Project Title	Prototype of HPC/Data Infrastructure for On-demand Services
Project Acronym	PHIDIAS
Grant Agreement No.	INEA/CEF/ICT/A2018/1810854
Start Date of Project	01.09.2019
Duration of Project	36 Months
Project Website	www.phidias-hpc.eu

4.2.1 - Contribution to Common Information Model, and Beta version of User guide, metadata and interface description for pilot users

Work Package	**WP 4, Use Case 1 - Satellite data**
Lead Author (Org)	**Hervé THEVENON (SPASCIA)**
Contributing Author(s) (Org)	**Pascal PRUNET (SPASCIA)**
Due Date	**01.03.2021**
Date	**28.05.2021**
Version	**V0.4**

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

Versioning and contribution history

Version	Date	Author	Notes
0.1	21.03.2021	Hervé THEVENON, Pascal PRUNET (SPASCIA)	TOC and V0.1
0.2	21.04.2021	Jean-Christophe DESCONNETS (IRD)	Review
0.3	05.05.2021	Hervé THEVENON (SPASCIA)	Revised edition
0.4	28.05.2021	Florian PIFFET (CINES)	Final edition

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

Table of Contents

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

Executive Summary

This deliverable presents the user guide (features, examples, etc.) for the pilot users developped in PHIDIAS.

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

1 Output

- The output of the processing is a list of `(lon, lat)` suspected of being NO_2 emission sources, ranked by relative importance of the emissions.
- The geolocations are approximate.
- The likelihood that a source is found near a positive prediction (and conversely that no source is to be found where no prediction is made) depends on how much tolerance is added to the coordinates of the prediction:
- Likelihood is above 80% within the trapezoid defined by the lower left corner `(lon-3o, lat-3a)` and the upper right corner `(lon+3o, lat+3a)` with `o` and `a` the spacing of the projection grid in longitude and latitude respectively.
- Likelihood is above 90% within the trapezoid defined by the lower left corner `(lon-5o, lat-5a)` and the upper right corner `(lon+5o, lat+5a)`.
- Tolerances are provided for each list of results.
- The list is provided as a CSV file that can be visualised in QGIS. For this purpose, two style templates (.qml) files are provided in order to facilitate the visualisation.

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

2 User defined inputs

2.1 Area of interest

The area of interest is defined by its centre and a radius. The area of interest is the area where emission sources are to be found.

- Centre: two floats – `(lon, lat)` – representing the decimal degrees coordinates of the centre of the area of interest.

- Radius: an integer in the range [30, 300] that represents the approximate `radius` in kilometres of the area to be searched for.

2.2 Observation filters

Four optional values can be used in order to filter out observations deemed unsuitable for the calculations:

- `maximum_cloud_fraction` – a float in [0, 1]. Observations whose `cloud_fraction_crb_nitrogendioxide_window` variable is greater than `maximum_cloud_fraction` are filtered out. When not provided, no filtering is performed on the cloud fraction variable.

- The `minimum_qa_value` – an integer in [0, 99]. Observations whose `qa_value` variable is lower than `minimum_qa_value` are filtered out. When not provided, no filtering is performed on the quality assurance variable.

- `timestamp_start` and `timestamp_end`, two Unix epochs that describe the period of time for which observations are to be taken into account. When not provided, these values are respectively replaced by the lowest and greatest timestamps of all the observations stored in the database.

2.3 Predictor's parameters

The current version features two arguments:

- The `disaggregation_factor` is a float in [0.25, 1] that is multiplied to the native grid (in longitude and latitude independently) in order to define the projection grid of the output.

- The `number_of_sources` to be found. Practically, the range is [20,50].

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

3 User defined inputs

At Spascia, the current user interface consist of two parts:

- A bash shell script that performs 3 tasks presented later,
- A set of style files in order to visualise the data in QGIS.

3.1 Bash shell script

- It launches a SQL query on our database to produce a CSV of all the useful observations (based on all the user inputs but the last two).
- When the preparation of the observations' CSV is completed, it launches a program (written in C and compiled on that machine) that executes the algorithm described in the companion document (Phidias - ADEGENOS - algorithm v1.0).
- It sends a SMS to the requestor when the result is ready to be downloaded.

3.2 QGIS style files

- `square_spectral_35_equal.qml` is designed to visualise average concentrations at native S5P resolution (5.5 km x 3.5 km), with a colour ramp based on the inverted spectral scheme and modified as to increase the contrast of the strongest emissions against the background of the weakest emissions.

- `class_validation.qml` is designed to visualise the predicted emission sources and to show the areas impacted by these sources.

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

4 Worked example

4.1 *Extraction of the observations*

The following bash script is modified to create a CSV of all the observations that are available for the area of interest and comply with the filters.

```
#!/bin/bash

# Connection with PSQL PSQL_CONN='/usr/local/pgsql/bin/psql -h 127.0.0.1 -p 5433
-U postgres -d no2_offl -c '

#========
function
query {

OUTPUT="/Users/postgres/herve/${DATE}_${CITY}_OFFL_${RADIUS}km.csv"

${PSQL_CONN} "COPY (
WITH
hash AS (SELECT selecthashes($LON, $LAT, 1, ${RADIUS}000)),
obs  AS  (SELECT  uid,  netcdf_id,  pr_delta_time,  pr_longitude,  pr_latitude,  pr_no2_vcd,
pr_no2_vcd_precision, in_surface_altitude, in_eastward_wind, in_northward_wind
   FROM observations,hash
   WHERE geohash_1=selecthashes AND pr_no2_qa_value>${QA} AND dt_cloud_fraction_crb<=${CLOUD}
) SELECT
netcdfs.timestamp_start, obs.*
FROM obs, netcdfs WHERE netcdfs.uid=obs.netcdf_id
) TO '${OUTPUT}' WITH (FORMAT CSV, HEADER);"

./sendSMS.sh "$CITY completed"
gzip ${OUTPUT}
}

#========
#-- Kusile Coal power station - Afrique du sud

CITY="Kusile"
LON="28.903"
LAT="-25.944"
RADIUS=300
QA="60"
CLOUD="0.4"
DATE='20210220

' query

#========
```

4.2 *Processing*

The CSV produced on step 4.1 is processed with the following

command line:

```
./no2SourceFinder -i inputs/20210220_Kavala_qa60_cc04.csv -o outputs/20210220_Kavala -f 0.25 -
c 50
```

Effectively, the program takes 4 inputs:

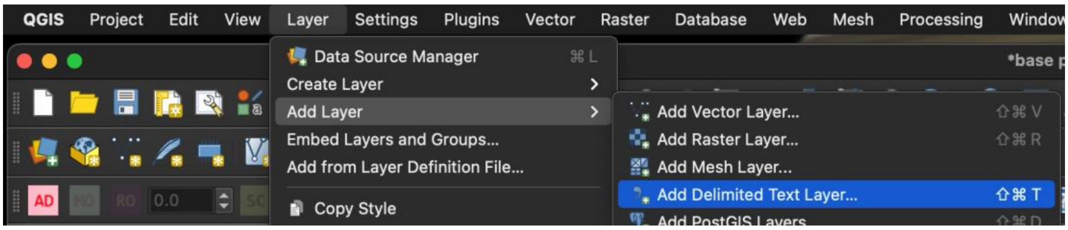
-i	the path to the CSV file used as input,

Co-financed by the Connecting Europe Facility of the European Union

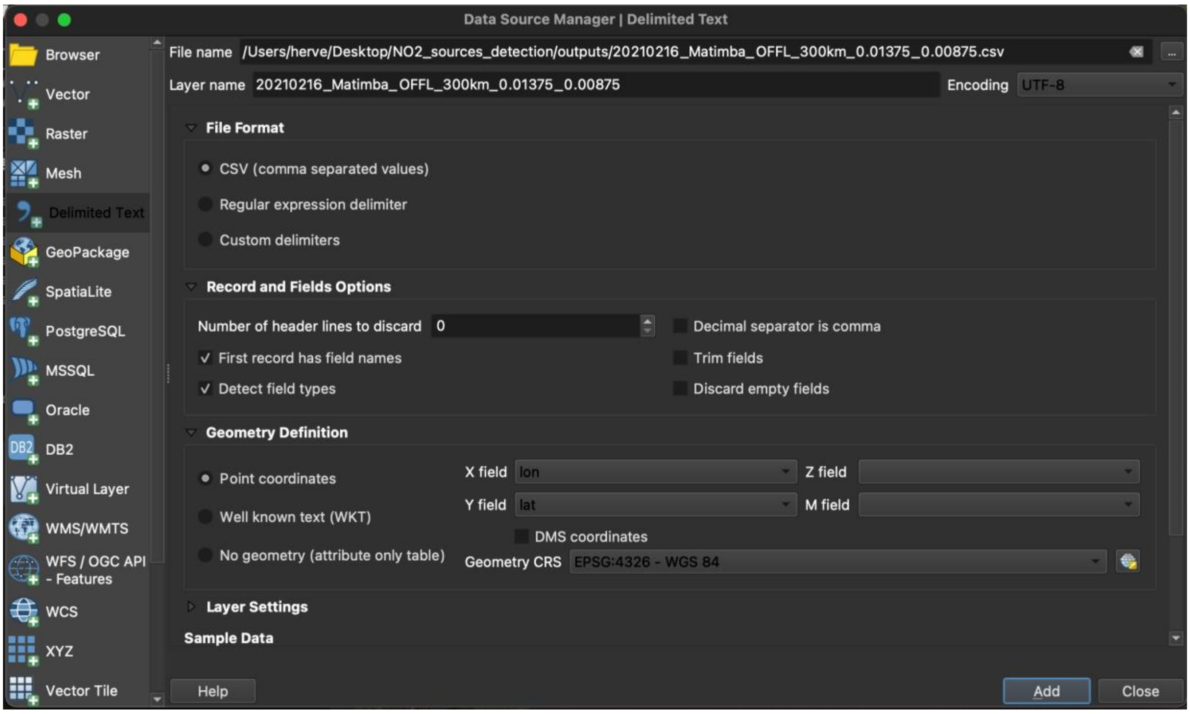
The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

-o	the path to the directory where the output file is to be stored, and the root of the filename. Given the filename provided in the example above, the program will complete the filename with the resolution of the grid (in decimal degrees) and the .csv suffix as in `20210220_Kavala_0.01375_0.00875.csv`
-f	the disaggregation factor described in section 2.3
-c	the number of sources to be found.

<table>
<tr><td>2</td><td>Open QGIS. The following snapshots are from QGIS version 3.16.1-Hannover.</td></tr>
<tr><td>3</td><td>Start a new project: from the Project menu, select New.</td></tr>
<tr><td>4</td><td>Add a CSV layer

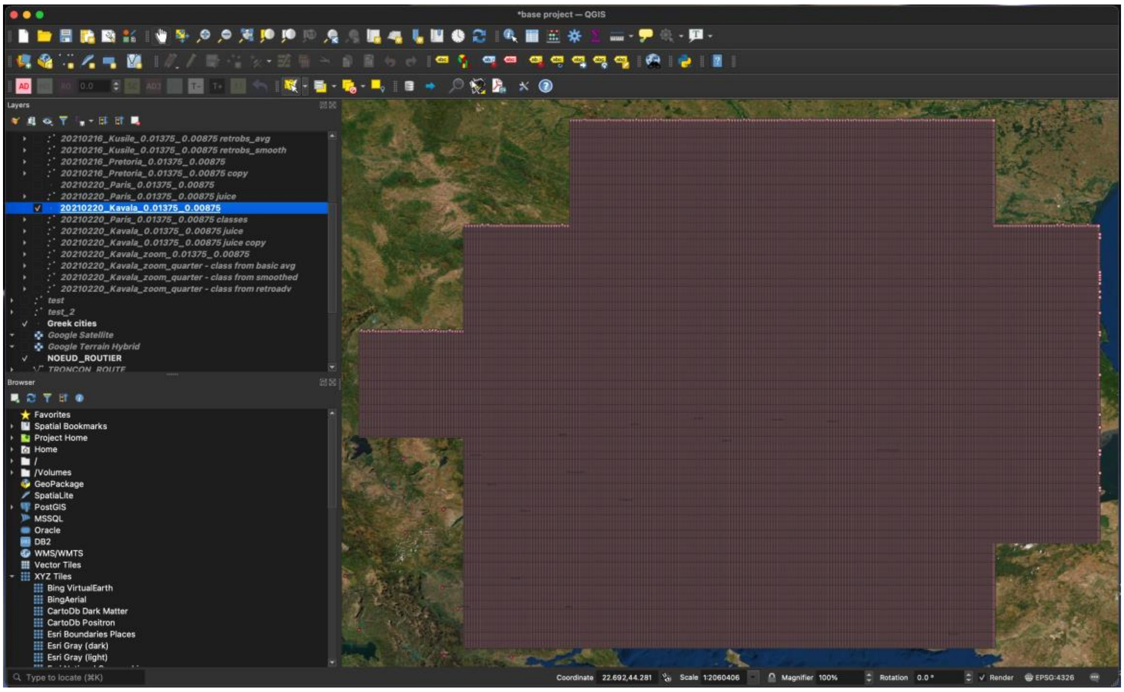


</td></tr>
<tr><td>5</td><td>Click the ellipsis button at the top right, then select the file produced in 4.2. Processing.
Make sure that First record has field names is ticked.
Further down the X field and Y field should show lon and lat respectively.
Click the Add button at the bottom right of the window, then Close.

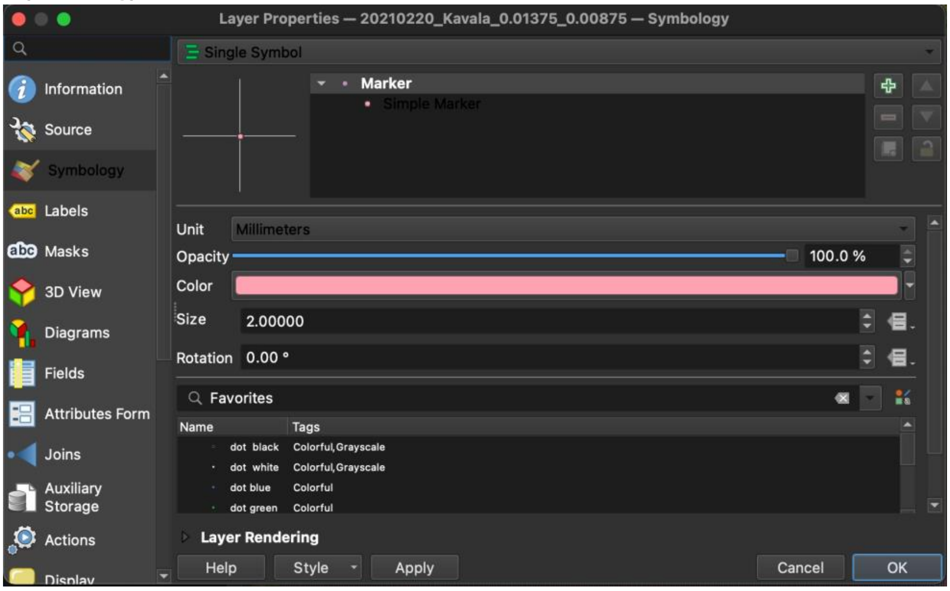


</td></tr>
</table>

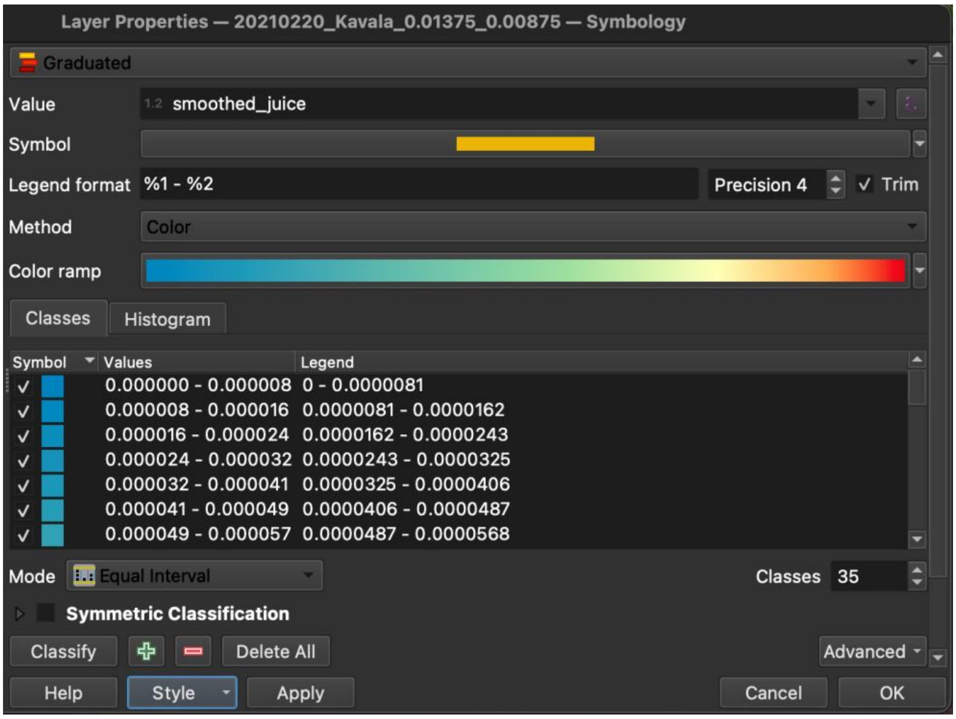
The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

6	Tick the new layer in the ***Layers*** panel.  The map is now centred on the dataset. The screenshot was taken with the ESRI basemap layer underneath. See https://opengislab.com/blog/2018/4/15/add-basemaps-in-qgis-30 in order to install basemaps.
	
7	Right click the new layer in the ***Layers*** panel, and select ***Properties...*** (located at the very bottom of the list).
8	Select the ***Source*** item on the left panel, and give a meaningful name to the layer in ***Layer name***

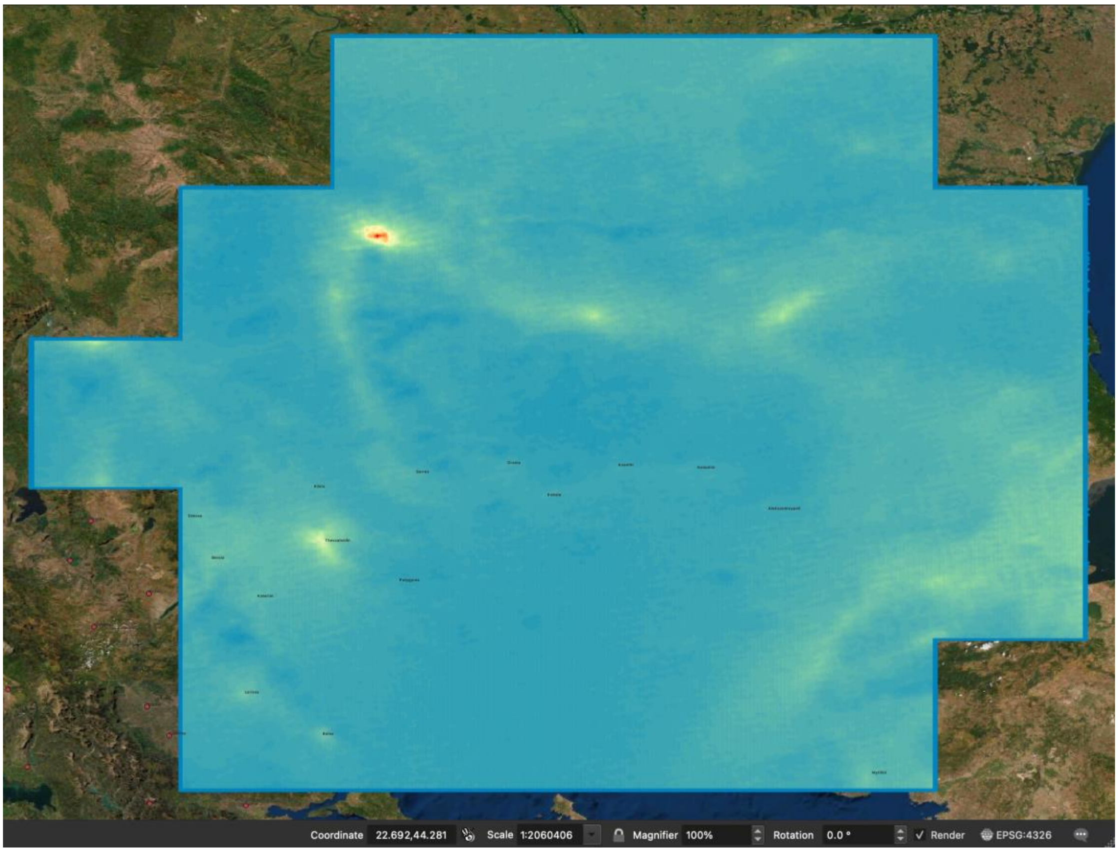
The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

9	Select the ***Symbology*** item on the left panel. 
10	Click on the ***Style*** button at the bottom of the window, and select ***Load style...***
11	In the new window that opened, click on the ***ellipsis*** button. Select the file `square_spectral_35_equal.qml` that was copied on your local drive on step 1. Click the ***Load style*** button at the bottom right. Note: - ***Single Symbol*** was replaced with ***Graduated*** - The ***Value*** selected by default is ***smoothed_juice*** - The ***Color ramp*** is the one described in section 3.2.

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.



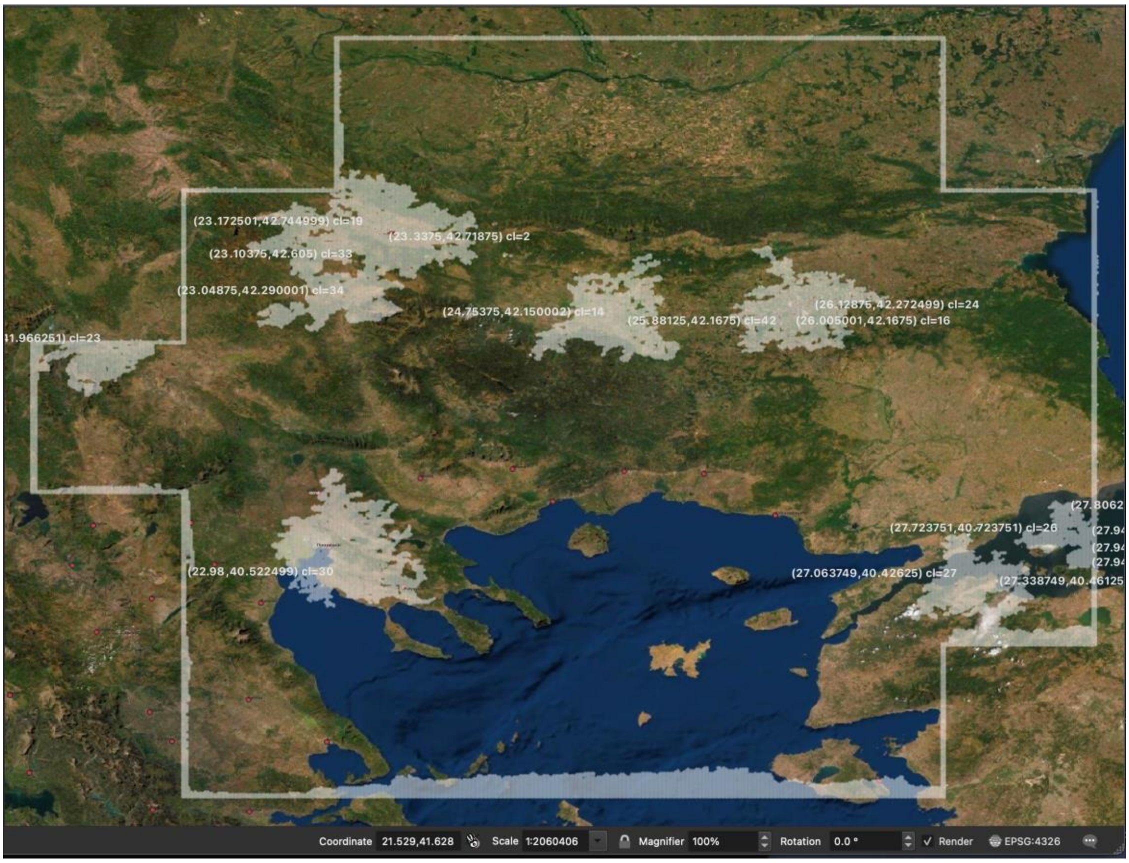
12 Click the ***Classify*** button, then ***Apply***, and ***OK***. The result is shown below.



The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

	Note #1: You may change the ***Value*** displayed in the ***Symbology*** panel. This style is specifically designed for the ***_juice*** columns found in the CSV. Note #2: By default, the style applies a pixel size equivalent to the quarter of the native S5P resolution. This should be changed according to the resolution of the grid that was used in ***4.2 Processing***. When in ***Symbology*** panel, click on the ***Symbol*** (not on the arrow at the right of the line), select ***Simple Marker***, and change the ***Size*** to suit the resolution of the grid. Quarter of native resolution is approximately 800 ***Meters at Scale*** on the smaller axis (latitude).
13	In the Layers panel, - untick the layer, - right click on the layer, select Duplicate Layer. - Select and tick the new layer - Right click the new layer, select Properties.
14	In the ***Symbology*** section, Click the ***Style*** button, select ***Load Style...***
15	Click the ***ellipsis*** button in order to find and select the file *class_validation.qml* Click the ***Load style*** button at the bottom right.
16	Click the ***Classify*** button, then ***Apply***, and ***OK***. The result is shown below. Sources are shown as red filled squares. The empty squares with a white outline are the locations that are considered as polluted by the emission sources.

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.



Note #3:
You may change the ***Value*** displayed in the ***Symbology*** panel.
This style is specifically designed for the ***_class*** columns found in the CSV.

Note #4:
The uncertainty related with the geolocation of the sources is discussed in the companion document.

Note #5:
Following the installation of the basemaps suggested in step 6, a basemap like the *Google Terrain Hybrid* map may be used to see names of urban areas and industries.

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

4.3 Examples files

The following files are included in the package:

`Phidias - ADEGENOS - User Guide.pdf`	This document
`Phidias - ADEGENOS - Algorithm.pdf`	This document's *companion* document.
`square_spectral_35_equal.qml`	The QGIS style sheets used in the worked example section.
`class_validation.qml`	
`20210220_Kavala_qa60_cc04.csv`	Sample result from the extraction step (section 4.1). CSV file. Approximate size: 280 MB.
`20210220_Kavala_0.01375_0.00875.csv`	Sample result from the processing step (section 4.2). CSV file. Approximate size: 50 MB. May be used from step 4 of the worked example (section 4.3).

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

5 Perspectives

5.1 FAIR principles

So far the outputs lack some characteristics in order to comply with the FAIR requirements. This will be addressed pragmatically while improving ease of use, functionality, and integration with PHIDIAS production system.

5.2 Binary serialisation of the outputs

The development will include binary serialisation of the outputs, while maintaining accessibility and interoperability.

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.