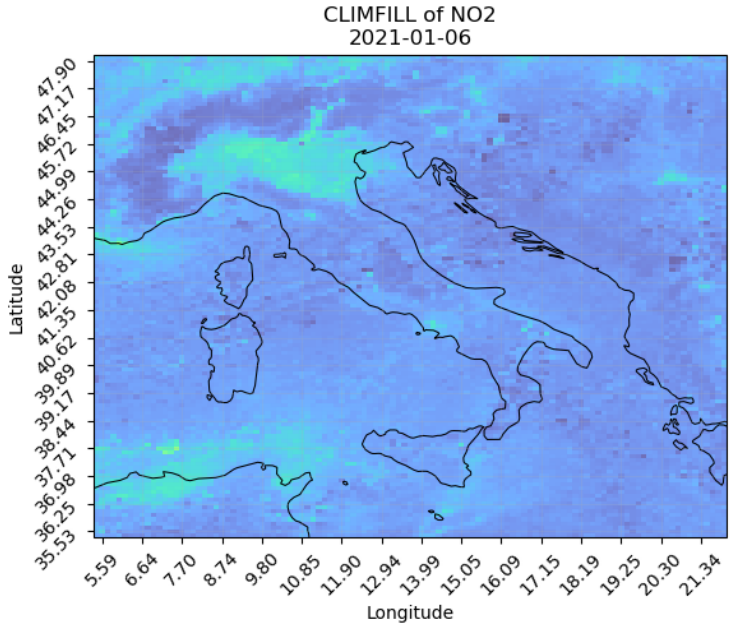
One month of Sentinel-5p gap-filled NO2, SO2, CO, HCHO observations

Mean RMSE	NO2	SO2	CO	HCHO
Random Forest	6.87E-06	2.97E-04	7.62E-04	2.92E-05
Simple Interpolation	6.42E-06	3.05E-04	7.19E-04	2.94E-05
RF added value	6.00%	-2.60%	5.58%	-0.63%

Validation performed by iteratively adding **fictitious but realistic missing values** over 10 different validation sets







1: Initial guess → **4: Regression step**

Future work

- ❑ **Extend the training dataset** and benchmark the results→ at least one full year of observations
- ❑ **Include other** physically relevant **gappy variables**, ideally with missingness patterns that only partially overlap
- ❑ **Test different ML models** (Gradient Boosting, Partial Convolutional Neural Networks, …)
- ❑ **Embed the technique in Mal-Air** to allow earth observations as predictors

POLITECNICO
MILANO 1863

Gap-filling techniques for air quality earth observations

Giorgia De Moliner – POLIMI PhD candidate

ARIANET