# A GIS-Based Methodology for Multi-Pollutant Analysis of Sentinel-5P TROPOMI Data with Spatial Intersection Masking

Yana Lipiyska, Adrian Yordanov, Tsvetelina Atanasova-Evdenova, Nikolay Najdenov

Department of Photogrammetry and Cartography, Faculty of Geodesy, University of Architecture, Civil Engineering and Geodesy, Sofia, Bulgaria
(yangelova_fgs@uacg.bg, a.yordanov_fgs@uacg.bg, tsatanasova_fgs@uacg.bg, naid_fgs@uacg.bg)

**Keywords:** air quality, remote sensing, nitrogen dioxide, carbon monoxide, sulfur dioxide

**Abstract**

Urban air quality monitoring using satellite remote sensing requires consistent, spatially comparable datasets across multiple pollutants. This paper presents an end-to-end, open-source GIS-based methodology for processing Sentinel-5P TROPOMI Level-2 products to generate synchronized multi-pollutant datasets suitable for urban-scale analysis. The workflow integrates automated data acquisition via the Copernicus Data Space Ecosystem OData API, quality-controlled preprocessing using the HARP toolbox and a QGIS-based spatial analysis pipeline developed in Python. The main methodological contribution is a spatial intersections masking approach, in which a daily mask of valid pixels is calculated as a logical intersection of valid observations for carbon monoxide (CO), nitrogen dioxide ($NO_2$), and sulfur dioxide ($SO_2$). This ensures that all statistical comparisons between pollutants are derived from spatially coinciding sets of pixels, thereby eliminating location-based sampling biases arising from differences in swath geometry, cloud cover, and gas-specific retrieval sensitivity. The methodology was applied to four consecutive winter seasons (2019–2023) over Sofia, Bulgaria, a city with significant air pollution driven by residential solid-fuel combustion. The resulting synchronized daily database enables temporal trend analysis (Mann-Kendall test), inter-pollutant correlation assessment, and pollution hotspot identification. The workflow relies entirely on open-source software (QGIS, GDAL/OGR, NumPy, HARP) and is designed to be easily adaptable to any urban area and any combination of TROPOMI gas products with minimal parameter adjustments.

## 1. Introduction

Improving air quality ranks among the highest priorities globally, being identified as one of the leading environmental causes of death. It has a direct impact on human health and necessitates serious action to reduce pollutants to acceptable levels. Since the early 1970s, the European Union has been working to improve air quality, yet it remains one of the most serious environmental and public health challenges in urbanized areas across Europe. Globally, elevated concentrations of atmospheric pollutants such as fine particulate matter, nitrogen dioxide ($NO_2$), and sulphur dioxide ($SO_2$) are associated with increased morbidity and premature mortality, which is even more pronounced in densely populated cities.

In Bulgaria, air pollution represents a serious problem: in 2019 alone, approximately 11,000 deaths were attributed to it, making it the seventh most significant mortality risk factor in the country. One of the main contributing factors to air pollution is residential heating. According to the State of Global Air report (State of Global Air, 2022), emissions from residential heating — particularly the combustion of solid fuels such as wood and coal — are among the primary sources of fine particulate matter ($PM_{2.5}$), alongside road transport emissions. Previous studies conducted in Bulgaria indicate a clearly defined seasonal trend in air pollution, with higher concentrations of particulate matter and nitrogen dioxide observed during winter months as a result of increased emissions and reduced atmospheric mixing (Dimitrova et al., 2019), when elevated heating demand coincides with unfavourable meteorological conditions.

In recent years, national and municipal policies have focused on introducing measures to reduce residential heating emissions through programs that replace inefficient heating systems with cleaner energy sources. Evaluating the effectiveness of the measures undertaken requires consistent, reliable, and spatially integrated — within a unified geographic information system (GIS) — observations of atmospheric pollutants over a period of several years.

Satellite remote sensing can serve as a significant complement to ground-based air quality monitoring. Unlike in-situ measurements, which are limited to specific locations, satellite observations offer the advantages of continuous spatial coverage and the ability to analyse long-term trends and variability in pollutant distribution.
Nevertheless, the use of Sentinel-5P TROPOMI L2 data in GIS-based studies of urban areas faces certain technical obstacles, such as uneven spatial coverage caused by cloud interference, differences in orbital swath geometry, differences in sub-dataset structure among various gases, and lack of a standardized spatial synchronization process for multiple pollutants.
In this paper, we propose an end-to-end, open-source methodology for processing Sentinel-5P TROPOMI data and creating synchronized spatial datasets of various gas products for GIS-based urban pollution analysis.
We have developed a workflow based on a study of the city of Sofia, Bulgaria, during four consecutive winters In particular, carbon monoxide (CO), nitrogen dioxide ($NO_2$), and sulfur dioxide ($SO_2$) were selected as target pollutants, and a spatially synchronized, quality-assured dataset of daily average concentration values for all three species was successfully obtained and prepared for subsequent statistical analysis. Our pipeline can easily be transferred to any other urban area of interest, as well as for any other combination of Sentinel-5P gas products, provided that the spatial and temporal extents are specified.

## 2. Related work

Several studies have used Sentinel-5P TROPOMI data for urban and regional air quality monitoring. Comparisons between TROPOMI column retrievals and ground-based $NO_2$ measurements have shown strong correlation, supporting the use of satellite data for urban-scale analysis (Oxoli et al., 2020; Tonion and Pirotti, 2022). More recent work has explored machine learning approaches for estimating surface-level $NO_2$ concentrations from TROPOMI observations (Karagkiozidis et al., 2024; Cedeno Jimenez et al., 2023). For Bulgaria specifically, Georgieva et al. (2021) applied satellite-derived data to air pollution modelling, and Dimitrova et al. (2019) documented seasonal pollution patterns linked to residential heating.

These studies, however, typically treat individual pollutants independently, without addressing the spatial coverage differences that arise when comparing multiple gas species observed under different retrieval sensitivities and cloud-masking criteria. To the best of the authors' knowledge, no existing workflow explicitly enforces spatial coincidence between pixel sets of different pollutants prior to statistical comparison — a gap this study addresses through the spatial intersection masking approach described in Section 3.4.

## 3. Materials and methods

### 3.1 Sentinel-5P Mission and TROPOMI Instrument

The Sentinel-5 Precursor (Sentinel-5P) satellite is part of the European Copernicus programme and is specifically designed for monitoring atmospheric composition at global and regional scales. Launched in 2017, it provides daily global coverage with a spatial resolution of up to 5.5 × 3.5 km² from August 2019 onwards (previously 7 × 3.5 km²), enabling detailed monitoring of air pollution at the urban scale (ESA/Copernicus, 2023). This high spatial resolution represents a significant improvement over previous satellite missions and allows for the detection of pollution patterns over individual cities.

Sentinel-5P is equipped with the TROPOspheric Monitoring Instrument (TROPOMI). TROPOMI measures solar radiation reflected by the Earth's atmosphere and surface based on hyperspectral observations in the ultraviolet, visible, near-infrared, and shortwave infrared spectral ranges. The retrieved products include vertical column densities (VCDs) of trace gases, which represent the integrated concentration of a pollutant throughout the atmospheric column. The foundational description of the mission and instrument is provided by Veefkind et al. (2012), which establishes TROPOMI as a key advancement in atmospheric remote sensing. Among the gases and aerosols measured by the spectrometer are nitrogen dioxide ($NO_2$), sulphur dioxide ($SO_2$), carbon monoxide (CO), ozone ($O_3$), methane ($CH_4$), and formaldehyde (HCHO) (ESA/Copernicus, 2023).

The accuracy of TROPOMI products depends on several factors such as cloud cover, snow, and surface reflectance, which can reduce data availability or increase uncertainty (ESA/Copernicus, 2023). Typical accuracy for $NO_2$ retrievals is within 10–30%, depending on atmospheric conditions, with higher reliability in polluted regions and increased uncertainty in cleaner environments (van Geffen et al., 2022).

Validation studies comparing satellite observations with ground-based networks (e.g., NDACC and Pandora) confirm that TROPOMI data are suitable for regional and urban-scale analyses (Verhoelst et al., 2021). Furthermore, comparisons with chemical transport models, such as the Copernicus Atmosphere Monitoring Service (CAMS), indicate that satellite-derived $NO_2$ data can effectively capture spatial patterns and temporal variability over European cities (Douros et al., 2023).

### 3.2 Data and study site

This analysis covers the winter period (December – February), 2019-2023, over Sofia, Bulgaria.

A total of 472 CO products, 448 $NO_2$ products, and an equivalent set of $SO_2$ products were retrieved using the Copernicus Data Space Ecosystem (CDSE) OData API. Data acquisition was automated through a custom Python script designed to filter and download datasets according to the study requirements.

An overview of the script is presented in Figure 1.
The script consists of three main functions, developed in accordance with the OData API guidelines (Copernicus Data Space Ecosystem, n.d.): get_request, get_authorization, and download_products.

**3.2.1 Get_request:** The get_request function of the script builds API queries by using specific parameters such as mission name, product type, processing mode, temporal range and area of interest (AOI). All queries are used to generate URL and search products in the CDSE database. The script iterates over multiple time intervals to ensure complete temporal coverage with one run instead of multiple. Priority is given to reprocessed products when building the queries since their quality is improved by ESA with additional processing. If no reprocessed products are available, the script switches to the next best available processing mode until no more are available. Following a successful request all product ids and names are collected for subsequent download.

**3.2.2 Get_authorization:** The get_authorization function handles authentication with the CDSE by generating an access token using the user's credentials. The function also generates refresh tokens using the access token in order to keep downloading products for long periods of time.

#### 3.2.3 Download_products

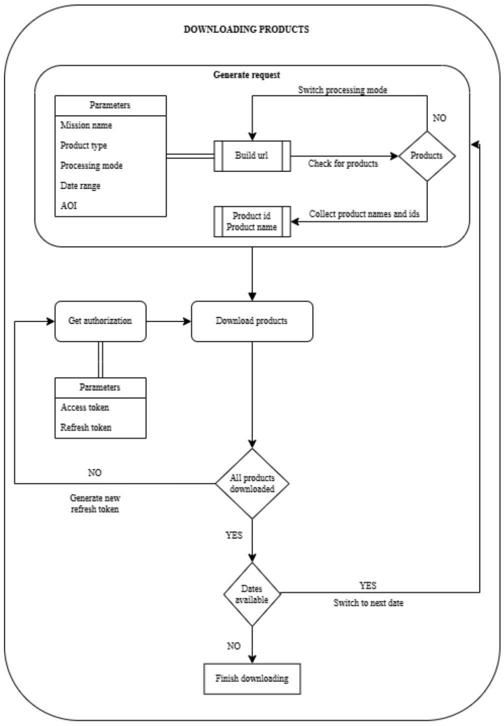


Figure 1. Downloading algorithm overview


https://doi.org/10.5194/isprs-annals-XII-4-W2-2026-121-2026  

The download_products function manages the retrieval of datasets by iterating through the list of product names and identifiers obtained from the get_request function. Each product is then downloaded from the CDSE system. The function also generates new refresh tokens when necessary. The overall script utilizes those three core functions and iterates over defined time intervals until all data requests are fulfilled. It also allows the users to define inputs in human readable format, which are then internally converted to appropriate structure for the get_request function.

### 3.3 Processing

All datasets were processed using the HARP toolbox in Python (S[&]T, n.d.). An overview of the processing chain is presented in Figure 2. The main function of the processing were to georeference the data and resample it onto a regular grid. The script was implemented following established guidelines (CloudFerro, n.d.). It includes data filtering, spatial binning and unit conversion.

Data filtering was based on the following criteria:

- **Quality flag** – different quality flag thresholds are used for each product type according to recommendations from ESA. For NO2 products, a threshold of > 0.75 was used (Eskes and Eichmann, 2023). For the CO and SO2, thresholds of > 0.5 were applied (S5P Mission Performance Centre, 2023; Pinardi and Hendrick, 2022).
- Wind speed – to reduce uncertainties due to high wind speed, all pixels with wind speed above 5 m/s were removed.

Resampling of the original products onto a regular gird is required since the Sentinel-5p products are swath observations. Those types of products are not suitable for GIS-based analysis. This operation is performed using the spatial binning operator in HARP (S[&]T, n.d.). The operator grids the original data onto a rectilinear latitude-longitude grid.

The operator reads the spatial extent of the input data along with user-defined grid parameters and the constructs the target grid as an array of coordinate pairs. The cell values of the created grid are calculated using area-weighted averaging. Each original pixel contributes to the new cell value based on its spatial overlap with the new grid. The exact value of the weight for each pixel is calculated using the following equation:

$$\omega(i,j,k) = \sum_i \begin{cases} \frac{area(P_t(j,k) \cap P_s(i))}{area(P_t(j,k))} & x_s(i) \neq nan \\ 0, & x_s(i) = nan \end{cases} \quad (1)$$

Where:
$w_{(i, j, k)}$ – weight value,
$P_{t\,(j, k)}$ – area of the new cell,
$P_{t\,(j, k)}$ ^ $P_{s\,(i)}$ – spatial overlap of the new cell and the original pixel,
j, k – indices of the current cell in the target grid,
i – index of the original pixel.
In case there are no pixels overlapping the new grid cell, the weight for that cell is 0.
The new cell values are determined using the following equation:

$$y_t(j,k) = \frac{\sum_i \omega(i,j,k)\, y_s(i)}{\sum_i \omega(i,j,k)} \quad (2)$$

Where:
$y_{t\,(j, k)}$ – cell value,
$w_{(i, j, k)}$ – weight value,
$y_{s\,(i)}$ – value of the original pixel,
j, k – indices of the current cell in the target grid,
i – index of the original pixel.

All datasets were projected onto a grid defined by the following boundaries:

- North – 43.00°
- South – 42.00°
- East – 24.00°
- West – 23.00°

The spatial bining parameters were derived from the grid boundaries using the following 2 equations:

$$A = \frac{N - S}{SR + 1} \quad (3)$$
$$B = \frac{E - W}{SR + 1}$$

Where,
A – number of latitude edge points,
B – number of longitude edge points,
N – coordinate of the north-most boundary,
S – coordinate of the south-most boundary,
E – coordinate of the east-most boundary,
W – coordinate of the west-most boundary,
SR – spatial resolution.
The resulting parameters are presented in table 1:

| Parameter | Value |
|---|---|
| Number of latitude edge points (A) | 101 |
| Latitude offset (S) | 42.00° |
| Number of longitude edge points (B) | 101 |
| Longitude offset (W) | 23.00° |
| Spatial resolution (SR) | 0.01° |

Table 1. Spatial binning parameters

All retrieved values are also converted to $10^{15}$ * molecules/cm$^2$.

The described workflow was used to derive 2 types of products with different temporal coverage – daily and monthly. Monthly averages were derived by aggregating spatially binned data and computing mean values. Those products can be used to determine general trends about pollutant concentrations in the study area. However, those products were not implemented in the later analysis.

Daily products were used for more precise analyses. These datasets exhibit higher variability and contain missing data due to stricter filtering criteria and sensitivity to atmospheric conditions such as wind and cloud coverage.

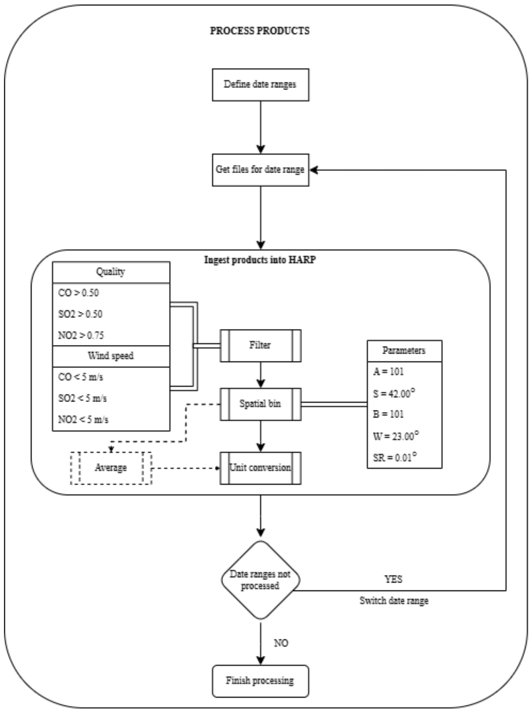


Figure 2. Processing algorithm overview

### 3.4 GIS-Based Spatial Analysis

After the initial pre-processing of Level-2 products, a spatial workflow was developed. This section describes the spatial workflow implemented in QGIS, culminating in the spatial intersection masking approach, detailed in Section 3.4.3 QGIS was used for comparing temporally and spatially the concentrations of all three target pollutants – carbon monoxide (CO), nitrogen dioxide ($NO_2$), and sulfur dioxide ($SO_2$).

The workflow was created in the form of a Python script using the QGIS Python Console and utilized the following Python libraries: GDAL/OGR (GDAL/OGR contributors, 2024), NumPy (Harris et al., 2020). Everything was done automatically to be able to process a considerable number of daily files (about 263–347 per gas type). Gemini AI was used for help with script writing.

**3.4.1 Data Loading and Temporal Aggregation via Virtual Raster Tiles:** Pre-processed NetCDF (.nc) files were imported to QGIS through a custom python script. This was done, because the files were not correctly positioned in QGIS, due to the Sentinel-5P .nc file not being a regular raster. Its coordinates are irregular arrays (swaths).

The fact that the Sentinel-5P satellite collects data in swaths means that in one particular day over the area of study, several files may overlap; thus, to combine those files into a single file on daily basis without having duplicate data stored on disk, GDAL Virtual Raster Tile (VRT) files were used (GDAL/OGR contributors, 2024).

The grouping of files temporally wise was done by regular expressions (Python re module) by reading the acquisition date of the files from the name of the file itself (e.g., S5P_OFFL_L2__CO_____20221202T...). In other words, the name of the file contained an eight-digit date that was extracted from it using the regular expression and then the files having the same acquisition date were grouped together in a single VRT file. In this regard, the VRT files' coordinate reference system was set to WGS 84 (EPSG:4326), same as HARP-processed products.

**3.4.2 Quality Assessment and Coverage Filtering:** The main difficulty is posed by the highly variable spatial coverage due to clouds and tight filtering criteria implemented during preprocessing (QA ≥ 0.75 for $NO_2$ and QA ≥ 0.5 for CO and $SO_2$, describes in Section 2.3). In days with extensive cloud coverage (typical in winter months), only a small portion of valid pixels are left, and including such a day in the dataset could significantly contribute to statistical uncertainty.

In order to evaluate the spatial representativeness of each daily image, the total number of valid pixels within the study area VRT was calculated. It was done using the NumPy isfinite() function on raster band array retrieved by GDAL. Daily images with less than 1,000 valid pixels were not included in further spatial analysis and marked as "Excluded" in statistical results but kept in the dataset, nonetheless. This threshold was chosen based on the empirical observation of daily valid pixels distribution that ranged between less than 10 and over 8,400 valid pixels for $SO_2$ and between about 50 and close to 10,000 valid pixels for CO.

For $SO_2$, using the 1,000 pixels cut-off, about 183 days with good data remained out of the possible 266 days. For CO, using the same pixel cut-off, a similar number of days was retained. The $NO_2$ gas had higher coverage compared to the other two on a daily basis. This is supported by research that notes that the $NO_2$ retrieval is less sensitive to sub-pixel clouds than $SO_2$ and CO (Eskes and Eichmann, 2023).

It should be noted, that the presence of negative values in the $SO_2$ daily products is in accordance with the existing properties of the TROPOMI $SO_2$ retrieval algorithm. In clean-air conditions, due to the noise from the instrument and some systematic errors of the spectral fitting, slightly negative values may appear. This phenomenon, is called the measurement noise floor, is described in the Sentinel-5P $SO_2$ Product Readme File (Pinardi and Hendrick, 2022) and is an expected artefact of the differential optical absorption spectroscopy (DOAS) retrieval approach. In compliance with the common recommendation (Theys et al., 2017), negative values were not removed during the averaging of temporal series.

**3.4.3 Spatial Intersection Masking for Multi-Pollutant Comparability:** A base principle, for this research study, of the multi-pollutant assessment is that the statistical comparison between CO, $NO_2$, and $SO_2$ needs to be done by means of observations from the same region on any particular day. Each of the three gases is different upon meteorological conditions, cloud cover, and aerosols, the size of their respective footprints vary depending on the particular dataset used, despite the satellite orbit being the same (Borsdorff et al., 2018; Theys et al., 2017; van Geffen et al., 2022). Comparing mean concentrations of these gases measured within different subsets of the region would have resulted in location-based distortions in data interpretation.

In order to correct this problem, spatial intersection masking was applied. Thus, for each day when all three gases had more than 1,000 valid pixels observed, the following intersection mask was generated logically:

$$CommonMask(date) = Valid(SO_2) \cap Valid(CO) \cap Valid(NO_2) \quad (4)$$

where Valid(gas) is a Boolean array which takes the value 1 for pixels with non-NaN values in the daily VRT of the gas and 0 otherwise. The intersection was performed by element-wise logical AND operation with the use of the Python library NumPy (Harris et al., 2020). As a result, a binary raster with identical geographic extent and resolution was created and saved on disk as a GeoTIFF with the Byte data type (0/1). NoData value in the output raster was set to 0.


https://doi.org/10.5194/isprs-annals-XII-4-W2-2026-121-2026  

Finally, the common masks were used to create daily concentration grids for all gas species, with the values of all pixels located outside of the mask being replaced by NaNs. In this way, statistical operations (daily mean concentrations, standard deviations, and inter-pollutant correlation coefficients) were computed only using coinciding spatial pixels in the three products. Thus, spatial coincidence of data was guaranteed on each analysis day. Daily concentration grids were saved as Float32 GeoTIFF files and imported into QGIS for further visualization and processing.

**3.4.4 Synchronised Daily Statistics:** For those days where a valid common mask existed, spatially equal average concentrations were obtained for all three gas types. Daily average was calculated using the arithmetic average of all pixels within the common mask area via NumPy array manipulation performed directly on the GDAL imported raster images. This information was stored in a comma-separated values (CSV) file consisting of the following columns for each valid observation day: date, mean $SO_2$ column density, mean CO total column, mean $NO_2$ tropospheric column, and the total number of synchronous pixels.

This synchronised daily statistics file acts as the main source for the analysis of temporal trends and inter-pollutant correlations presented in the next two sections. The selection of spatially identical sets of pixels for all three gases ensures that there is no spatial sampling error when comparing observations drawn from products that have different spatial extents, which is in accordance with the recommended approach for fusing satellite data from multiple species in urban air quality assessments (Borsdorff et al., 2018; Georgieva et al., 2021).

The complete GIS processing workflow, from importing of the data through to synchronized computation of the statistics, was performed using a set of Python scripts that could be executed inside the QGIS Python Console module. The scheme of the workflow is presented in Figure 3.

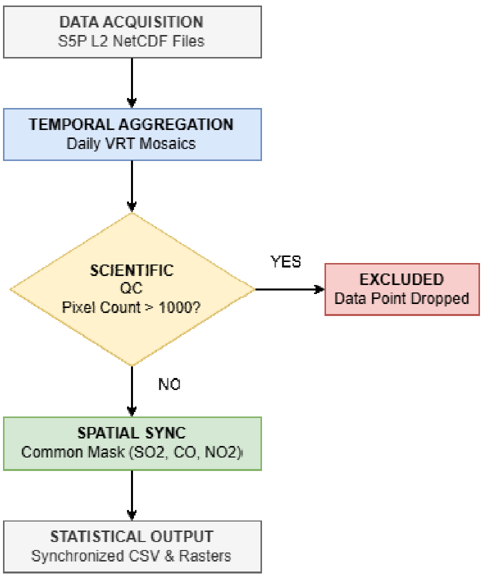


Figure 3. Automated data extraction

All intermediate and output raster maps were arranged into a QGIS project organized in layers for each day and gas type.

## 4. Results.

After creating the synchronized data set, various techniques of analysis, both in terms of space and time, can be used to determine the spatial and temporal variation in pollutant concentration. For this study we choose to use some of the most appropriate. A lot of other methods are applicable such as Inter-Pollutant Correlation and Source Apportionment, Zonal statistic etc.

### 4.1 Temporal Trend Analysis

The analysis of temporal trends was performed with the Mann–Kendall test (Mann, 1945; Kendall, 1975), which is a non-parametric method for detecting trends in time series data. This technique has been extensively applied in the atmospheric sciences and environmental studies as a means of testing for monotonic trends in time series, which does not require normally distributed data and is insensitive to missing data, a feature that can be especially important when dealing with satellite-derived time series, which may contain gaps owing to cloud contamination (Alexandersson, 1986). If a significant trend was detected, its magnitude was calculated through the Theil–Sen slope estimator (Sen, 1968). These statistical methods were implemented using the *kendall and trend* packages in R, within the QGIS processing system through the *Processing R Provider plugin.*

The Mann-Kendall trend test and Theil-Sen slope were performed for the synchronised daily mean concentration time series of each gaseous species, using n = 116 days of valid observations spread over four winters (2019/2020-2022/2023). The findings are summarised in Table 2.

| **Gas** | **z-statistic** | **Kendall's τ** | **p-value** | **Sen's slope (units day⁻¹)** | **95% CI** | **Trend** |
|---|---|---|---|---|---|---|
| $SO_2$ | –4.05 | –0.255 | < 0.001 | –0.564 | [–0.818; –0.322] | Significant decrease |
| CO | 1.39 | 0.088 | 0.164 | 0.572 | [–0.293; +1.382] | Not significant |
| $NO_2$ | –0.80 | –0.050 | 0.424 | –0.002 | [–0.008; +0.004] | Not significant |

Table 2. Mann-Kendall test for all 3 gases

Statistical significance was assessed at the $\alpha = 0.05$ level; p-values below this threshold indicate that the detected trend is unlikely to be attributable to random variation in the data.

The grey zone bordering the trend line is referred to as the 95% confidence interval of the linear fit. The width of this zone signifies the degree of confidence in the estimate of the slope. The narrow the zone, the higher the level of confidence; the wider the zone, the lower the level of confidence.

A significantly monotonically decreasing trend was found for $SO_2$ ($z = -4.05$, $\tau = -0.255$, $p < 0.001$). The value of -0.564 units day$^{-1}$ for the Theil-Sen slope translates to an average decrease of about 50.8 units over the course of the winter season. This is indicative of a steady decline in the levels of sulphur dioxide column density. However, further validation is necessary to determine specific causative factors.

Neither a statistically significant increasing (CO) nor a decreasing ($NO_2$) trend was observed in the latter two gases (CO, $z = +1.39$, $\tau = +0.088$, $p = 0.164$; $NO_2$, $z = -0.80$, $\tau = -0.050$, $p = 0.424$). In particular, Sen's slope for CO (+0.572 units day$^{-1}$) falls within a 95% confidence interval that includes zero (-0.293 to +1.382).

As for $NO_2$, the estimate is close to zero (–0.0024 units day$^{-1}$) meaning that no changes in column densities of nitrogen dioxide have been registered, i.e., that the tropospheric pollution by $NO_2$ stayed at the same level throughout the study period.

The difference between the statistically significant decreasing trend in $SO_2$ and the lack of trend in CO and $NO_2$ is quite intriguing. It can be interpreted as a hint that any decrease in the studied pollutants happened due to lessened combustion of fuels rich in sulphur compounds but without noticeable impact on road transport activities ($NO_2$) and biomass burning activities (CO). Such conclusions align well with those drawn by Georgieva et al. (2021) for Bulgarian cities. On Figures 4 to 6 results are shown for all the gases.

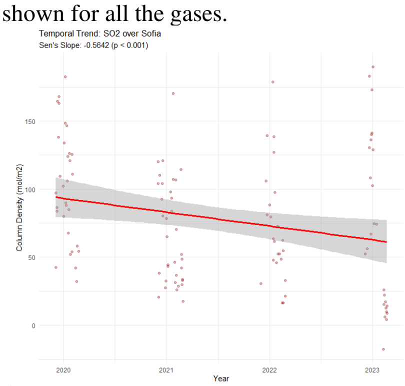


Figure 4. Temporal trend of $SO_2$ concentrations over Sofia, showing a slight decreasing slope with variability across the study period and confidence interval indicated.

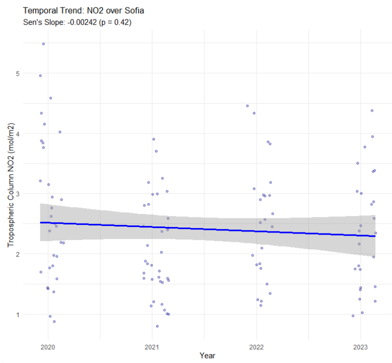


Figure 5. Temporal trend of $NO_2$ concentrations over Sofia, showing a slight decreasing slope with variability across the study period and confidence interval indicated.

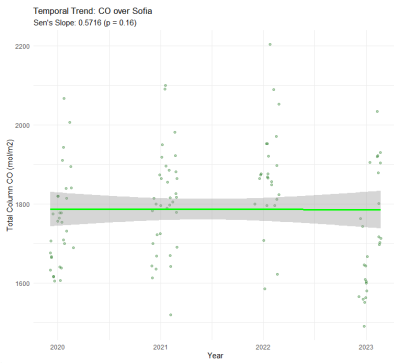


Figure 6. Temporal trend of CO concentrations over Sofia, showing almost constant slope with variability across the study period and confidence interval indicated.

### 4.2 Spatial hotspot analysis

The hotspot analysis technique was implemented through the Getis-Ord $G_i^*$ index (Getis and Ord, 1992). This method is used to detect statistically significant concentrations of high and low values (hotspots and coldspots) in space. The technique enables us to identify pollution hotspots on a smaller than urban level, and thus, the results obtained will be more reliable compared to mere concentration maps. The Hotspot Analysis plugin in QGIS was utilized for this analysis. Figure 7 shows an example of hotspot analysis. Where white spots represent no significant correlation and red so called „hotspot“ with correlation above 1.76.

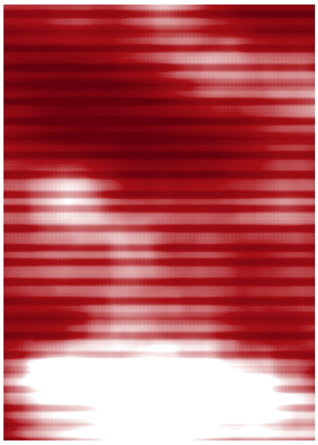

Figure 7. Hot spot analysis of CO data

## 5. Discussion

The main innovation (primary contribution) in this study comes from the GIS-based analysis workflow and, in particular, the spatial intersection masking technique. Although the techniques used in the workflow mosaic formation based on virtual rasters, quality filtering, and zonal statistics — are standard procedures in satellite data processing (GDAL/OGR contributors, 2024; Borsdorff et al., 2018), the implementation of these procedures as an automatic workflow for multiple pollutants with spatial coincidence enforced before comparison is a methodological advancement over those presented elsewhere in the literature.

Practical importance of the operation becomes clear if we look at the result of neglecting the procedure in question. Under conditions of partly cloudy skies, for instance, $SO_2$ valid pixels may end up located in the industrialized southern part of the region under analysis, whereas $NO_2$ valid pixels will likely be gathered around the center of the city. In case no masking had been performed, the difference in daily means of the two species would be explained by a possible source effect. However, using an identical spatial sample by construction, which removes location-based sampling bias as a source of discrepancy between pollutants.

The workflow is designed to be transferable to an analysis in any other (urban) region, requiring only minor adjustments.

The elements of the workflow that need to be set again when implementing it in another region include: the bounding box of the study area, the period under investigation, the gas of interest, quality filter threshold values (based on the guidelines specific to ESA products), and output grid spatial resolution. They are specified as variables at the very start of each script and do not contain any hardcoded geographic information. Automated detection of the subdataset names additionally guarantees that the workflow will work with any new version of the product.

The workflow used exclusively open-source software, namely QGIS, GDAL/OGR, NumPy, and the HARP library. All of them are free software and are available on Windows, macOS, and Linux platforms. Moreover, they are packaged as individual modules whose input/output specifications are documented clearly.

There are several limitations associated with the workflow that need to be acknowledged. Firstly, the application of spatial coverage filtering described, and the masking of intersections discussed causes a loss of information, especially when the weather conditions are not stable. In the final results 116 days of 266 were used.

This problem is typical for any kind of passive optical remote sensing and cannot be prevented by the proposed workflow itself.

Secondly, the concentration values that were examined using the TROPOMI products correspond to integrated concentrations in the columns above the surface, not surface concentrations. The conversion factor between the column density and the surface concentration is determined by the vertical distribution profile for each gas, which varies according to meteorological conditions and the height of the sources. As such, this workflow did not have a step of applying the vertical profile factor, and the interpretation of the output values should be done based on the total load of the atmosphere.

Thirdly, with a 0.01° grid size (≈11km), this resolution is not sufficient to capture urban pollution variations. The resolution is set according to the highest resolution that the instrument can acquire, meaning that the output will not get better through the processing steps, without the use of downscaling techniques.

## 6. Conclusion

The current work describes a workflow for automatable, modular and open source multi-pollutant GIS-based analysis of the Sentinal-5P TROPOMI Level-2 products that solves three main technical issues: (1) the aggregation of multi-swath daily observations into meaningful daily composites, (2) the screening out of days where the spatial coverages are not sufficient for further analysis and (3) the spatial synchronisation of observations of different gases for reliable multi-pollutant statistical comparison.

Spatial mask intersections, the core methodological approach in this study, ensure by design, that the generated daily statistics are based only on spatially consistent pixel sets without any location-based sampling biases affecting the inter-pollutant correlations and trends in pollution concentrations. The proposed methodology was derived based on data for four consecutive winter seasons (2019-2023) in a case study region of Sofia, Bulgaria. The resulting synchronised daily database of CO, NO2 and SO2 concentration maps allows for the generation of time trends (Mann-Kendall test), inter-pollutant correlation analysis, pollution hotspots and source attribution via CO/NO2 ratio etc.

The process is easily transferrable to other (urban) locations and different sets of TROPOMI gas concentrations. The approach utilizes entirely open-source software packages (QGIS, GDAL/OGR, NumPy, HARP), and it has been designed such that minimal changes of several location-specific parameters would suffice to implement the process in a new setting.

In future research, it is planned to use the process to analyze further Central and Eastern European regions where residential wood/coal burning is one of the major sources of air pollution during wintertime, as well as to integrate ground-level measurements to confirm trends obtained through the satellite approach. In addition, including meteorological covariates (e.g., temperature, wind speed, planetary boundary layer height) based on products like ERA5 is considered as an option for differentiating emission-driven changes from those caused by meteorological variations.

## References

Alexandersson, H., 1986. A homogeneity test applied to precipitation data. *Journal of Climatology*, 6(6), 661–675. https://doi.org/10.1002/joc.3370060607.

Borsdorff, T., aan de Brugh, J., Hu, H., Hasekamp, O., Sussmann, R., Rettinger, M., Hase, F., Schneider, M., Garcia, O.E., Kouremeti, N., Remmers, J., Landgraf, J., 2018. Mapping carbon monoxide pollution from cities and wildfires using the Sentinel-5P TROPOMI instrument. *Geophysical Research Letters*, 45(7), 3507–3515. https://doi.org/10.1002/2018GL077987.

Cedeno Jimenez, J.R., Pugliese Viloria, A.d.J., Brovelli, M.A., 2023. Estimating daily $NO_2$ ground level concentrations using Sentinel-5P and ground sensor meteorological measurements. *ISPRS International Journal of Geo-Information*, 12(3), 107. https://doi.org/10.3390/ijgi12030107.

CloudFerro, n.d. Processing Sentinel-5P data on air pollution using Jupyter Notebook on Creodias. *CloudFerro / CREODIAS Documentation*. https://creodias.docs.cloudferro.com/en/latest/cuttingedge/Processing-Sentinel-5P-data-on-air-pollution-using-Jupyter-Notebook-on-Creodias.html.

Copernicus Data Space Ecosystem, n.d. OData API. *Copernicus Data Space Ecosystem Documentation*, European Space Agency (ESA). https://documentation.dataspace.copernicus.eu/APIs/OData.html.

Dimitrova, M., Trenchev, P., Georgieva, E., Neykova, N., Neykova, R., Nedkov, R., Gochev, D., Syrakov, D., Veleva, B., Atanassov, D., Spassova, T., 2019. Seasonal changes of aerosol pollutants over Bulgaria. *Proceedings of the Fifteenth International Scientific Conference "Space, Ecology, Safety"*, Sofia, Bulgaria.

Douros, J., Eskes, H., van Geffen, J., Boersma, K.F., Compernolle, S., Pinardi, G., Blechschmidt, A.-M., Peuch, V.-H., Veefkind, J.P., 2023. Comparing Sentinel-5P TROPOMI $NO_2$ observations with the CAMS European ensemble. *Geoscientific Model Development*, 16(2), 509–534. https://doi.org/10.5194/gmd-16-509-2023.

ESA/Copernicus, 2023. *Sentinel-5P TROPOMI Product User Manual*. European Space Agency and Copernicus Programme.

Eskes, H.J., Eichmann, K.-U., 2023. *Sentinel-5P nitrogen dioxide [L2__NO2___] level 2 product readme file* (S5P-MPC-KNMI-PRF-NO2, Issue 2.4, product version 02.05.00). Royal Netherlands Meteorological Institute (KNMI) / Institute of Environmental Physics Bremen (IUPB) / European Space Agency (ESA). https://sentinels.copernicus.eu/documents/247904/3541451/Sentinel-5P-Nitrogen-Dioxide-Level-2-Product-Readme-File.

GDAL/OGR contributors, 2024. *GDAL/OGR Geospatial Data Abstraction software Library*. Open Source Geospatial Foundation. https://gdal.org.

Georgieva, E., Syrakov, D., Atanassov, D., Spassova, T., Dimitrova, M., Prodanova, M., Veleva, B., Kirova, H., Neykova, N., Neykova, R., Hristova, E., Petrov, A., 2021. Use of satellite data for air pollution modeling in Bulgaria. *Earth*, 2(3), 586–604. https://doi.org/10.3390/earth2030034.

Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3), 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, F.A., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. *Nature*, 585(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2.

Karagkiozidis, D., Alexandri, G., van Geffen, J.H.G.M., Eskes, H., Segers, A., van der A, R., Balis, D., 2024. Machine learning-based estimation of surface $NO_2$ concentrations over Europe using Sentinel-5P data. *Remote Sensing of Environment*, 300, 113890. https://doi.org/10.1016/j.rse.2023.113890.

Kendall, M.G., 1975. *Rank Correlation Methods*. Griffin, London.

Mann, H.B., 1945. Nonparametric tests against trend. *Econometrica*, 13(3), 245–259. https://doi.org/10.2307/1907187.

Oxoli, D., Cedeno Jimenez, J.R., Brovelli, M.A., 2020. Assessment of Sentinel-5P performance for ground-level air quality monitoring: preparatory experiments over the COVID-19 lockdown period. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIV-3/W1-2020, 111–116. https://doi.org/10.5194/isprs-archives-XLIV-3-W1-2020-111-2020.

Pinardi, G., Hendrick, F., 2022. *Sentinel-5P sulphur dioxide [L2__SO2____] level 2 product readme file* (S5P-MPC-BIRA-PRF-SO2, Issue 2.4, product version 02.04.01). Royal Belgian Institute for Space Aeronomy (BIRA-IASB) / European Space Agency (ESA). https://sentinels.copernicus.eu/documents/247904/3541451/Sentinel-5P-Sulphur-Dioxide-Readme.pdf (accessed 28 March 2026).

S5P Mission Performance Centre, 2023. *Sentinel-5P carbon monoxide [L2__CO____] level 2 product readme file* (S5P-MPC-SRON-PRF-CO, Issue 2.6, product version 02.06.00). Netherlands Institute for Space Research (SRON) / European Space Agency (ESA). https://sentinels.copernicus.eu/documents/247904/3541451/Sentinel-5P-Carbon-Monoxide-Level-2-Product-Readme-File (accessed 28 March 2026).

S[&]T, n.d. *HARP manual, Version 1.29*. ESA Atmospheric Toolbox (BEAT), S[&]T / European Space Agency (ESA). https://stcorp.github.io/harp/doc/html/index.html.

Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63(324), 1379–1389. https://doi.org/10.1080/01621459.1968.10480934.

State of Global Air, 2022. *Air Quality and Health in Bulgaria*. Health Effects Institute. https://stateofglobalair.org/sites/default/files/documents/2022-06/soga-southeast-europe-bulgaria-report-bulgarian_0.pdf.

Theys, N., De Smedt, I., Yu, H., Danckaert, T., van Gent, J., Hörmann, C., Wagner, T., Hedelt, P., Bauer, H., Romahn, F., Pedergnana, M., Loyola, D., Beirle, S., Ding, J., Peters, E., Hendrick, F., Van Roozendael, M., 2017. Sulfur dioxide retrievals from TROPOMI onboard Sentinel-5 Precursor: algorithm theoretical basis. *Atmospheric Measurement Techniques*, 10(1), 119–153. https://doi.org/10.5194/amt-10-119-2017.

Tonion, F., Pirotti, F., 2022. Sentinel-5P $NO_2$ data: cross-validation and comparison with ground measurements. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B3-2022, 749–756. https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-749-2022.

van Geffen, J., Eskes, H., Boersma, K.F., Maasakkers, J.D., Veefkind, J.P., 2022. TROPOMI $NO_2$ retrieval algorithm update (version 2.2). *Atmospheric Measurement Techniques*, 15(7), 2037–2060. https://doi.org/10.5194/amt-15-2037-2022.

Veefkind, J.P., Aben, I., McMullan, K., Forster, H., de Vries, J., Otter, G., Claas, J., Eskes, H.J., de Haan, J.F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., Levelt, P.F., 2012. TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of atmospheric composition. *Remote Sensing of Environment*, 120, 70–83. https://doi.org/10.1016/j.rse.2011.09.027.

Verhoelst, T., Compernolle, S., Pinardi, G., Lambert, J.-C., Eskes, H.J., Eichmann, K.-U., Fjæraa, A.M., Granville, J., Niemeijer, S., Cede, A., Tiefengraber, M., Hendrick, F., Pazmiño, A., Bais, A., Bazureau, A., Boersma, K.F., Bognar, K., Dehn, A., Donner, S., Elokhov, A., Gebetsberger, M., Goutail, F., Grutter de la Mora, M., Gruzdev, A., Gratsea, M., Hansen, G.H., Irie, H., Jepsen, N., Kanaya, Y., Karagkiozidis, D., Kivi, R., Kreher, K., Levelt, P.F., Liu, C., Müller, M., Navarro Comas, M., Piters, A.J.M., Pommereau, J.-P., Portafaix, T., Prados-Roman, C., Puentedura, O., Querel, R., Remmers, J., Richter, A., Rimmer, J., Rivera Cárdenas, C., Saavedra de Miguel, L., Sinyakov, V.P., Stremme, W., Strong, K., Van Roozendael, M., Veefkind, J.P., Wagner, T., Wittrock, F., Yela González, M., Zehner, C., 2021. Ground-based validation of the Copernicus Sentinel-5P TROPOMI $NO_2$ measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks. *Atmospheric Measurement Techniques*, 14(1), 481–510. https://doi.org/10.5194/amt-14-481-2021