

CO2MVS RESEARCH ON SUPPLEMENTARY OBSERVATIONS



D2.1 List of CO₂, NO₂ and CO hot spot locations for the year 2021 identified in satellite observations

| | |
|---|--|
| Due date of deliverable | December 31, 2023 |
| Submission date | March 2024 |
| File Name | CORSO-D2.1-V1.2 |
| Work Package /Task | WP2/Task 2.1 |
| Organisation Responsible of Deliverable | UT3 |
| Author name(s) | T. Doumbia, C. Granier, G. Kuhlmann, M. Guevara |
| Revision number | 1.2 |
| Status | Issued |
| Dissemination Level / location | Public www.corso-project.eu |



The CORSO project (grant agreement No 101082194) is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Commission. Neither the European Union nor the granting authority can be held responsible for them.

1 Executive Summary

This work has been done as part of WP2, which focuses on the use of co-emitted species to better estimate anthropogenic emissions in the future CO₂MVS capacity. The anthropogenic signal of the observations of NO₂ from satellite is generally much better than that of CO₂.

This report contributes to the objectives of WP2, as it describes the methodology and results of the use of the TROPOMI satellite observations of tropospheric NO₂ columns obtained from the Copernicus Sentinel-5P satellite for identifying hotspot locations. Areas with high NO₂ concentrations can indicate regions of high anthropogenic activity and pollution, and they will be used as proxies for the locations of hotspot of CO₂ and CO emissions. A CSV file is attached to this deliverable, which contains hotspot locations worldwide, specifically the latitude and longitude coordinates of local NO₂ maxima in TROPOMI maps.

The methodology described in this report, which as been developed as part of CORSO WP2, will be applied to the observations from geostationary observations from the GEMS satellite over Asia, TEMPO over North America and in 2024-2025 Sentinel-4 in Europe.

Table of Contents

| | | |
|-------|--|----|
| 1 | Executive Summary | 2 |
| 2 | Introduction | 4 |
| 2.1 | Background..... | 4 |
| 2.2 | Scope of WP2..... | 4 |
| 2.2.1 | Objectives of this deliverable | 5 |
| 2.2.2 | Work performed in this deliverable..... | 5 |
| 2.2.3 | Deviations and counter measures..... | 5 |
| 2.3 | Project partners involved in this deliverable: | 5 |
| 3 | Methodology and results | 6 |
| 3.1 | Gathering the satellite observations | 6 |
| 4 | Future work..... | 12 |
| 5 | References | 12 |

2 Introduction

2.1 Background

To enable the European Union (EU) to move towards a low-carbon economy and implement its commitments under the Paris Agreement, a binding target was set to cut emissions in the EU by at least 40% below 1990 levels by 2030. European Commission (EC) President von der Leyen committed to deepen this target to at least 55% reduction by 2030. This was further consolidated with the release of the Commission's European Green Deal on the 11th of December 2019, setting the targets for the European environment, economy, and society to reach zero net emissions of greenhouse gases in 2050, outlining all needed technological and societal transformations that are aiming at combining prosperity and sustainability. To support EU countries in achieving the targets, the EU and European Commission (EC) recognised the need for an objective way to monitor anthropogenic CO₂ emissions and their evolution over time.

Such a monitoring capacity will deliver consistent and reliable information to support informed policy- and decision-making processes, both at national and European level. To maintain independence in this domain, it is seen as critical that the EU establishes an observation-based operational anthropogenic CO₂ emissions Monitoring and Verification Support (MVS) (CO2MVS) capacity as part of its Copernicus Earth Observation programme.

The CORSO research and innovation project will build on and complement the work of previous projects such as CHE (the CO₂ Human Emissions), and CoCO₂ (Copernicus CO₂ service) projects, both led by ECMWF. These projects have already started the ramping-up of the CO2MVS prototype systems, so it can be implemented within the Copernicus Atmosphere Monitoring Service (CAMS) with the aim to be operational by 2026. The CORSO project will further support establishing the new CO2MVS addressing specific research & development questions.

The main objectives of CORSO are to deliver further research activities and outcomes with a focus on the use of supplementary observations, i.e., of co-emitted species as well as the use of auxiliary observations to better separate fossil fuel emissions from the other sources of atmospheric CO₂. CORSO will deliver improved estimates of emission factors/ratios and their uncertainties as well as the capabilities at global and local scale to optimally use observations of co-emitted species to better estimate anthropogenic CO₂ emissions. CORSO will also provide clear recommendations to CAMS, ICOS, and WMO about the potential added-value of high-temporal resolution ¹⁴CO₂ and APO observations as tracers for anthropogenic emissions in both global and regional scale inversions and develop coupled land-atmosphere data assimilation in the global CO2MVS system constraining carbon cycle variables with satellite observations of soil moisture, LAI, SIF, and Biomass. Finally, CORSO will provide specific recommendations for the topics above for the operational implementation of the CO2MVS within the Copernicus programme.

2.2 Scope of WP2

The work presented in this report is part of WP2 of CORSO, which deals with "Use of co-emitted species (correlations, improved emission ratios, uncertainties) in data assimilation systems. The aim of WP2 is to improve the use of observations of co-emitted species (CO₂, NO₂, CO) to better estimate anthropogenic emissions in the future CO2MVS capacity. This is based on the recognition that anthropogenic CO₂ emissions cannot completely be constrained with CO₂ concentration observations alone, and the detectability of the anthropogenic signal of co-emitted species is often much better than that of CO₂. For the emission estimation development at local scale, this WP focuses on the development of methods to increase the

accuracy of annual CO₂ emission estimates of hot spots, industrial and urban areas by integrating satellite observations of co-emitted species (NO₂ and CO) in data assimilation systems. Since CO₂ satellite observations are temporally sparse (even with the future CO2M constellation), temporal sampling biases are a significant source of uncertainty in annual CO₂ emission estimates of hot spots. Co-emitted species such as CO and NO₂ are and will be available at sub-diurnal temporal coverage from current and future LEO and GEO satellites. They can therefore be used to improve the constraint on the temporal variability of CO₂ emissions and hence for reducing the uncertainty in annual estimates. The local and regional studies will focus on three regions: Europe, Africa, and Southeast Asia.

Among the objectives of this WP, the work presented in this report focuses on the identification of emission hot spots in NO₂ observations from LEO and GEO satellites for validating bottom-up inventories, on the improvement of data-driven methods for detecting and quantifying hot spot emissions and their application to quantify NO_x and CO emissions from LEO and GEO satellites.

2.2.1 Objectives of this deliverable

The objectives of this deliverable are to improve the accuracy of annual CO₂ emission estimates of hotspots power plants using satellite observations of the NO₂ co-emitted species. Among the multiple satellite products that will be considered in this WP2, we have used TROPOMI data from the Copernicus Sentinel-5P satellite and we will start to use soon the observations from the Geostationary Environment Monitoring Spectrometer (GEMS) data over East and Southeast Asia. This work directly contributes to the provision of a dataset containing hotspot locations worldwide (i.e., latitude and longitude coordinates of local NO₂ maxima in TROPOMI maps).

The title of this deliverable is “List of CO₂, NO₂ and CO hot spot locations for the year 2021 identified in satellite observations”. The deliverable contains two files, this report which describes the methodology and the list of hot spots, which is given separately as a CSV file.

2.2.2 Work performed in this deliverable

This deliverable was accomplished through a series of activities detailed in Section 3:

- Gathering the satellite observations
- Formulating the methodology for hotspots identification
- Compiling the file containing information on the identified hotspots

2.2.3 Deviations and counter measures

As discussed with the CORSO coordination, this deliverable has been slightly delayed due to recruitment delays and in reviewer availability.

2.3 Project partners involved in this deliverable:

| Partners | |
|---|----------|
| BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION | BSC |
| UNIVERSITE PAUL SABATIER TOULOUSE III | UT3-CNRS |
| EIDGENOSSISCHE MATERIALPRUFUNGS- UND FORSCHUNGSANSTALT | EMPA |

3 Methodology and results

3.1 Gathering the satellite observations

The satellite data used in this work consist of gridded tropospheric NO₂ data (Level 3 data) obtained from the S5P-PAL (Sentinel 5P Product Algorithm Laboratory) data portal, at <https://maps.s5p-pal.com/>. The product that can be downloaded from this website is available at a resolution of about 0.022° x 0.022° (about 2 km x 2 km, representing 134,217,728 pixels, as shown in Figure 1, which shows the dump of a file downloaded from this site.

```

dimensions:
  time = 1 ;
  latitude = 8192 ;
  longitude = 16384 ;
variables:
  double datetime_start(time) ;
    datetime_start:description = "start time of the measurement" ;
    datetime_start:units = "days since 2000-01-01" ;
  float tropospheric_NO2_column_number_density(time, latitude, longitude) ;
    tropospheric_NO2_column_number_density:description = "tropospheric vertical column of NO2" ;
    tropospheric_NO2_column_number_density:units = "umol/m2" ;
  float tropospheric_NO2_column_number_density_uncertainty(time, latitude, longitude) ;
    tropospheric_NO2_column_number_density_uncertainty:description = "uncertainty of the tropospheric vertical column of NO2 (standard error)" ;
    tropospheric_NO2_column_number_density_uncertainty:units = "umol/m2" ;
  double datetime_stop(time) ;
    datetime_stop:units = "days since 2000-01-01" ;
  int count(time) ;
  float weight(time, latitude, longitude) ;
  double latitude(latitude) ;
    latitude:units = "degree_north" ;
  double longitude(longitude) ;
    longitude:units = "degree_east" ;

// global attributes:
:Conventions = "HARP-1.0" ;
:datetime_start = 8815.04893849769 ;
:datetime_stop = 8829.03878580102 ;

```

Figure 1: Dump of a NO₂ TROPOMI file at 0.022° x 0.022° resolution.

The observations, averaged over the year 2021, have been acquired in both netcdf and image formats. Both files feature a high spatial resolution of approximately 0.022° x 0.022° (about 2 km x 2 km, representing 134,217,728 pixels), making them well-suited for investigating NO_x emissions hotspots from space globally (Figure 2).

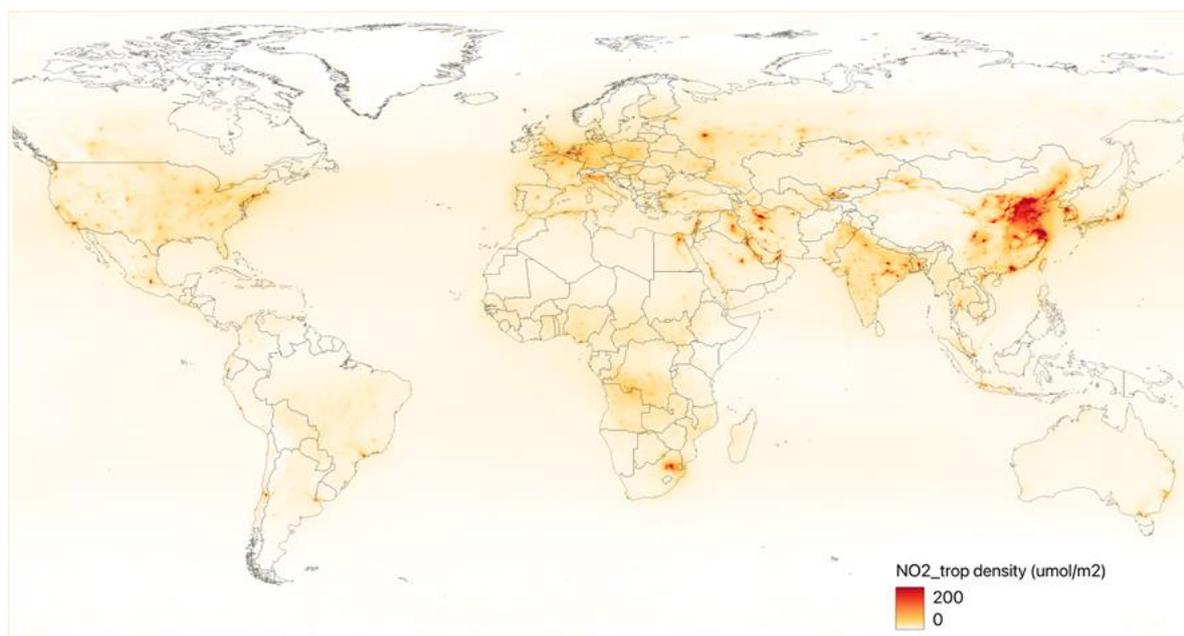


Figure 2: TROPOMI NO₂ tropospheric column density annual average in μmol/m² for the year 2021.

3.2 Methodology to identify hotspots

In this work, the objective is to detect high-probability hotspots in TROPOMI NO₂ tropospheric column density ($\mu\text{mol}/\text{m}^2$). To accomplish this goal, three different statistical methods were utilized:

- One straightforward approach called the 2*sigma (2*STD) method, which is illustrated in Figure 3: the standard deviation (STD) and the mean values are calculated at the global level. An upper bound (UB) is calculated by adding twice the standard deviation to the mean ($\text{Mean} + 2*\text{STD}$). The pixels showing values higher than the UB value are determined as being hotspots.

In Figure 3, the standard deviation and mean values are computed as $9.76 \mu\text{mol}/\text{m}^2$ and $6.51 \mu\text{mol}/\text{m}^2$, respectively. The upper bound value is calculated as $26.03 \mu\text{mol}/\text{m}^2$. A mask using the UB values is then applied, resulting in the hotspots shown in Figure 3.



Figure 3: Map of tropospheric NO₂ column density resulting from the application of the 2*sigma method to TROPOMI data.

- We have also used another straightforward approach, i.e. the Interquartile Range (IQR) method. The IQR defines the middle 50% of values when ordered from lowest to highest. It is calculated as the difference between the third quartile (Q3, median of the upper half of the data) and the first quartile (Q1, median of the lower of the data). Typically, hotspots are identified where the observed values exceed a certain multiple of the IQR value. We have opted for a factor of 1, as factors of 2 or higher tend to remove some hotspots from the analysis. The pixels where the tropospheric column is above the difference between Q3 and Q1 are considered as hotspots.

The result of this method is shown in Figure 4. At global scale, we obtain a Q1 value of $208 \mu\text{mol}/\text{m}^2$ and a Q3 value of $254 \mu\text{mol}/\text{m}^2$, resulting in an IQR of $46 \mu\text{mol}/\text{m}^2$. It should be noted that the filtering out of pixel values below $46 \mu\text{mol}/\text{m}^2$ is more stringent compared to the 2*sigma method, where pixels below $26 \mu\text{mol}/\text{m}^2$ are filtered (Figure 4).

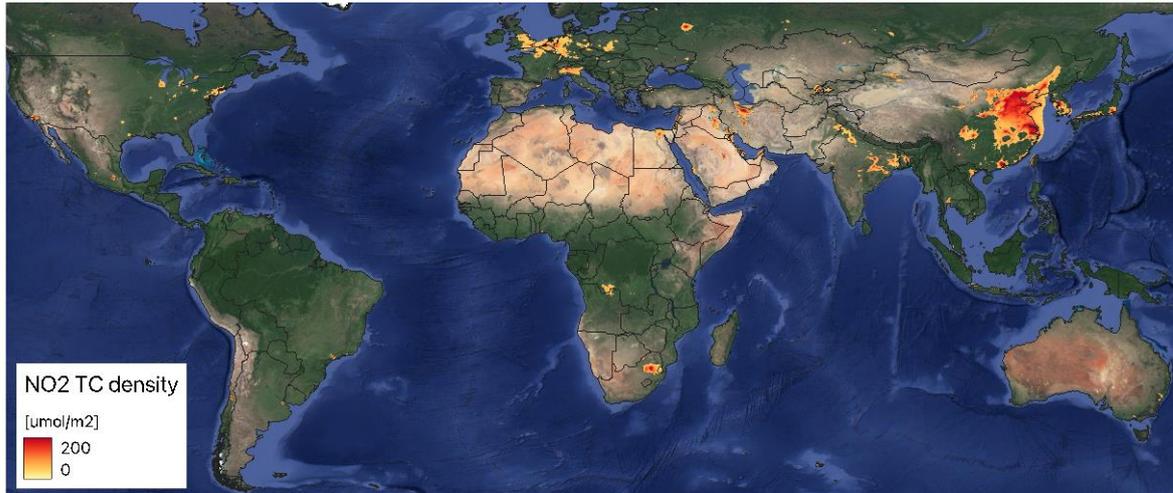


Figure 4: Map of tropospheric NO₂ column density resulting from the application of the IQR method to TROPOMI data.

A comparison between the results provided by both methods and the power plant locations sources from the CoCO₂ project (Guevara et al., 2024) and the World Research Institute (WRI, Yin et al., 2020) reveals that the 2*sigma method (red color palette) offers a more extensive spatial coverage of the identified hotspots over the selected South Africa region. However, it tends to overestimate the spatial distribution by generating more hotspots. Conversely, the IQR method (green color palette) tends to underestimate the distribution by potentially eliminating hotspots (Figure 5). These disparities in results are influenced by regional variations and the intensity of NO₂ column density.

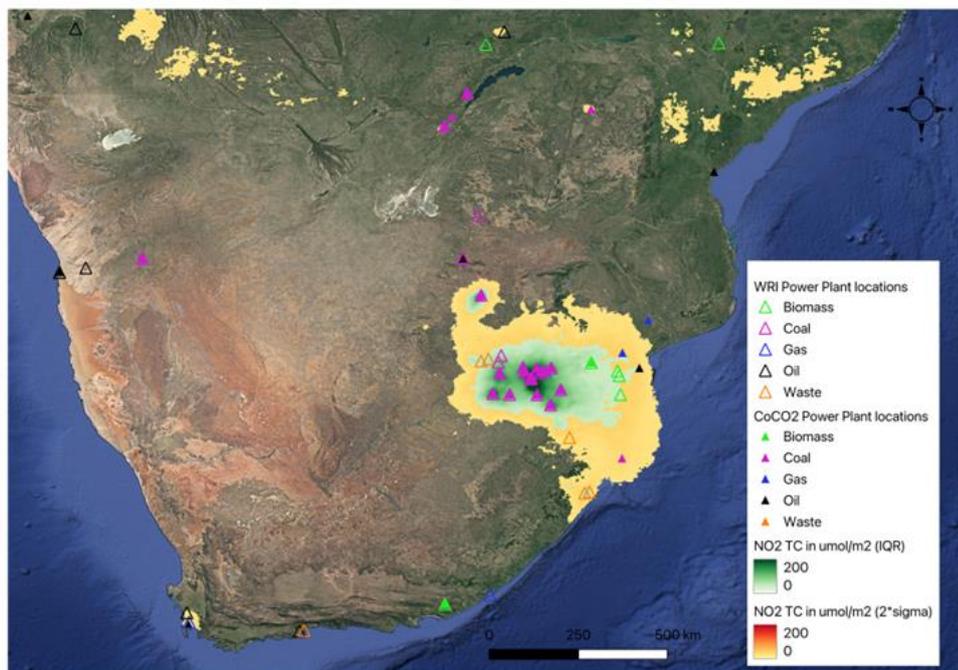


Figure 5: Map of tropospheric NO₂ column density following the application of the 2*sigma method (green color palette) and IQR method (red color palette) on the TROPOMI data in South Africa region. It also illustrates a comparison with power plant locations from CoCO₂ (triangles) and WRI (open triangles) databases. The colored markers represent the fuel type used in the power plants.

- Statistical method for detecting clusters: i.e. the Exploratory Spatial Data Analysis (ESDA) for hotspot analysis. This method utilizes LISA statistics (Local Indicators of Spatial Association) to evaluate the existence of clusters in the spatial distribution of a given variable. Unlike the previous methods, this approach offers the capability to determine statistically significant levels of high NO₂ tropospheric column density in a given area. We will evaluate in the coming months the possibility of quantifying the significance of regionally defined thresholds using the to the first two approaches.

In this report, the hotspot analysis aimed at identifying significant local clusters within a spatial dataset is conducted using the Getis-Ord G_i^* statistic (Getis and Ord, 1992; Caliskan and Anbaroglu, 2023). The Getis-Ord* method computes hotspots by taking the defined neighborhood of each spatial variable and calculating whether the values within that neighborhood are significantly higher or lower than across the entire area. The result of this analysis is a G_i^* value. Positive values indicate that the values within the neighborhood are notably higher when compared to the average regional value, indicating a hotspot. Negative G_i^* values suggest the opposite.

The output of this analysis includes a p-value variable, which represents the significance level allowing us to either reject or accept the null hypothesis of our analysis. If our null hypothesis stated that there is no spatial clustering of NO₂ tropospheric column levels in the studied area, then when cells fall below a certain threshold, we can reject this hypothesis. Typically, this threshold will be set at 0.01, 0.05, or 0.1, corresponding to confidence levels of 99%, 95%, or 90%, respectively, indicating the degree of confidence in rejecting the null hypothesis.

This approach, which involves employing the QGIS Hotspot Analysis tool on the dataset is illustrated in Figure 6. Initially, to manage the large number of pixels, the global TROPOMI NO₂ tropospheric column density is regridded from a resolution of 0.022° x 0.022° to 0.125° x 0.125° and processed separately for the five continents (Africa, Europe, Asia, America, and Oceania). It should be noted that the regridding process does not eliminate hotspots, but it consolidates the determination of hotspots: this conclusion has been derived from analyses over small areas.

Each resultant raster file is converted into a points shapefile containing the latitude, longitude, and associated NO₂ tropospheric column density value for each cell. To compute the local Getis-Ord G_i^* statistic, a fixed distance band of 10 km was chosen, representing approximately the minimum distance between the center of two adjacent pixels.

Following the specification of the input layer and distance band, the spatial weights matrix was determined using row-standardized spatial weights, where each element of the matrix rows is divided by the sum of the corresponding row. The output file comprises the extracted coordinates for hotspot pixels, along with their corresponding NO₂ tropospheric column density, G_i^* value (or Z-score), and related p-values for any entity in the dataset.

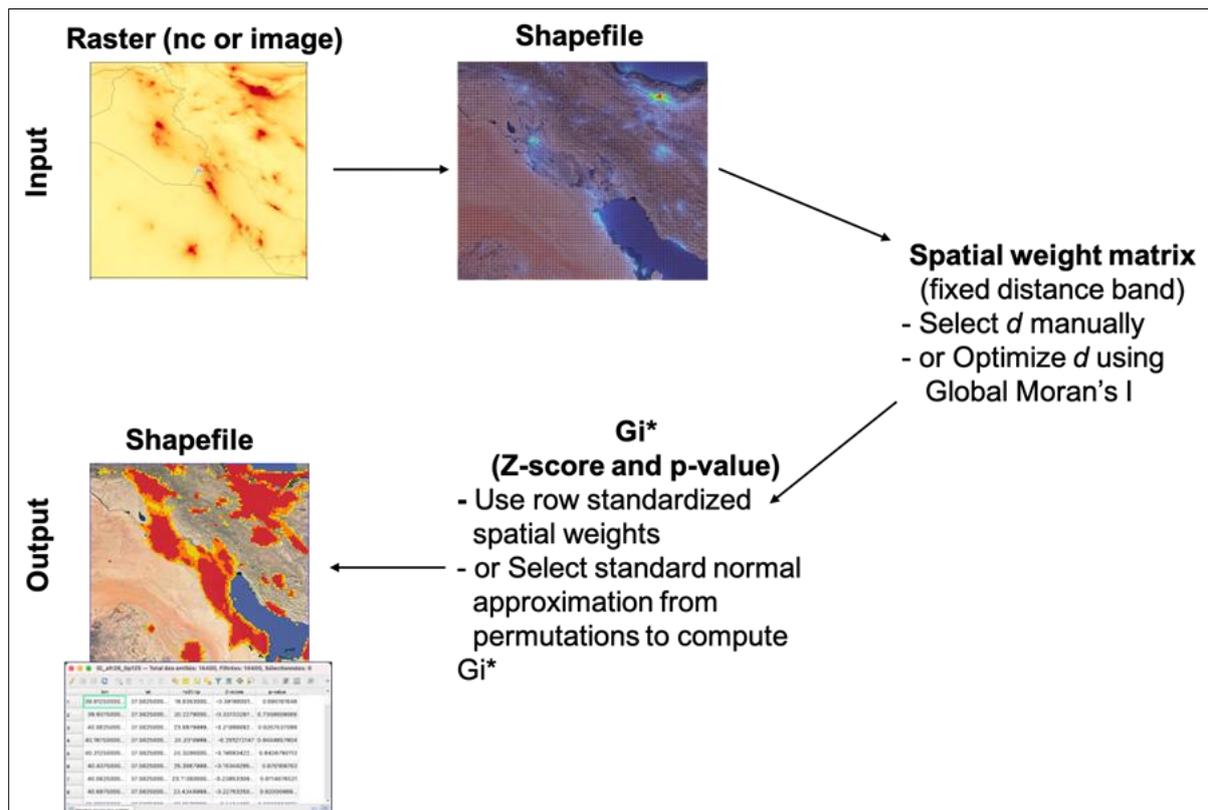


Figure 6: TROPOMI data processed with the IQR method to create a map of tropospheric NO₂ column density.

The identified hotspots using calculated G_i^* values are displayed in Figure 7. This figure is for the same region as for Figure 5, but it also includes hotspot values corresponding to the 90% ($1.65 \leq G_i^* < 1.96$ and $0.05 < p\text{-value} \leq 0.1$), 95% ($1.96 \leq G_i^* < 2.58$ and $0.01 < p\text{-value} \leq 0.05$), and 99% ($G_i^* \geq 2.58$ and $p\text{-value} \leq 0.01$) confidence levels. As depicted in Figure 6, the G_i^* method yields results that fall between those obtained from the 2*sigma and IQR (Interquartile Range) methods.

Notably, there is a good concordance between the hotspots identified through the G_i^* approach and the locations of power plants. Hence, the results obtained using the G_i^* are considered more appropriate for hotspot identification (Figure 7).

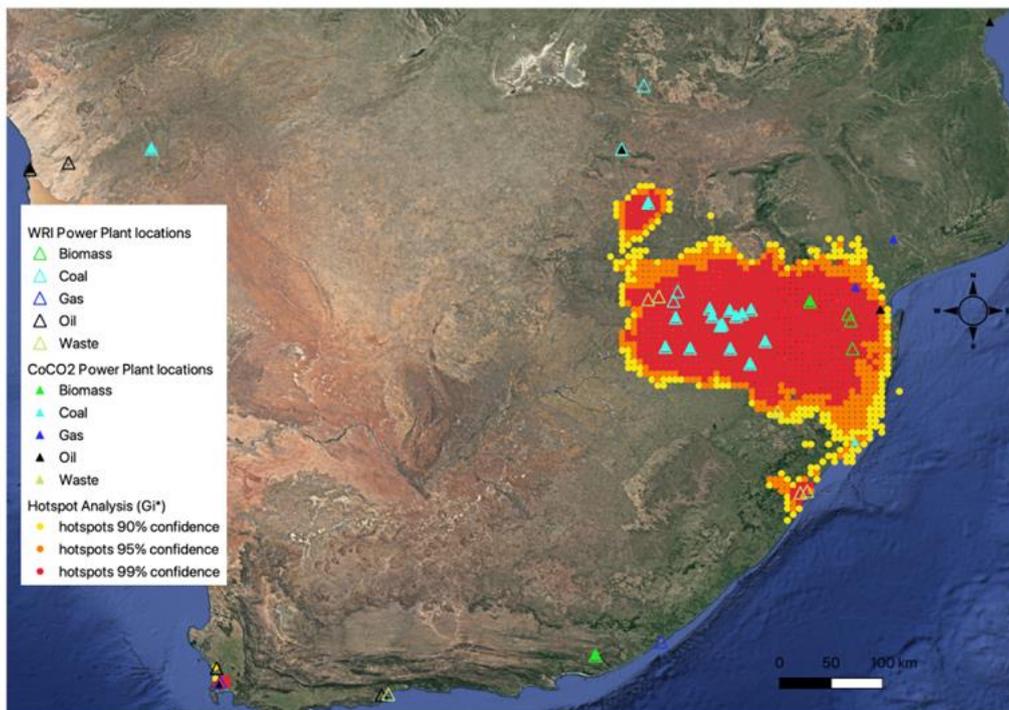


Figure 7: Hotspots of NO₂ tropospheric column density derived from the hotspots analysis tool (Gi*).

The hotspots detected through this method have been exported to a CSV file, provided with this report. Figure 8 provides an example of the output file format. To manage the extensive number of points beyond the confidence levels, only hotspots with significant confidence levels (namely, those at 90%, 95%, and 99%) are included in the final CSV file. Including the Gi* and p-values along with the other parameters will allow the further users to select threshold confidence levels in accordance with their specific objectives. The regional datasets have been merged into a single file to include hotspots at a global level.

| | A | B | C | D | E | F | G | H |
|----|-----------|-----------|----------|----------------------------|--------|---------|---|---|
| 1 | Continent | longitude | latitude | NO2 trop. Column (umol/m2) | Gi* | p-value | | |
| 2 | Africa | 14.938 | 37.563 | 33.090 | 1.774 | 0.07610 | | |
| 3 | Africa | 15.063 | 37.563 | 38.765 | 2.297 | 0.02161 | | |
| 4 | Africa | 15.188 | 37.563 | 31.922 | 1.666 | 0.09568 | | |
| 5 | Africa | 27.313 | 37.563 | 34.430 | 1.897 | 0.05778 | | |
| 6 | Africa | 27.438 | 37.563 | 33.260 | 1.790 | 0.07353 | | |
| 7 | Africa | 27.938 | 37.563 | 34.430 | 1.897 | 0.05778 | | |
| 8 | Africa | 28.063 | 37.563 | 39.958 | 2.407 | 0.01608 | | |
| 9 | Africa | 36.813 | 37.563 | 39.562 | 2.371 | 0.01776 | | |
| 10 | Africa | 36.938 | 37.563 | 65.542 | 4.766 | 0.00000 | | |
| 11 | Africa | 37.063 | 37.563 | 46.554 | 3.015 | 0.00257 | | |
| 12 | Africa | 44.813 | 37.563 | 52.564 | 3.569 | 0.00036 | | |
| 13 | Africa | 44.938 | 37.563 | 88.879 | 6.918 | 0.00000 | | |
| 14 | Africa | 45.063 | 37.563 | 122.452 | 10.014 | 0.00000 | | |
| 15 | Africa | 45.188 | 37.563 | 60.429 | 4.295 | 0.00002 | | |
| 16 | Africa | 45.313 | 37.563 | 50.521 | 3.381 | 0.00072 | | |
| 17 | Africa | 45.438 | 37.563 | 41.417 | 2.542 | 0.01103 | | |
| 18 | Africa | 45.563 | 37.563 | 42.225 | 2.616 | 0.00889 | | |
| 19 | Africa | 45.688 | 37.563 | 38.731 | 2.204 | 0.02170 | | |

Figure 8: Sample of output layer with the latitude, longitude, NO₂ tropospheric column density, Z-score and p-value field in the attribute table.

4 Future work

This deliverable focuses on the compilation of a list of hotspots detected using the TROPOMI NO₂ tropospheric column observations at the global scale. This work will be complemented during the coming months by the following activities:

- The generation of a list of hotspots for the Southeast Asia region using the GEMS geostationary observations of tropospheric NO₂ columns. The GEMS dataset has already been acquired, and the Gi* method will be employed for the analysis.
- Evaluation of the use of satellite observations of CO for hotspot detection, knowing that the longer lifetime of CO does not make hotspots easy to detect

5 References

Caliskan, M., and B. Anbaroglu, pace Time Cube analytics in QGIS and Python for hot spot detection, SoftwareX, <https://doi.org/10.1016/j.softx.2023.101498>, 2023.

Getis, A. and Ord, J.K., The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, 24, 189-206. <https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>, 1992.

Guevara, M., Enciso, S., Tena, C., Jorba, O., Dellaert, S., Denier van der Gon, H., and Pérez García-Pando, C.: A global catalogue of CO₂ emissions and co-emitted species from power plants, including high-resolution vertical and temporal profiles, *Earth Syst. Sci. Data*, 16, 337–373, <https://doi.org/10.5194/essd-16-337-2024>, 2024.

Yin, L., L. Byers, L. Malaguzzi Valeri, and J. Friedrich, Estimating Power Plant Generation in the Global Power Plant Database.” Technical Note. Washington, DC: World Resources Institute. Available online at: wri.org/upload/estimating-power-plant-generation-in-the-global-powerplant-database, 2020.

Document History

| Version | Author(s) | Date | Changes |
|---------|---|---------------|---|
| 1.1 | T. Doumbia, C. Granier, G. Kuhlmann, M. Guevara | February 2024 | |
| 1.2 | T. Doumbia | March 2024 | Minor updates incorporating reviewer comments |
| | | | |
| | | | |
| | | | |

Internal Review History

| Internal Reviewers | Date | Comments |
|--------------------|---------------|---|
| Richard Engelen | 11 March 2024 | Various comments for clarification of the text. |
| Thomas Kaminski | 14 March 2024 | In lined in document |
| | | |
| | | |
| | | |