



D4.16.2

Report on optimal strategy for integrating the identified tracers



Deliverable number:	D4.16.2
Work package:	WP4 – Atmosphere
Intermediate Objective:	IO4.5
Deliverable type:	<input checked="" type="checkbox"/> Document, report
	<input type="checkbox"/> Websites, patent filings, videos, etc.
	<input type="checkbox"/> Other: please specify
Dissemination level:	<input checked="" type="checkbox"/> Public
	<input type="checkbox"/> Restricted
Estimated delivery (bimester):	B10
Actual delivery date:	30-06-2024
Author(s) (Partner-OU):	Salvatore Sinopoli, Claudia Roberta Calidonna, Giorgia De Benedetto, Malacaria Luana; Lo Feudo Teresa; Ammoscato, Ivano (ISAC LT)
Reviewed by:	Lucia Mona (CNR-IMAA), Gianluca Di Fiore (CNR-IMAA)
Note:	

*IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System - CUP B53C22002150006 (D.D. n. 130/2022)
Funded by EU - Next Generation EU
Mission 4 “Education and Research” - Component 2: “From research to business” -
Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”*

Index

<i>LIST OF FIGURES</i>	4
1. <i>INTRODUCTION</i>	5
2. <i>MATERIALS</i>	6
2.2 Satellite data	11
2.3 Data from weather-climate model	12
3. <i>ANALYSIS METHODOLOGIES</i>	15
3.1 Homogeneous Time Aggregation of Data Series	16
3.2 Analysis instruments of background data for the historical series	20
3.2.1 Smoothed Minima Approach	21
3.2.2 Background Data Selection – BaDS	25
3.2.3 Wind-Driven Background Selection	27
4. <i>RESULTS AND FUTURE WORKS</i>	32
<i>BIBLIOGRAPHY</i>	35

LIST OF FIGURES

Figure 1: Example of a RAW data from a Picarro, showing a row with sample time at each five seconds.	6
Figure 2: Example of a scatterplot for CO2 data at LMT site.....	7
Figure 3: Example of data in spatial map format. In this case a CO2 map from satellite products....	7
Figure 4: Example of transformation from a figure to an actual raster, with the actual loss of definition	8
Figure 5: Location of the LMT site in the Mediterranean Basin	8
Figure 6: Location of the LMT site, the yellow line is the A2 Highway, and the yellow dots are the municipalities of Lamezia Terme and Catanzaro.....	9
Figure 7: Picarro G2401 at GAW/WMO LMT site	10
Figure 8: Thermo 5012 MAAP at GAW/WMO LMT site	10
Figure 9: Wind Rose for LMT Site based on data from 2015-2022 period of observations	11
Figure 10: Main fire events in Calabria between August 8th and 12th 2021, tracked from MODIS satellite imaging (red spots).	12
Figure 11: WRF (2-km resolution) 10 m wind speed (m/s) and wind vectors, for August 8th, 2021 at 08:00 pm UTC (a) and 11:00 pm UTC (b); the red circle approximately indicates the LMT observatory	13
Figure 12: Representation of Regional Registry of Wild-Fires for Calabria in 2021. Red dots are the fire occurrences, yellow dots are the cities of Lamezia Terme and Catanzaro, and in blue the LMT site.	14
Figure 13: From data to Knowledge through Information.....	15
Figure 14: Example of two contemporaneous peaks in two different variables	16
Figure 15: Loop for the pre-analysis of the month folder.....	17
Figure 16: Loop for the analysis and aggregation of the daily file and aggregation in a new time format	19
Figure 17: Example of the Data-Base for LMT-site with mean hour aggregation;	19
Figure 18: CO2 measurements s in CMN site and in BIS site. The CMN is a very clean site, and the BIS site instead has seen a lot of disturbances.	21
Figure 19: Example of the comparison criterion for the smoothed minima approach.	22
Figure 20: Example of variation in technique for the maximum values.	23
Figure 21: in the upper figure the pre-analysis on LMT-CO2 dataset, in the bottom figure the selection of the completed procedure.	24
Figure 22: CH4 Background Selection made by BaDS at LMT Site.	26
Figure 23: CO Background Selection made by BaDS at LMT Site.	26
Figure 24: CO2 Background Selection made by BaDS at LMT Site.	27
Figure 25: correlation of wind direction and CO2 levels in July 2019, the air coming from seaside is much more clean than the part coming from land;.....	28
Figure 26: Comparison between the wind selection (green) and the Bads selection (red) for CO2 at LMT site.....	29
Figure 27: Comparison between the wind selection (green) and the Bads selection (red) for CH4 at LMT site.....	30
Figure 28: Transformation of the vectorial data in a gridded data. In a) the complete grid system, in b) the shapes of the fire areas near LMT site.	31
Figure 29: Loop of analysis to mark time, direction, location and type of an emission responsible of instrumental anomalies.....	32
Figure 30: Example of application of the method to track the open fires.....	34

1. INTRODUCTION

The main objective of this report is to identify which methods and strategies are suitable for integrating atmospheric data from different sources to support the identification of fire emission tracers in the framework of ITINERIS.

Among the types of data available, the most common is the collected time series format from most of the distributed instruments placed within a climate and weather data observatory.

This type of data can refer to one or more observation points distributed over the national territory. The previous deliverable (D4.3.2) reported: i) a list of instruments; ii) the duration of the collected time series; iii) the spatial location of the observation (lon, lat), and the reference to the operational status related to the OU available to share their measurement programme for the Fire Pilot activity within the ITINERIS project.

The main purpose of the pilot is to make the analysis related to fire emissions, both anthropogenic and/or natural, but together with instrumental atmospheric data, it is also important to relate them to the spatial localisation of the fire event occurrence. In this way, we can investigate the possibility of correlating the occurrence of events with anomalies of selected data recorded in the time series.

In general, stations in or near anthropised areas show massive fluctuations in the time series of measured data throughout the year. This is because emissions from anthropogenic activities are continuous and always consistent, and they can overlap with the actual trace of fire activity.

Therefore, the study of historical instrumental data without the aid of a fire occurrence log could lead to errors, and it is not possible to clearly identify the presence of a fire event within the signals of each instrument. This could lead to the assumption that anomalies recorded by the instruments are data from fire trails, when in fact they may be an overlap of the effects of different events. In the following sections of this report, it will be clearly shown that it is necessary to correlate the anomaly in detected data of each instrument with the simultaneous presence of one or more fire emission sources.

In particular, the analysis carried out focuses on the Lamezia Terme (LMT) site. By analysing the available data collected here, it is possible to integrate instrumental data of two different types: gaseous species and/or aerosol-type particles. Furthermore, it has been possible to reconstruct a fire register with a coverage of years comparable to that of the instrumental data.

On the basis of such a case, it is possible to compare the data collected in a site when simultaneous presence of fire events occurs, thus creating a real space-time correlation pattern on the dataset.

2. MATERIALS

Based on the review of available fire related data reported in Deliverable D4.16.1, the Lamezia Terme site has been identified as a pilot site for fire emission analyses, studies and methodologies as it provides a comprehensive and extensive dataset in terms of the variety of available instrumentation [1].

There are two types of different available information:

- **In situ instrumental data**, collected at the Regional WMO/GAW observatory (LMT) in Lamezia Terme. This provides a continuous monitoring of several chemical-physical parameters since 2015 (ongoing) [2, 3]. All the equipment provides data on the different parameters measured at the LMT site. These datasets are in the form of text files consisting of different columns, each of which refers to the specific variable studied. The rows of the text files are generally referred to as successive time steps, which can be divided into different sampling intervals for each species. These are specific to each instrument and depend on the sampling interval chosen.

```
date time daydec ALARM_STATUS INST_STATUS CavityPressure CavityTemp DasTemp EtalonTemp WarmBoxTe
2015-02-03 00:00:00 33.000000 0 963 1.4001382446E+02 4.4999961853E+01 3.6625000000E+01 4.4992271
2015-02-03 00:00:05 33.000058 0 963 1.4001141357E+02 4.5000053406E+01 3.6687500000E+01 4.4992362
2015-02-03 00:00:10 33.000116 0 963 1.3999583185E+02 4.5000073675E+01 3.6645988806E+01 4.4992362
2015-02-03 00:00:15 33.000174 0 963 1.3999636691E+02 4.5000083923E+01 3.6687500000E+01 4.4992362
2015-02-03 00:00:20 33.000231 0 963 1.3996664429E+02 4.5000114441E+01 3.6687500000E+01 4.4992332
2015-02-03 00:00:25 33.000289 0 963 1.3996799400E+02 4.4999565125E+01 3.6687500000E+01 4.4992362
2015-02-03 00:00:30 33.000347 0 963 1.3998316329E+02 4.4999624110E+01 3.6687500000E+01 4.4992379
2015-02-03 00:00:35 33.000405 0 963 1.4002130127E+02 4.4999717712E+01 3.6687500000E+01 4.4992362
2015-02-03 00:00:40 33.000463 0 963 1.3999775696E+02 4.4999717712E+01 3.6687500000E+01 4.4992362
2015-02-03 00:00:45 33.000521 0 963 1.3999092508E+02 4.4999778748E+01 3.6687500000E+01 4.4992424
2015-02-03 00:00:50 33.000579 0 963 1.3999473572E+02 4.4999809265E+01 3.6750000000E+01 4.4992576
2015-02-03 00:00:55 33.000637 0 963 1.4000170523E+02 4.4999975290E+01 3.6750000000E+01 4.4992651
2015-02-03 00:01:00 33.000694 0 963 1.3996671535E+02 4.4999992371E+01 3.6750000000E+01 4.4992668
2015-02-03 00:01:05 33.000752 0 963 1.4001992030E+02 4.4999900818E+01 3.6812500000E+01 4.4992668
2015-02-03 00:01:10 33.000810 0 963 1.3999583435E+02 4.4999992371E+01 3.6812500000E+01 4.4992851
2015-02-03 00:01:15 33.000868 0 963 1.3997348956E+02 4.4999881004E+01 3.6812500000E+01 4.4992772
2015-02-03 00:01:20 33.000926 0 963 1.4001884503E+02 4.4999809265E+01 3.6812500000E+01 4.4992820
2015-02-03 00:01:25 33.000984 0 963 1.4001258850E+02 4.4999870300E+01 3.6812500000E+01 4.4992820
2015-02-03 00:01:30 33.001042 0 963 1.3998875427E+02 4.5000083923E+01 3.6875000000E+01 4.4992851
2015-02-03 00:01:35 33.001100 0 963 1.4002267456E+02 4.5000419617E+01 3.6875000000E+01 4.4992942
2015-02-03 00:01:40 33.001157 0 963 1.4000916488E+02 4.5000083923E+01 3.6930069930E+01 4.4992973
```

Figure 1: Example of a RAW data from a Picarro, showing a row with sample time at each five seconds.

This type of data provides particularly high resolution, with a very small time step (even in the order of a second), but with reference to a specific, individual location (the Lat/Lon coordinate). The data validation procedure requires the data to be cleaned (looking for errors in the writing of the file, when the instrument was fully operational, in the calibration routine procedure, electrical malfunctions, etc.) before proceeding to the next analysis phases [4]. The cleaned data can be presented as scatter plots, as shown in the figure below, and are therefore suitable for extrapolation of information;

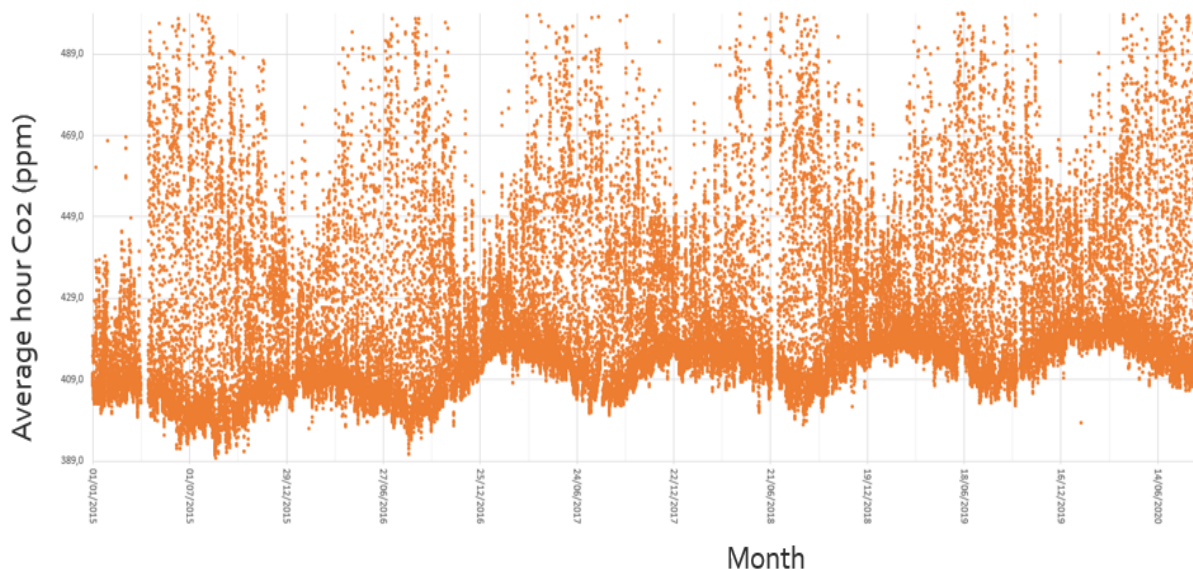


Figure 2: Example of a scatterplot for CO2 data at LMT site

- **Gridded data from models or satellites observations**, not derived directly from in situ observations, but from models, satellites, or local mapping of events. These data are in the shape of spatialised values, also time-varying (i.e. multiple maps of the same species at different times). This type of data is usually in raster or vector format, processed by GIS or specific information extrapolation algorithms;

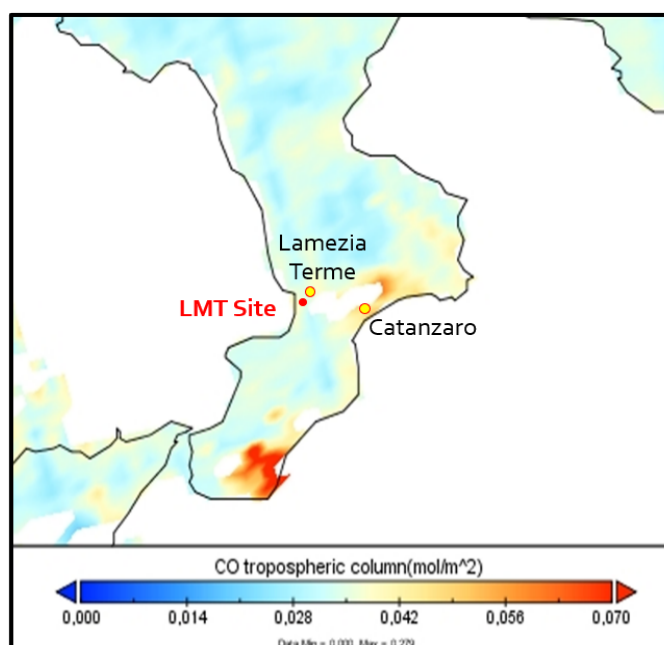


Figure 3: Example of data in spatial map format. In this case a CO2 map from satellite products.

It should be noted that, in this case, the choice is to have an overview of the variable over a large area, while accepting the compromise of losing information on its spatial distribution. Maps cannot provide precise punctual data, but they are the graphical representation of a single value attributed to the whole area represented by the spatial unit. By choosing the dimension of the spatial unit that will

cover the area of representation (cell), it is possible to decide how large the area is for the given value.

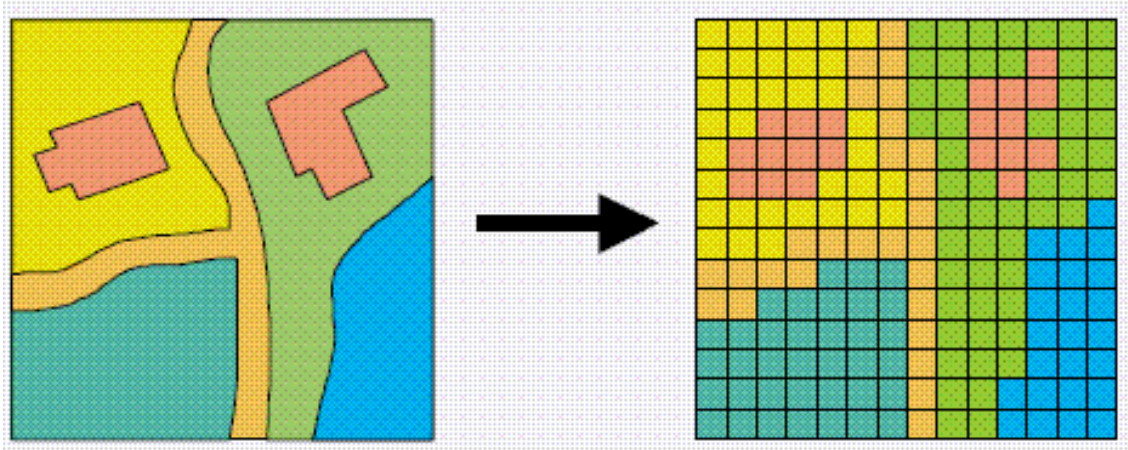


Figure 4: Example of transformation from a figure to an actual raster, with the actual loss of definition

The larger the cell dimension, the greater the approximation of the data expressed by the value paired to the cell, and consequently the worse the approximation. By worsening the spatial approximation, the phenomena are "averaged" over space, thus reducing the extreme values [5]. The size of the representative cells to be used depends strongly on the computational resources available and the extension of the spatial domain to be interpreted.

3.1 Geographical characterisation ad instrumental data of the pilot site LMT

The CNR-ISAC observatory in Lamezia Terme (Code: LMT, 38.88 N 16.23 E; 6 m.a.s.l.) is a regional WMO/GAW station.

The site is located near the coast, about 1000 metres from the coast of the southern Tyrrhenian Sea. The position of the measuring point with respect to Calabria places the observatory in the middle of the Mediterranean Sea, making it a special observation point [2,3].



Figure 5: Location of the LMT site in the Mediterranean Basin

At the same time, the area is subject to strong winds, almost always E-W or W-E, due to the presence of the channel between the two seas. This is a depression in the orographic barrier represented by the Calabrian Apennines, which connects the air masses coming from the Ionian coast with those of the Tyrrhenian side. The area is therefore characterised by the presence of strong and persistent winds; this characteristic has favoured the choice of this site as a location for numerous wind turbines over the years [3].

The site is not located in a remote area, but rather in an area with a high anthropic presence, and therefore polluted by activities and traffic. In particular, the area is dedicated to industry and agriculture and is affected by the nearby presence of the A2 motorway and the city of Lamezia Terme (approx. 70,000 inhabitants), both located to the north-east of the coast.



Figure 6: Location of the LMT site, the yellow line is the A2 Highway, and the yellow dots are the municipalities of Lamezia Terme and Catanzaro

The observatory is characterised by the existence of different data acquisition equipment, installed in 2015 and still present today, with a particularly consistent and long-lasting data-set available (about 9 years).

The equipment is briefly listed and described below:

- **Picarro G2401:** provides simultaneous and accurate measurements of carbon monoxide (CO), methane (CH₄), carbon dioxide (CO₂) and water vapour (H₂O). The patented CRDS technology (Cavity Ring-Down Spectroscopy) allows an effective measurement path length of up to 20 kilometres in a compact cavity, which results in high sensitivity and precision in a compact analyser;



Figure 7: Picarro G2401 at GAW/WMO LMT site

- Thermo 5012 MAAP** operates in accordance with WMO/GAW specifications [6] to monitor the concentrations of equivalent black carbon (eBC) at 637 nm continuously with a time resolution of 1 minute. The sample is inhaled into the instrument through the inlet from the internal pump, a visible light source at 670 nm is directed onto the deposited aerosol measuring the light transmission reduction, reflection intensities and the air sample volume, in order to provide real-time eBC concentration.



Figure 8: Thermo 5012 MAAP at GAW/WMO LMT site

- Vaisala WTX520**, is an automatic weather station that collects meteorological variables such as barometric pressure, temperature, wind speed and direction, accumulated rain (averaged over 10 minutes) and relative humidity, located at about 10 mt a.s.l.

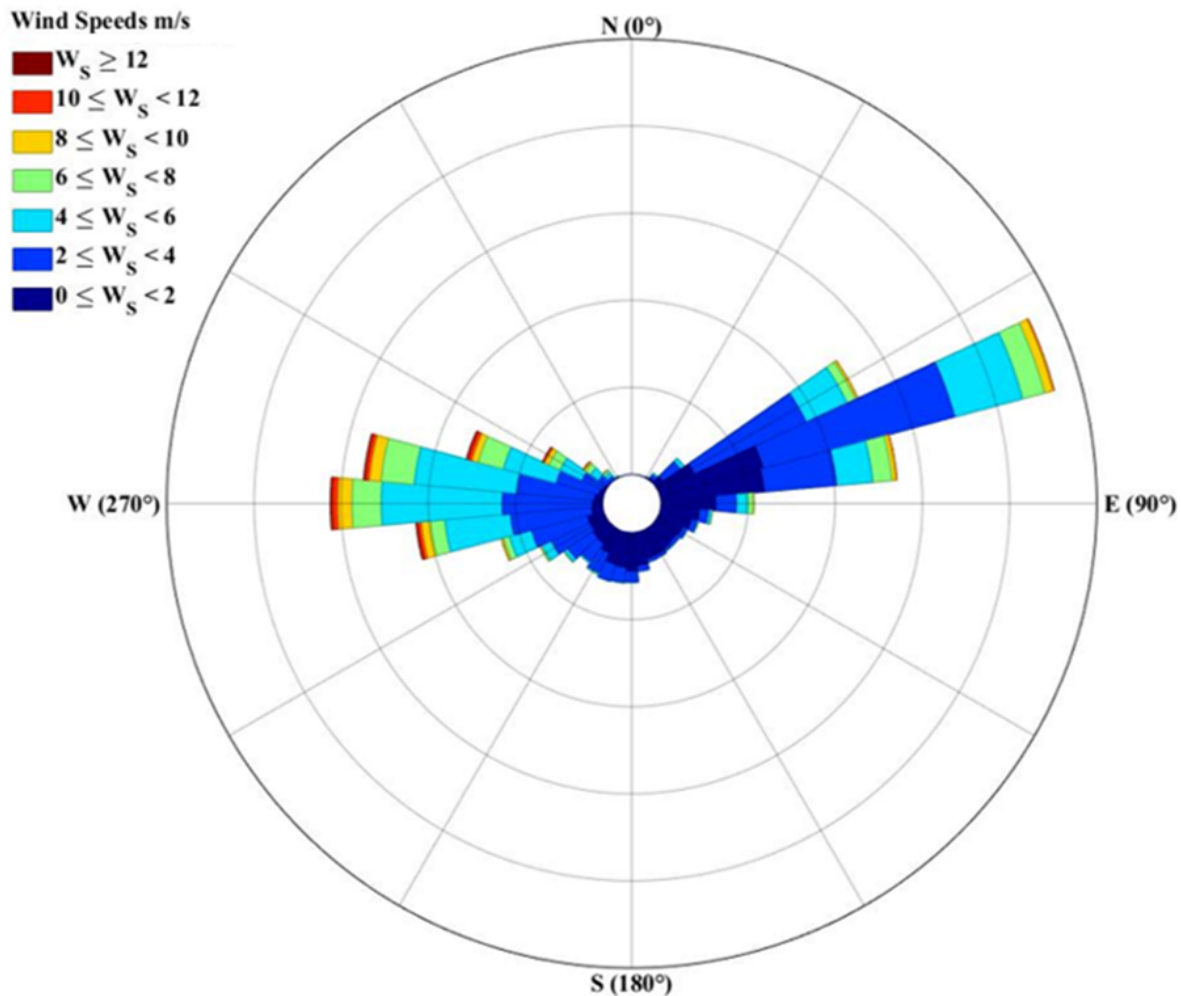


Figure 9: Wind Rose for LMT Site based on data from 2015-2022 period of observations

In particular, it should be noted that the area is buffeted by winds coming mainly from the E and W, which can reach speeds of more than 10 m/s.

There are other instruments at the site to collect data on both reactive gases (NO_x, NO₂, etc.) and aerosol particles. At this stage they have not been taken into account directly, since the main products of fire emissions are CO and black carbon, and these two species have been identified as the main tracers [1]. The other data are also aggregated in the analysis, but in the case of instruments to be installed in the future, their presence will serve to confirm and further characterise the type of event registered.

At this stage, however, it is important to isolate the primary tracers to distinguish fires within the data set.

2.2 Satellite data

For summer fire counts, Global Fire Emission Dataset-Version 4 (GFED4) and the online GFED Analysis Tool are available, derived from MODIS (Moderate Resolution Imaging Spectroradiometer) L2 fire products, with data resolution of 1 km (at NADIR).



Figure 10: Main fire events in Calabria between August 8th and 12th 2021, tracked from MODIS satellite imaging (red spots).

In the Figure 3-10 an example of detected fire events occurring during 8th -12th August 2021 from MODIS is reported.

This satellite dataset is in raster format, and so the red cells represent the area that undergo wildfire events in the previous days.

2.3 Data from weather-climate model

Numerical Weather Forecasting models can be used to evaluate the prevailing circulation conditions during a fire occurrence, this to help in identification of possible burnt area contributing with emission. In particular, the non-hydrostatic Weather Research and Forecasting (WRF) model version 4.2, was considered [7]. Two nested domains are used: a 10 km domain covering Southern Italy adequate for the mesoscale in the study area, which is orographically complex, and a 2 km domain for the detailed area near the observatory. The model is implemented with 40 vertical levels following the terrain, more strictly spaced in the limiting layer, 15 of them below 2000 m of altitude. The initial and boundary conditions have been derived from the analyses/previsions of the National Center for

Environmental Prediction - Global Forecast System (NCEP - GFS) (horizontal resolution of 0.25 degrees) and the model configuration is derived from those operationally adopted by the CNR-ISAC and borrowed from recent works concerning other extreme meteorological event regarding wind in Italy [7]. The format of these files is a raster aggregation of the same spatial extension but at different time instants, in the type of NetCDF format, typical for this kind of analysis.

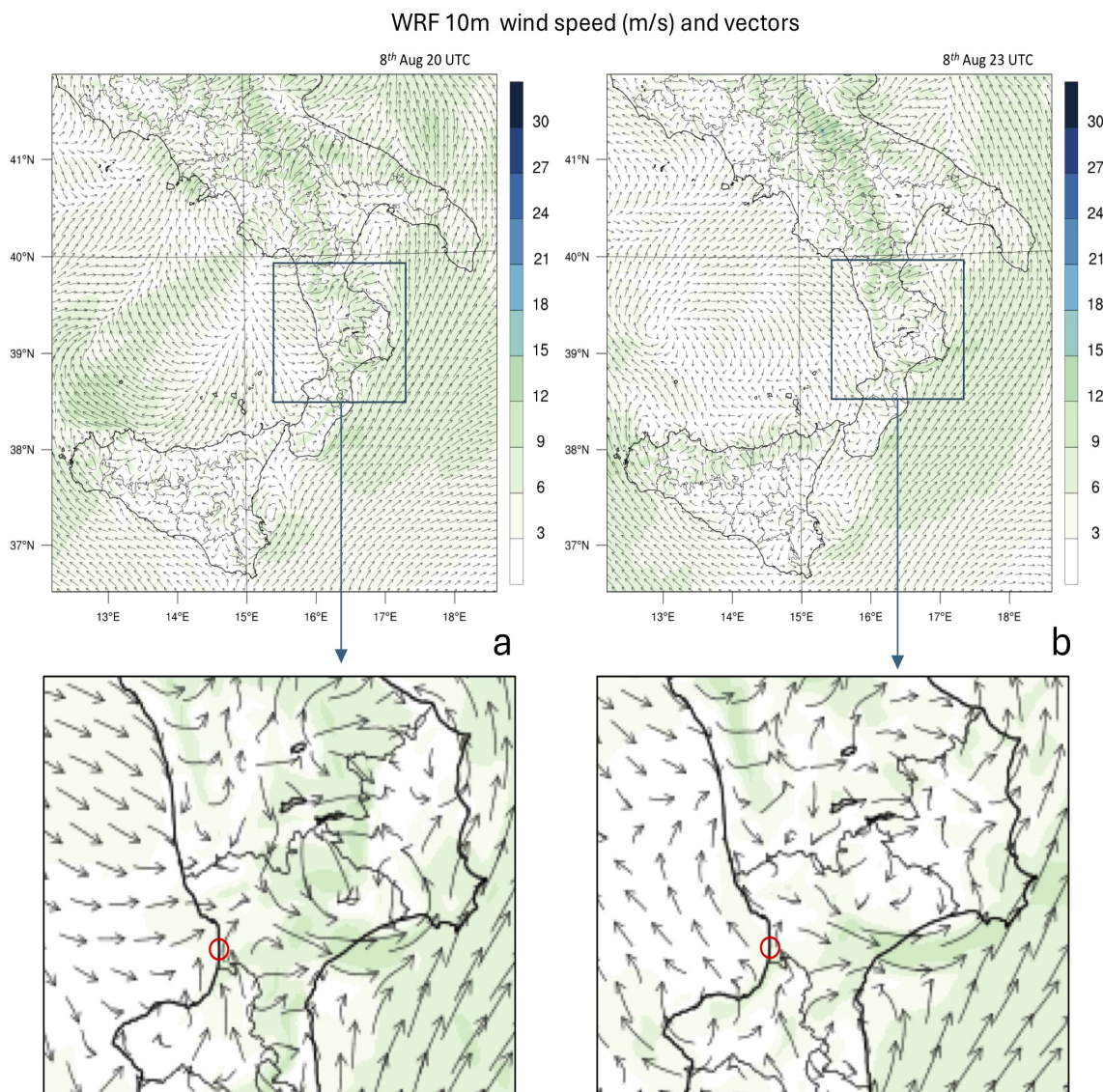


Figure 11: WRF (2-km resolution) 10 m wind speed (m/s) and wind vectors, for August 8th, 2021 at 08:00 pm UTC (a) and 11:00 pm UTC (b); the red circle approximately indicates the LMT observatory

3.4 Data from regional fire register

In addition to the instrumental and modelling types of data, records of the regional fire register are useful. We used such a register, available from Regional Calabrian Centre of Civil Protection Fire Division, with the similar time coverage of the LMT dataset. In the Calabrian Region the fires registry has a record of more than 800 fire occurrences for each year (fires in wooded or anthropised

areas) is available as a polygon-type vector shapefile (centroids are shown in the figure for a better visualisation for the 2021 year).

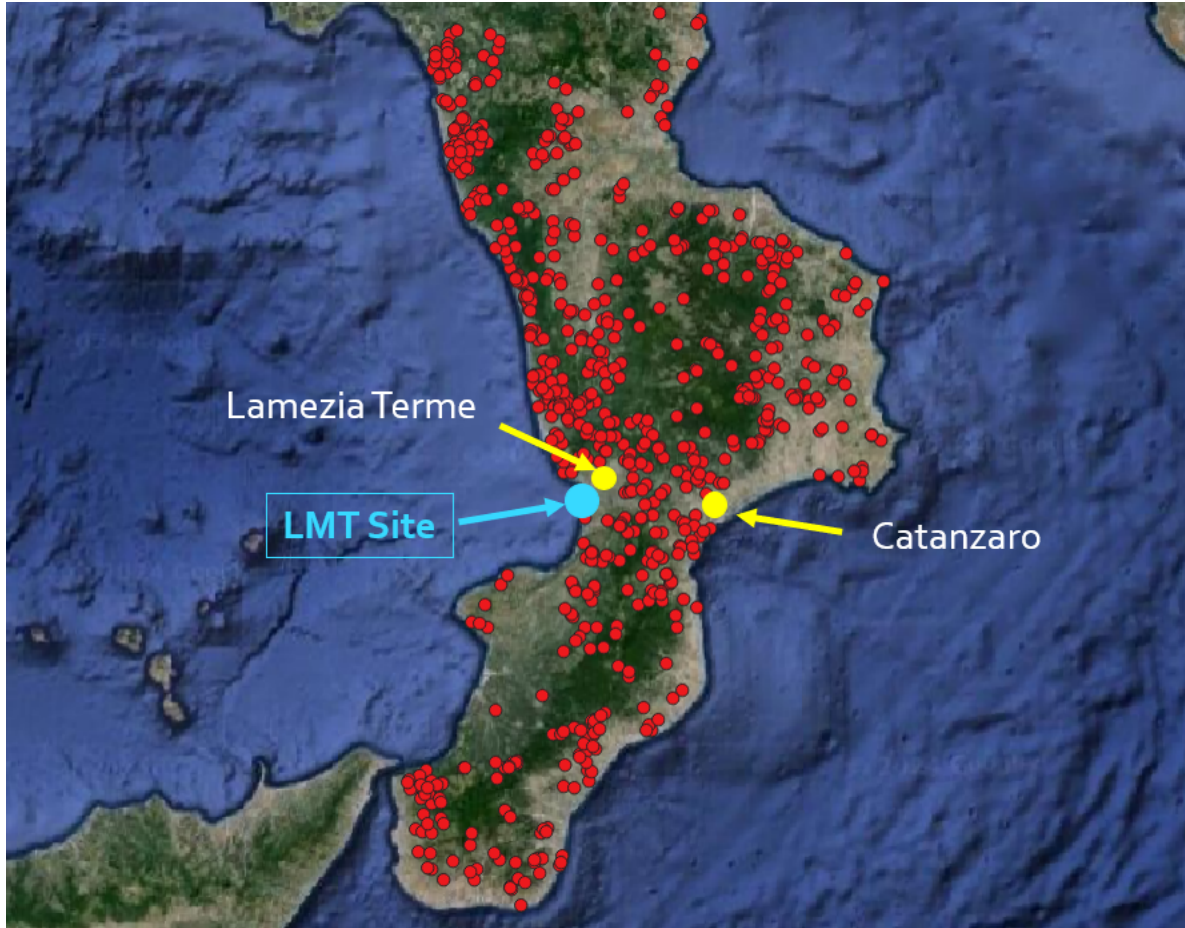


Figure 12: Representation of Regional Registry of Wild-Fires for Calabria in 2021. Red dots are the fire occurrences, yellow dots are the cities of Lamezia Terme and Catanzaro, and in blue the LMT site.

Such information provides feedback on the presence of a fire event near the LMT site and could be useful for the validation of the procedures used in the analysis of some cases.

3. ANALYSIS METHODOLOGIES

The methodology used here consists of comparing data from different instruments, models, satellites or digital event log systems, to identify tracers and study correlations between the different variables observed during a given event. The aim is the “data integration” in order to extract the best information from the whole dataset collected by different instruments detecting different sources (tracers)[8]. this is a particularly complex challenge, and requires a series of technical steps in order to prevent and resolve errors. Particularly when examining environmental data, one needs to discriminate between background and event data, but also to be able to read overlapping effects from different sources [9, 10, 11].

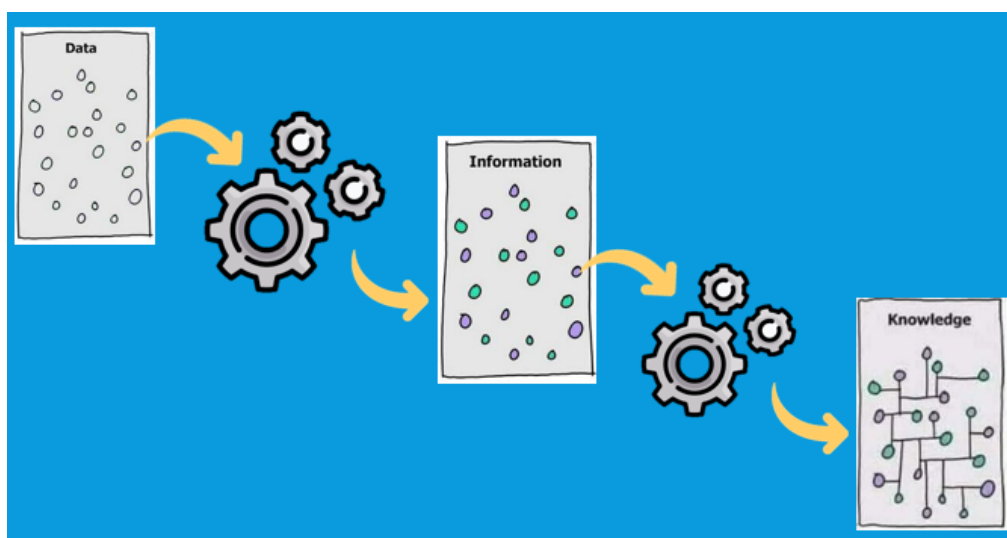


Figure 13: From data to Knowledge through Information

As shown in the previous section, all the data are in various formats and with different aggregations, both spatial and temporal. So, the first effort was time and space *homogenisation* in order to aggregate the various datasets into a format suitable for temporal or spatial comparison.

With the intention of carrying out correlation analyses, the heterogeneous data coming from the instruments, need to be made temporally homogenous in order to be consequently aggregated and analysed. This operation requires several steps, each of which prepare the next. Since the amount of data to be processed is particularly wide, and the iterative process has to be repeated thousands of times, it has been chosen to proceed through the development of different calculation codes. In this report, developed procedures are shown in a synthetic and schematic way, which will be detailed with the algorithms in the deliverable D4.16.3, in relation to data acquisition and processing.

As this is an analysis of fire emission data, the aim is to obtain a continuous sequence of more directly comparable variables, to verify the presence of almost simultaneous anomalies in the data sequences. Simultaneity is not to be understood as two events at the exact moment in time, but as anomalies of different variables in the same "event window". This is due to the fact that gases and aerosols do not travel in exactly the same way in air masses.

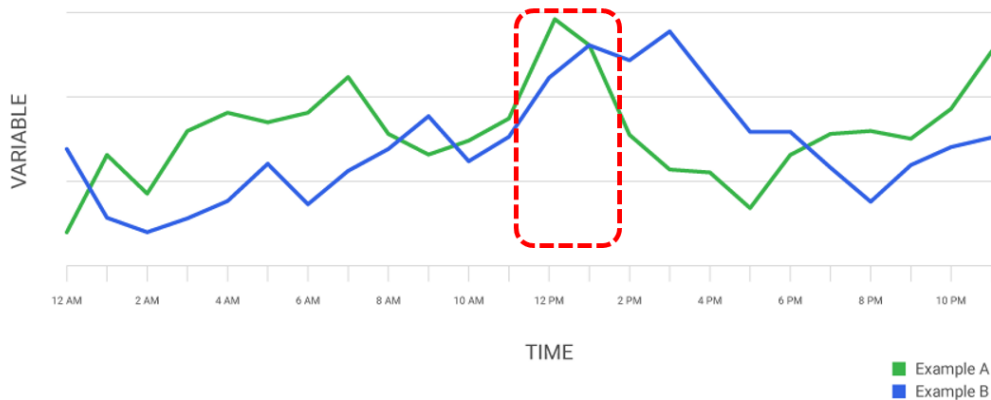


Figure 14: Example of two contemporaneous peaks in two different variables

The increase in fire-specific tracers can therefore be interpreted as a signal of the presence of a more or less intense and more or less nearby emissive source, that due to the increase of events in the Mediterranean [12, 13] is becoming a phenomenon increasingly impactful but increasingly difficult to interpret [14].

3.1 Homogeneous Time Aggregation of Data Series

To obtain a different sequence for each species, it is necessary to aggregate in situ data registered in time sequence format from each instrument, but at the same time guarantee that each sequence has the same time-discretization step.

Registered measurements from each instrument are not directly usable, as each data is stored within a specific file usually in **raw** format (text type). Such file type consists of a line for each sampled measure at each time-step, comprehensive of measured data, and other parameters necessary to register at the same time the status of the equipment.

The additional information allows us to understand the validity of measured data, or if it has to be discarded for some reason, either due to an equipment fault or a calibration stage with air cylinder input (so not sampling from the atmosphere). In both cases, such data are discarded and not analysed.

In this phase we focused on analysing greenhouse gases (CO₂, CH₄ and CO), black carbon (eBC) and meteorological data from previously described equipment.

Each of these three devices produced for every day of the year, starting from January 2015 to 31 December 2023, a file containing all the data sampled at a specific time. The following table summarises the sampling resolution for each considered instrument:

Table 1: Instrumentation considered for the main tracers for open fires

Instrumentation	Sample Time	Variables for Analysis
Picarro G2401	5 seconds	CO; CO ₂ ; CH ₄ ; H ₂ O
Thermo MAAP	1 minute	eBC
Vaisala WTX520	1 minute	Wind Speed, Wind Direction, Air temperature, Rain, Humidity, Atmospheric Pressure

The daily data from each instrument was first divided by year and then organised into monthly folders to make it easier to manage.

The first step to be performed in order to successfully carry out all subsequent steps is a data cleaning procedure [15, 16, 17]. This part is preliminary, serving to identify every possible type of error (incorrect data, partially written, missing files, etc.). For each folder, a first analysis cycle was developed, consisting of:

- A check on the number of raw files present inside each folder. This check allows to verify if the data for each day of the year have actually been saved;
- A check on the header format for every daily file present in the folder. This check allows to verify that the format in which the data is saved has not changed;
- A check on the correspondence between the number of actual sampling time resolution foreseen by the procedure and the number of rows stored in each file. This check is used to verify that the integrity of the individual file has been respected, and whether there are any problems within the file itself;

For each of these three steps, specific logs are created to speed up the subsequent analysis procedures. The following figure shows a schematic of the analysis loop used for each monthly folder (Figure 4-3).

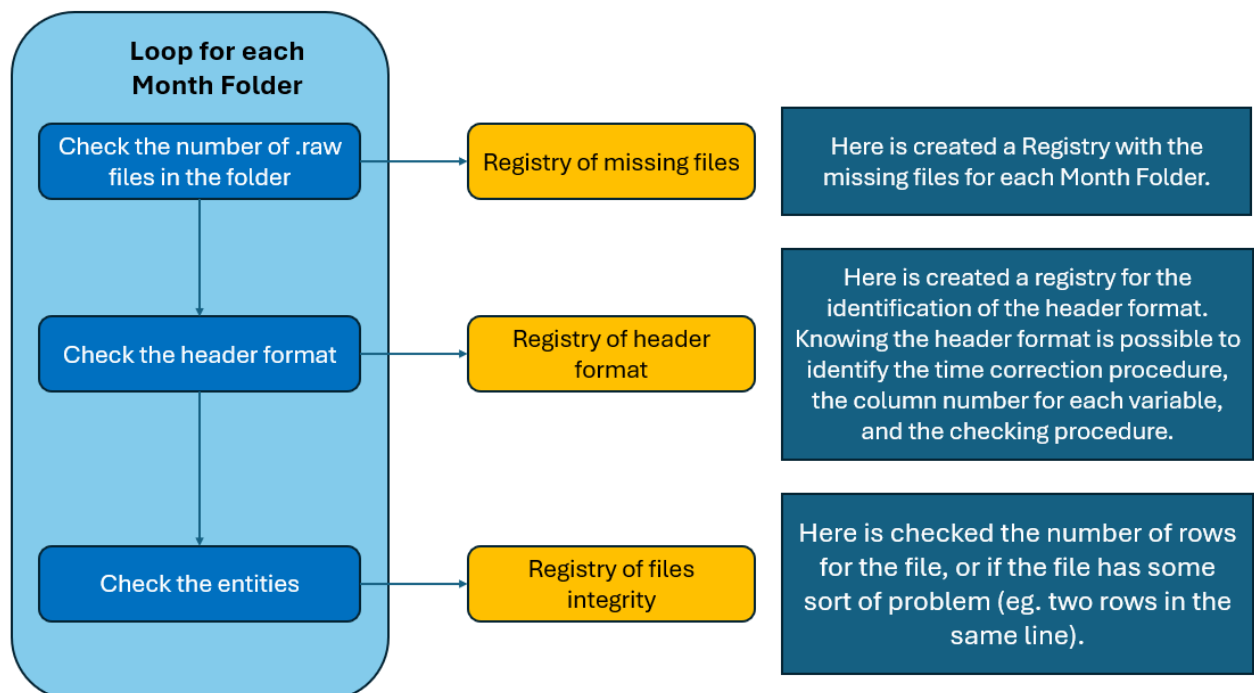


Figure 15: Loop for the pre-analysis of the month folder

At this stage it is possible to decide at which time step the data should be aggregated. This part is very well known in literature as “Data Aggregation” [18], turns out to be a particularly important step because the choice of parameters by which the data are aggregated turns out to be significant for the conduct and significance of subsequent statistics with respect to the data [19, 20]. Data aggregation if conducted as an automated procedure, without assessing its impacts on data quality,

can lead to major losses on the significance of the information produced (signal flattening), and should therefore be calibrated to individual study needs. For the majority of these analyses, a useful aggregation format for monitoring fire emissions is hourly mean aggregation.

Starting with data sampled every 5 seconds by the Picarro and every minute by the MAAP and Weather Station, hourly data can be derived for each of the variables of interest, taking into account their aggregated means.

The analysis process in this case is composed of 4 steps plus an optional one:

- An empty file is created containing an indication of how many rows must be present according to the chosen aggregation step (e.g. in a file for a day with 1 minute time steps, the number of data rows must be 1440, or in case of an hourly step, every file will have 24 lines). To each row an hourly value considering the beginning of the year, is assigned. For example, 2 January will have the number 25 as its first hour is actually the 25th hour of the year. In a non-leap year, the last hour of 31 December will have the value 8760 (365 days * 24 hours) as an indicative number. In the case of the LMT observatory, it was chosen to give the first hour of 1 January 2015 the value of zero hour and to continue with an hourly increasing ordinal number from year to year, so that the first hour of 1 January 2016 can be represented by the number 8761, the last one of 2016d one by 17544 (2016 is a leap year), and so on. This permits to have unambiguous recognition for each hour of the observatory looking at all its variables in the same time;
- In the second step, the integrity of the starting data is verified; if there is data useful to create the hourly mean value, the data will be written into the file, if the data are not present the row will remain empty;
- In the third step, the presence of calibration data is checked; if the instrument reports a calibration phase, the data are saved in an additional file and are not written in the hourly mean data file, leaving empty the corresponding lines;
- In the fourth step, the necessity of realigning the data to UTC+ 0 or UTC+1 format is evaluated; if needed the appropriate correction is applied;
- The optional fifth step is for the calibration; this phase depends on the single equipment because it is not present in all of them; if an internal calibration is requested, the process is done and the calibrations are applied to the hourly mean data;

Within this reconstruction process, the standard deviation values for each hour are also reconstructed along with the mean data, and the number of units used to calculate the mean and standard deviation for that particular hour is also stored. The output file is stored in a folder containing all the processed files for each instrument in each year. A diagram of the system is shown in the following figure.

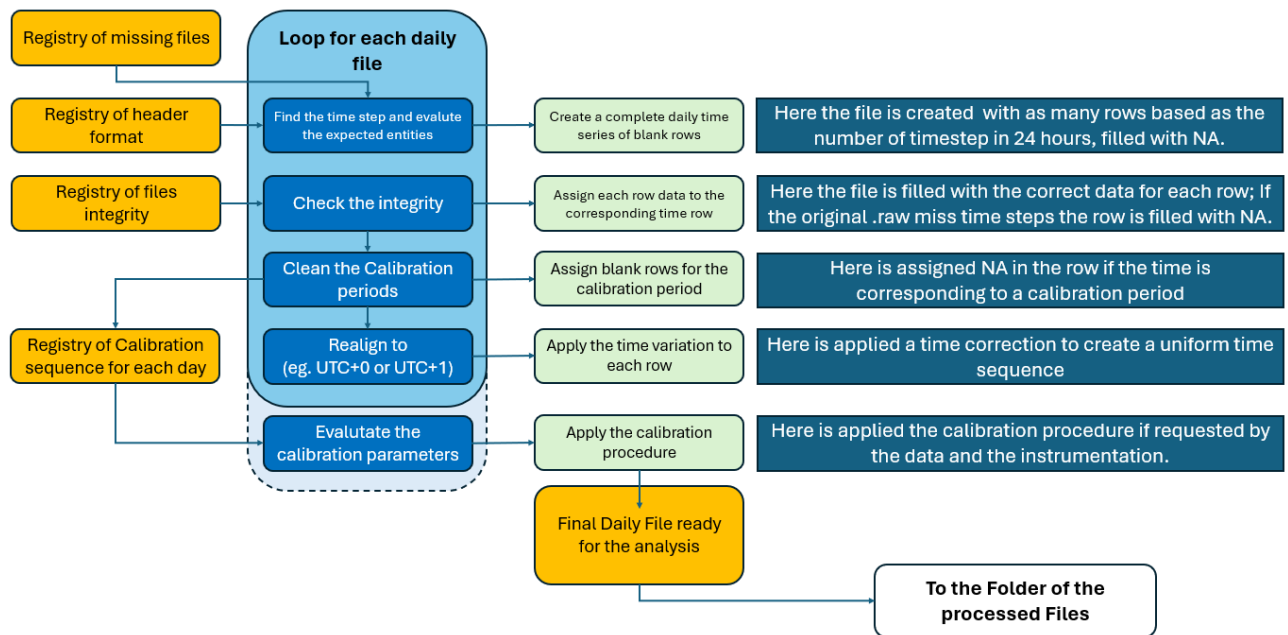


Figure 16: Loop for the analysis and aggregation of the daily file and aggregation in a new time format

At this point, three folders were created containing all the data, corrected for errors and aggregated into hourly averages for each of the three instruments previously mentioned. In this way it is possible to create a unique database for the entire recording period of the observatory. In the case of the LMT observatory, the enumeration carried out during the first step of the daily loop can be used to easily realign the hourly averages with respect to the observatory's hourly data. A representation of the resulting database is shown in the following figure:

LMT h num	date	time	Vaisala ent num	wind speed	wind speed std	wind direction	wind direction std	Picarro ent num	CO val	CO std	MAAP ent num	BC mass val	BC mass std
1	01/01/2015	00:00:00	60	1,59	0,73	106,35	28,49	720	343,79	23,93	60	1,46	0,19
2	01/01/2015	01:00:00	60	3,56	1,35	72,5	18,28	720	291,10	28,64	60	1,09	0,21
3	01/01/2015	02:00:00	60	4,42	1,03	64,08	8,04	720	295,63	45,25	60	0,44	0,24
4	01/01/2015	03:00:00	60	3,87	1,16	62,18	45,4	720	223,18	13,59	60	0,27	0,08
5	01/01/2015	04:00:00	60	3,15	1,45	70,37	126,26	720	209,59	8,88	60	0,29	0,09
6	01/01/2015	05:00:00	60	1,53	0,72	143,77	157,51	720	204,65	18,92	60	0,30	0,09
7	01/01/2015	06:00:00	60	2,12	0,78	213,7	146,42	720	190,05	19,94	60	0,26	0,10
8	01/01/2015	07:00:00	60	2,69	0,96	260,25	138,68	720	192,26	23,99	60	0,33	0,11
9	01/01/2015	08:00:00	60	2,63	0,74	311,92	82,1	720	199,04	28,39	60	0,40	0,11
10	01/01/2015	09:00:00	60	3,01	0,8	277,62	134,62	720	190,73	24,96	60	0,38	0,11
11	01/01/2015	10:00:00	60	4,09	1,18	33,97	74,94	720	204,18	27,61	60	0,39	0,11
12	01/01/2015	11:00:00	60	4,8	0,92	70,7	119,91	720	232,13	10,44	54	0,45	0,13
13	01/01/2015	12:00:00	60	4,69	0,88	38,43	74,19	720	213,61	7,98	60	0,34	0,03
14	01/01/2015	13:00:00	60	4,51	0,98	41,55	85,34	720	218,71	4,90	60	0,38	0,02
15	01/01/2015	14:00:00	60	3,94	0,96	94,18	150,71	720	218,73	8,49	60	0,38	0,04
16	01/01/2015	15:00:00	60	4,16	1,05	48,68	13,82	720	223,89	14,26	60	0,47	0,03
17	01/01/2015	16:00:00	60	2,47	0,5	65,3	11,48	720	231,28	5,38	60	0,47	0,05
18	01/01/2015	17:00:00	60	2,96	0,55	74,4	7,92	720	252,11	12,58	60	0,77	0,13

Figure 17: Example of the Data-Base for LMT-site with mean hour aggregation;

The database created in this way is suitable for the detection of anomalies, peaks, and overlapping effects in more than one of the tracers during the same time interval.

3.2 Analysis instruments of background data for the historical series

The evaluation of the background concentrations of an observation site is fundamental for the identification of all available measurements at a location, this makes possible the correct identification of events that occurred as anomalies with respect to concentration levels usually measured near the observation point [21].

Background data are the representative concentrations of the atmosphere unaffected by local conditions such as anthropogenic sources near the observation point. However, the definition of the background is not straightforward; each site will be affected by a different amount of external sources depending of their proximity to the measurement point and the weather conditions around it [5]. Once the background of the polluters under exam has been identified, it is possible to analyse in a more detailed way all those things that contribute to the increase of concentrations. It will be possible to identify the contribution of pollutant sources near the monitoring sites, to comprehend how meteorological phenomena influence the transport and the dynamics of evolution of pollutant concentrations around the area of interest [22, 23], or to retrieve the relative growth trends.

For this purpose, it was decided to develop after the reconstruction of the single database a system to identify background data from the “event” ones, also to understand the effective rate in data variation throughout the historical series.

To develop a robust yet versatile approach, it was decided to test the general algorithmics on more than one site, as there was a risk of running into forms of calibration that were too specific to the Lamezia Terme site. Therefore, it was decided to invest some of the time to study some of the databases of information on data and atmospheric emissions, such as the ICOS CO₂ monitoring circuit [24], of which several sites participating in WP4.16 are members.

which the LMT site is a part of along with other WP4.16 adherents.

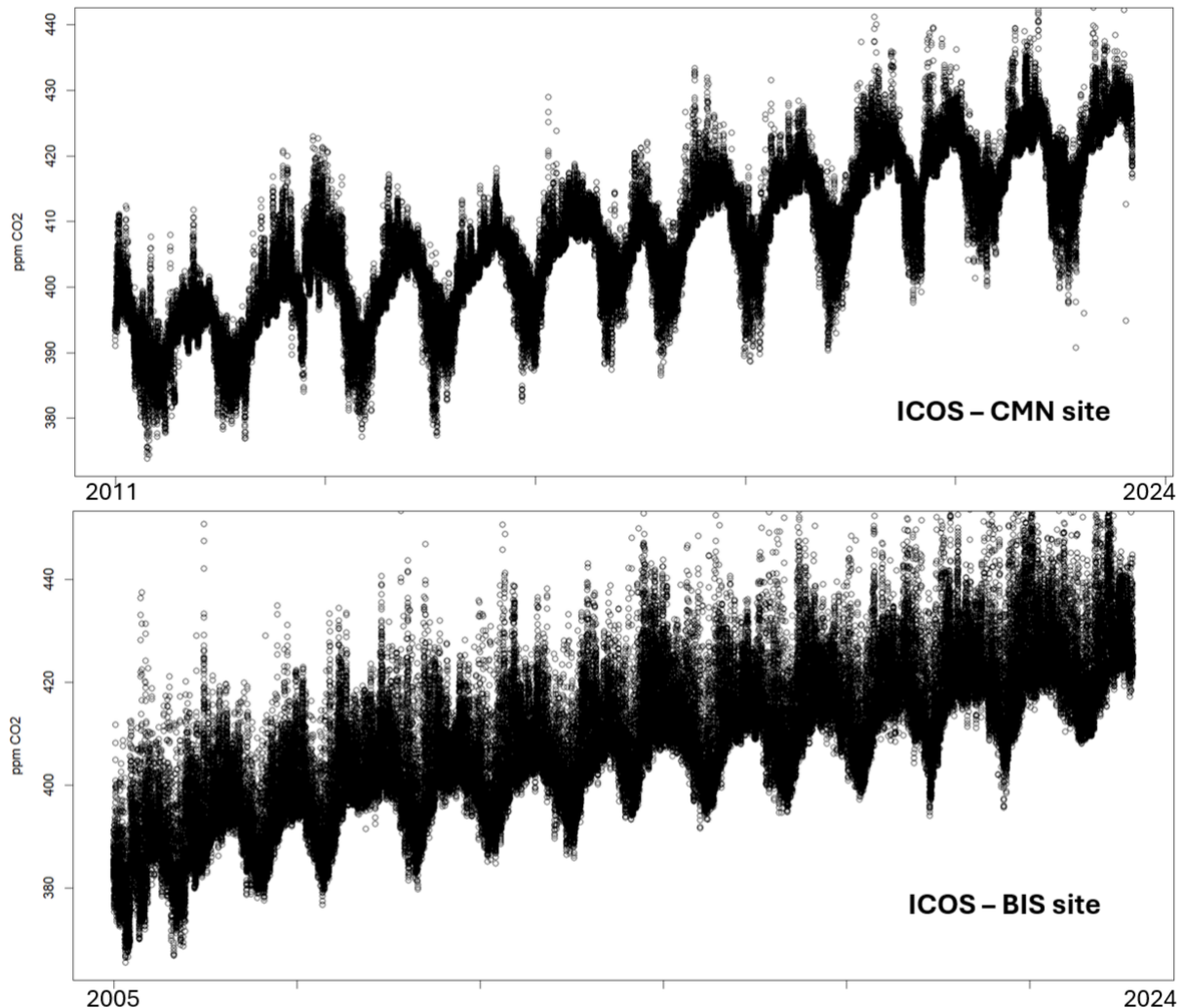


Figure 18: CO₂ measurements in CMN site and in BIS site. The CMN is a very clean site, and the BIS site instead has seen a lot of disturbances.

In the figure 4-6 above the CO₂ measurements observed at a global site, such as Monte Cimone CMN (ITALY) are reported. These are usually considered a background site with little anthropic influences, and Biscarosse BIS (France) a highly anthropic influenced site. This strong differentiation in data trends is very pronounced for CO₂ parameters but also for other parameters, such as particulate concentration values [25, 26].

This operation of background data analysis is necessary because the data trace recorded is unique for each one of the various sites. To better understand the differences, and individuate more efficiently the event within the dataset, it is necessary to create an analysis for each site. We will briefly describe some of the procedures used in the data analysis in relation to the background identification.

3.2.1 Smoothed Minima Approach

The approach described below was initially designed to be applied to the analysed time series of water flows in order to separate base levels from daily rapid flows [27].

The smoothed minima technique is applied in three different steps:

- a) The entire measurement series is divided into non-overlapping intervals of 5 days;

- b) The minimum daily streamflow value Q_i is identified for each window;
- c) A check is carried out between all the previously found minima based on the comparison of each daily value Q_i , the previous Q_{i-1} and the subsequent one Q_{i+1} to identify the turning points (TP);

these “Turning Points” (TP) will be all those daily values Q_i that fulfill the condition:

$$0.9 * Q_i \leq \min(Q_{i-1}, Q_{i+1})$$

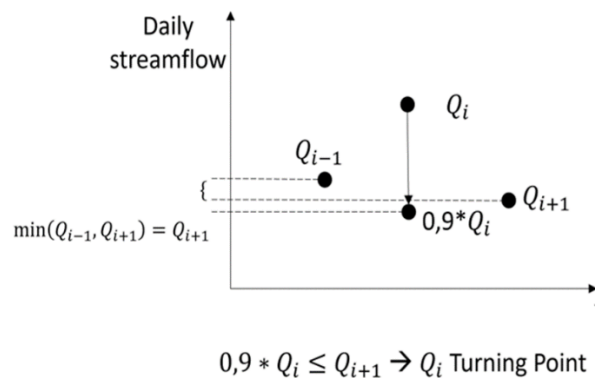
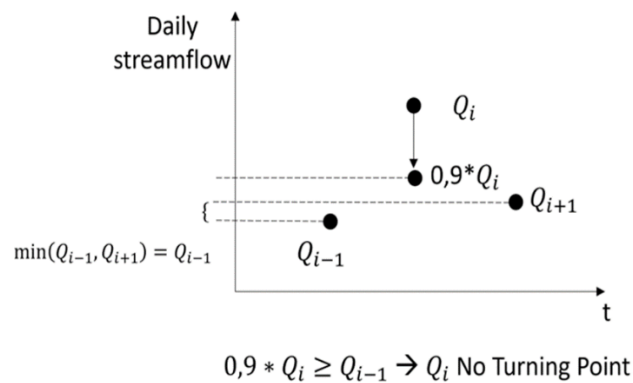


Figure 19: Example of the comparison criterion for the smoothed minima approach.

a subsequent linear interpolation allows to obtain a baseline through which it is possible to identify the trend of the water flow analysed net of events that occurred during the analysis period.

This method aims at the determination of daily flows of data that could be labelled as background levels. It does not use a specific characterisation of the parameters attributed to each site on which it is applied and not even provide an upper limit of concentrations. The inability to obtain an upper limit from this technique determined the need to add a second stage of analysis for the identification of an upper trend, within which the background level of the pollutants under study is defined. In this way, it is possible to characterise the analysed site by distinguishing between background concentrations and the evolution of concentrations due to anthropic activities.

VARIATIONS IN TECHNIQUE

Datasets produced by LMT observatory give information on the concentrations of the pollutants (CO₂, CH₄, CO) with a measurement periodicity of 5 seconds, these were aggregated on an hourly basis to be adherent with the databases available for the other sites. The first difference with the initial algorithm is therefore the use of data on an hourly basis, with which it will be less frequent to include concentrations due to external phenomena that we want to isolate, compared to the daily analyses.

The identification of the upper limit for the definition of the background is carried out with a criterion similar to the previous one, but related to the maximum concentration of each observation window: once the upper extremes of each interval have been defined, triads of adjacent values are considered and all the points for which the condition is valid are marked as Turning Points related to the maxima::

$$k * Q_i \geq \max(Q_{i-1}, Q_{i+1})$$

with k increasing factor of the central point of the triad.

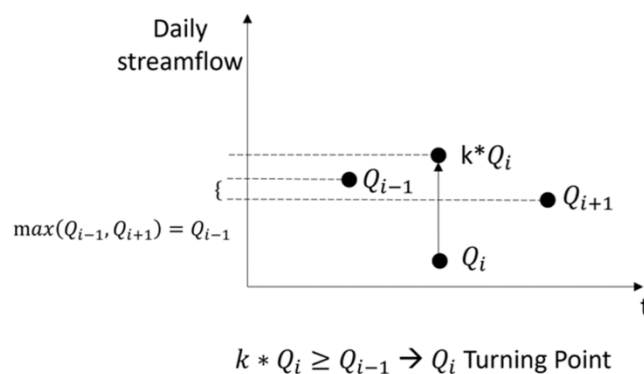
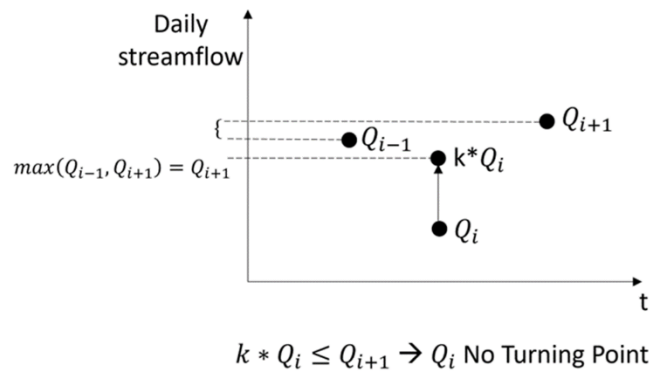


Figure 20: Example of variation in technique for the maximum values.

However, the drawback of the above analysis is the consideration of non-significant data as possible turning points (both maximum and minimum), including possible instrument errors, post-processing errors or measurements with high standard deviations. The latter are representative of rapid variations and thus contradict the background concept. In order to avoid the inclusion of useless data in the main analysis, it was necessary to implement prior dataset clean-up strategies:

- the hourly means computed with a low number of measurements are deleted from the set;
- the hourly means characterised by values of standard deviation higher than a threshold defined on a percentual basis are deleted from the set;
- the hourly means that, based on a percentile analysis, can be identified as “outliers” and that represent atypical situations for the site are deleted from the set.

In the following images the comparison between the pre-analysis and the applied technique results: is shown

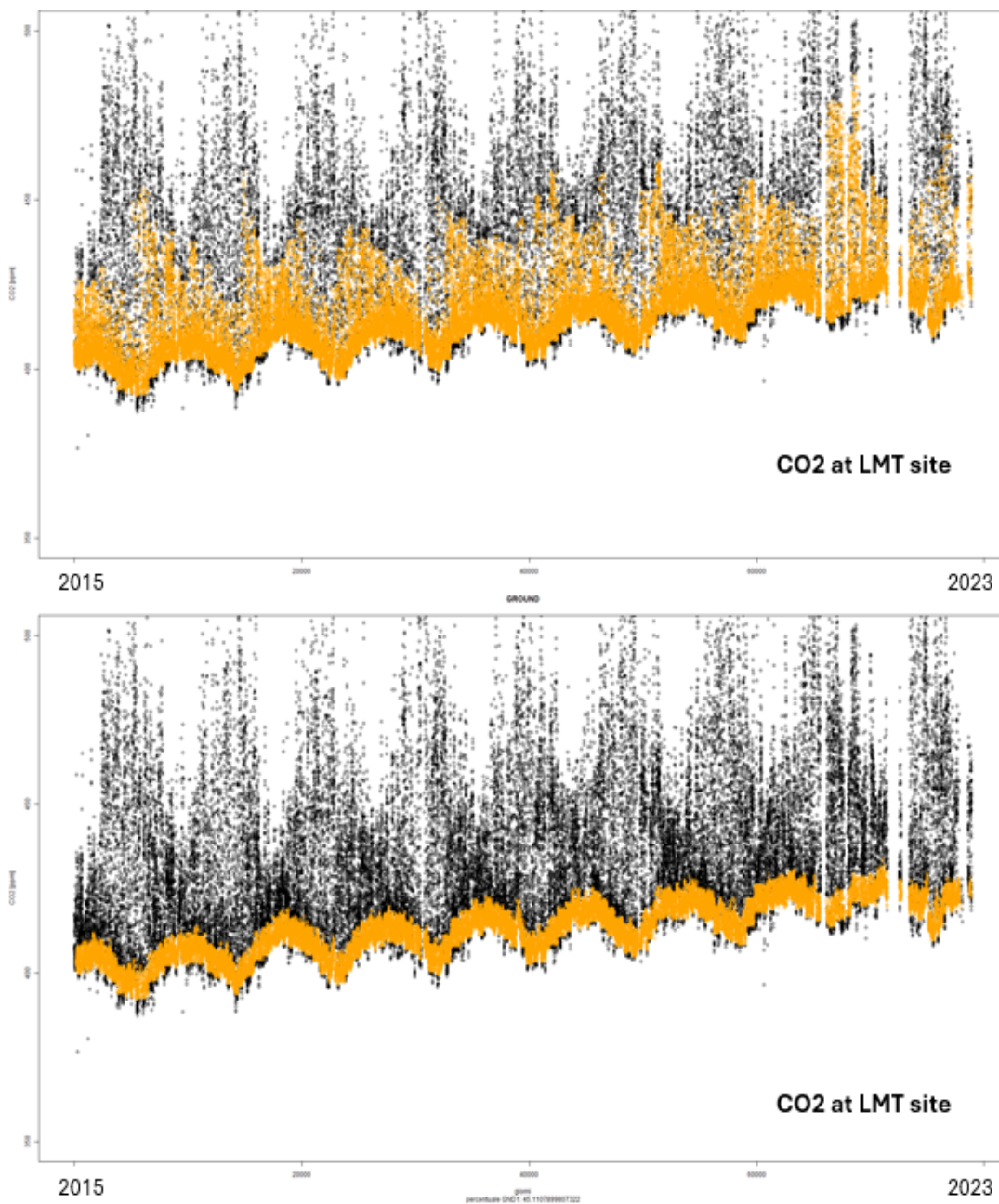


Figure 21: in the upper figure the pre-analysis on LMT-CO2 dataset, in the bottom figure the selection of the completed procedure.

3.2.2 Background Data Selection – BaDS

The BaDS methodology (BaDs-Background Data Selection method) is a statistical method based on the assumption that, for the background conditions of a species, very little variability is allowed, both within hourly means (i.e. with a small sub-hourly standard deviation) and between pairs of consecutive means (i.e. slow variation over consecutive hours) [28, 29]. It consisted essentially of two steps to control: the variability of the hourly data and the difference between consecutive hourly means.

The choice of the threshold parameters was established on the sensitivity test starting from the complete dataset of the three gases analysed, so that only background measurements were retained and, at the same time, only a small number of hourly values were removed [1].

The BaDS is applied to the datasets in the following steps [28-29]:

- This methodology examined at first the standard deviation assigned to each hourly average and the datum was flagged if the value exceeded a given threshold cut-off value σ i.e., for CO 10 ppb, for CO₂ 3 ppm and for CH₄ 16 ppb.
- Secondly, each datum, hourly mean value, was compared with the previous one, and the datum was flagged if the difference exceeded a given threshold $\delta = 1.5$.
- Thirdly, each hourly measurement that passed the previous steps is compared with the corresponding moving median value. To do so is evaluated a “504 hours” time window (21 days * 24 hours), 252 before and 252 after each hourly measurement. If the time window is composed of at least of 126 valid data (so in the previous steps survived at least the 25 % of data population) the moving median is computed. If the difference between the measurement and the computed median does not exceed the threshold $\rho = 7 * \delta = 10.5$, the datum is considered as background;
- Fourthly, similarly to the previous step, each hourly measurement that passed the previous steps is this time compared with the corresponding moving mean value. This moving mean for each hour is again computed on a window of 504 hours, but this time only if in the window at least the 10% of data passed the previous steps. If the difference between the measurement and the computed mean exceeds the threshold $\rho = 7 * \delta = 10.5$, the datum was considered as a non-background datum.
- Finally, a readmitting procedure was made comparing all the discarded hourly mean values that did not pass the previous steps with the respective moving mean. If the difference was lower than the ρ threshold, the hourly datum was reintegrated and considered part of the background set.

The results of this type of analysis are satisfactory, the following images are given as example:

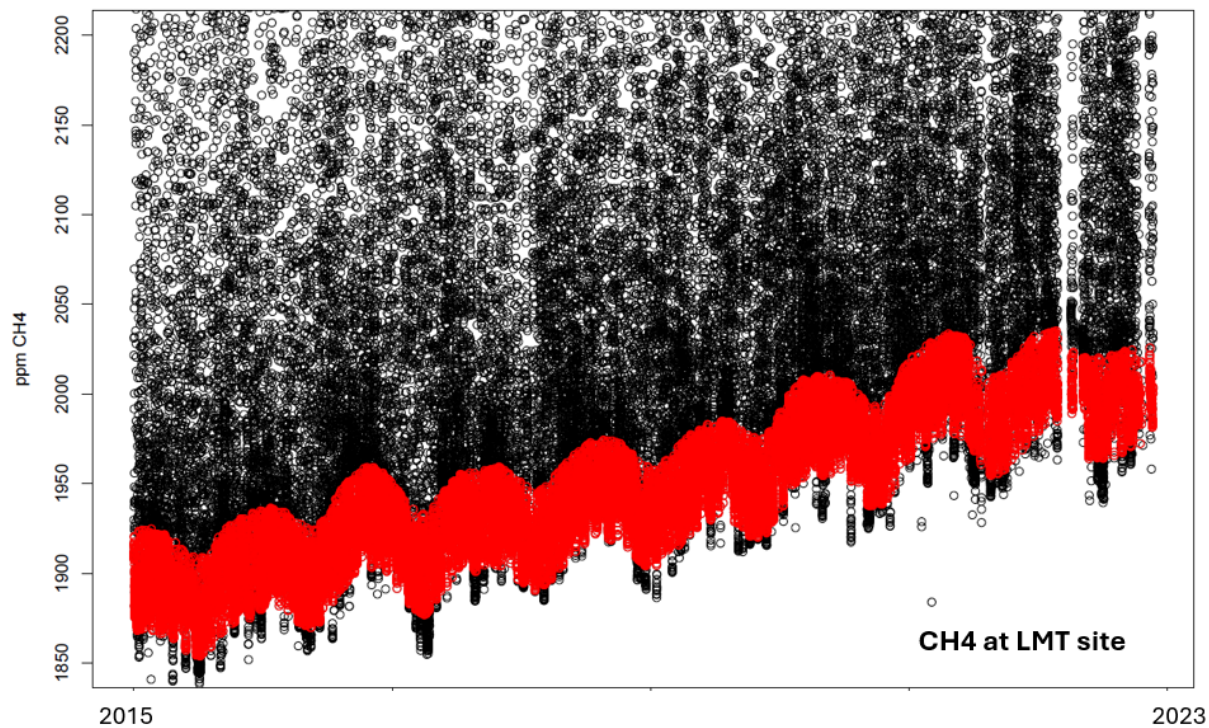


Figure 22: CH4 Background Selection made by BaDS at LMT Site.

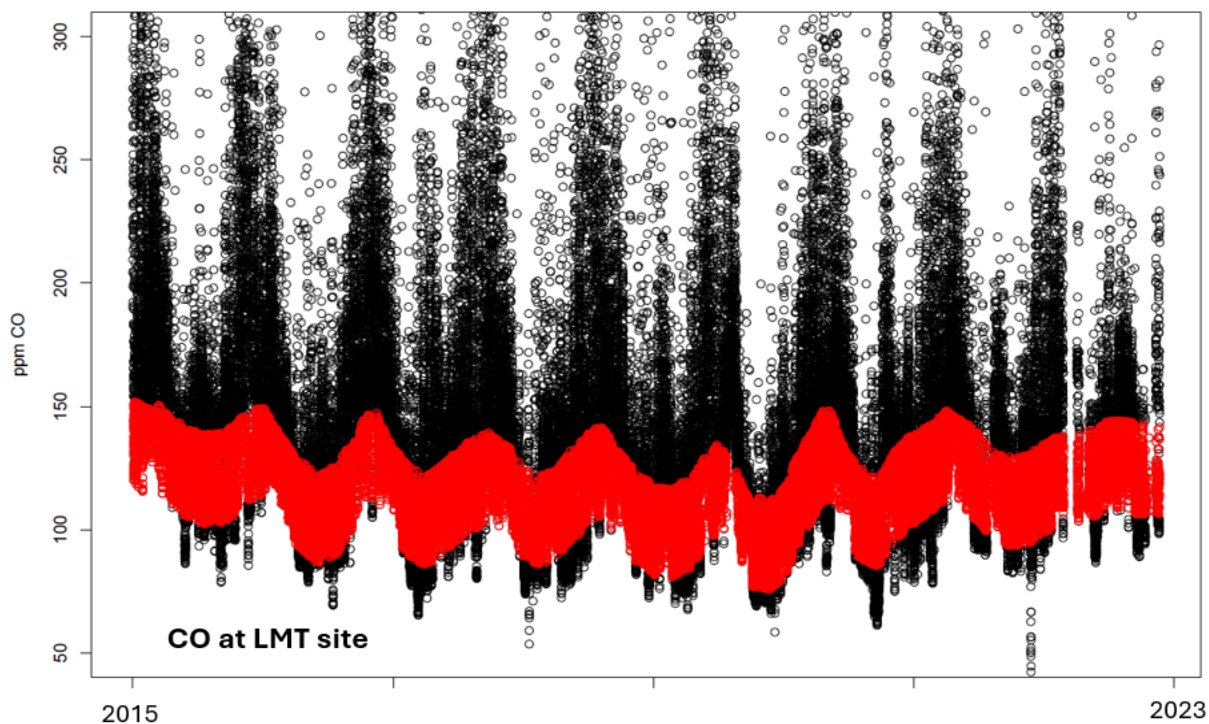


Figure 23: CO Background Selection made by BaDS at LMT Site.

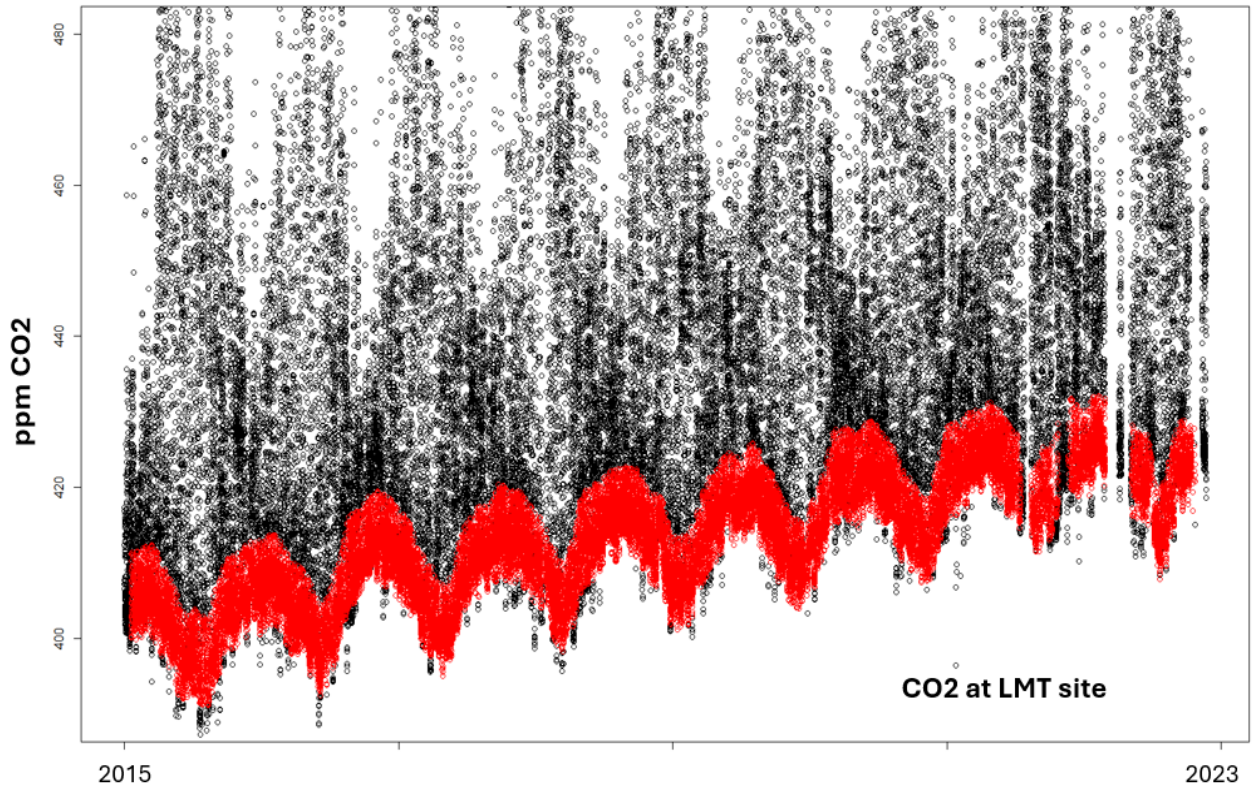


Figure 24: CO2 Background Selection made by BaDS at LMT Site.

3.2.3 Wind-Driven Background Selection

In this case, a selection method was used that took into account the analysis of meteorological parameters by evaluating the direction of air masses towards the site and the wind intensity. The meteorological data were obtained as hourly averages from an automatic weather station. In this second criterion, the geographical location and characteristics of the observation point in relation to the coastline were used. The experimental site was located about 600 metres from the Tyrrhenian Sea, with a gas sampling point at 4 metres above the ground.

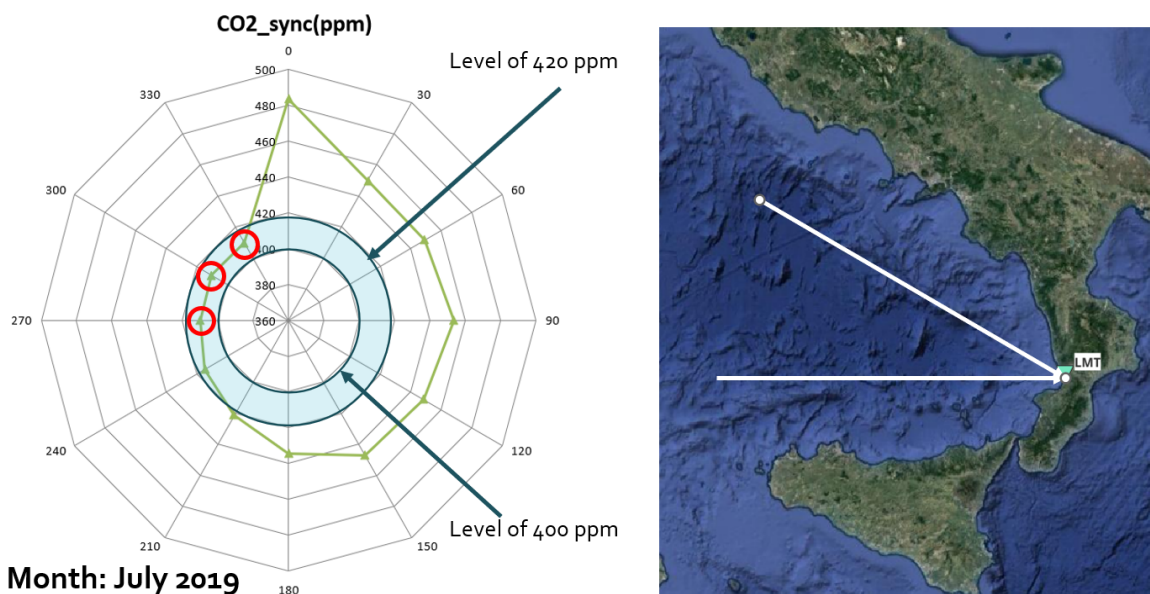


Figure 25: correlation of wind direction and CO2 levels in July 2019, the air coming from seaside is much more clean than the part coming from land;

The observation point is in the Tyrrhenian Gulf of Lamezia Terme, and the wind have no orographic obstacles up to the southern coast of Sardinia at 600 km. Considering this last information, it can be plausible to assert that the air masses coming from the West-Northwest (W-NW) sectors were less affected by anthropic factors respect to other directions. Consequently, we adapted for this method the following factors.

In the first selection phase, hourly mean values of wind which having the following characteristics were evaluated:

- wind direction coming from W-NW, i.e., 240° to 330°;
- wind speed greater than or equal to 2 m/s, i.e. 7.2 km/h.

In the second phase, sequences of at least four consecutive hours were selected only for the data selected in the previous points. This condition allowed us to identify groups of at least four hours in which the average wind direction and wind speed were persistent and coming from the offshore direction, and therefore potentially clean. For these groups, only the hourly mean values from the third hour onwards were selected, leaving out the first two hours. The first hours were discarded because, in a location like LMT subject to daily reversal of the wind direction, a mass of "polluted" air can go towards the sea to be after brought back, creating of alterations.

In the third phase, the hours identified in the previous point were used to identify the corresponding hourly mean values concentrations of CH₄, CO and CO₂.

In this last phase, a further filter was applied, which was aimed at eliminating the concentration data of the gases which, despite respecting the wind field criteria, exceed the global mean ground concentration value by 1.25 (CH₄ 1922 ppb, CO₂ 419 ppm and CO 130) [30].

In the following images, the result of both the wind analysis method and BaDS are shown:

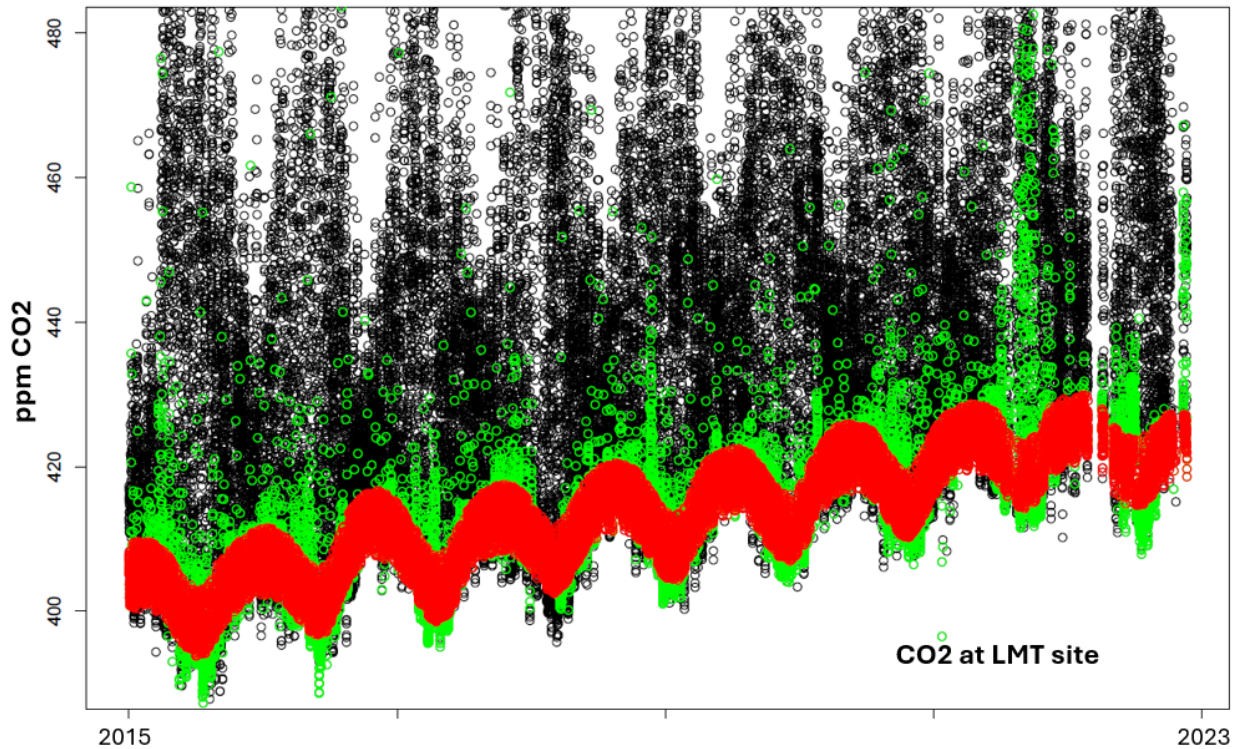


Figure 26: Comparison between the wind selection (green) and the BaDS selection (red) for CO₂ at LMT site.

The differences between the two methods are obvious to the eye. The wind-driven selection offers a dirtier data, and not without values that differ significantly from the lower part of the dataset.

This is despite the many considerations made to try to mitigate the effects due to recirculation, small and large local events, and influences also due to perhaps unidentified activities.

Relying solely on personal experience of the site, relying on local considerations, does not necessarily lead to directly appreciable results.

This is true even in a state like LMT, where the direction of "clean air" is easy to identify.

In the figures, the selection in red is that of the BaDS, and all data inconsistencies have been eliminated, so that the selection is very compact, with a narrow range of vertical variation and with a strong continuity in seasonal and multiannual cyclicity.

This analysis has made it clear that for these masses of data, with great spatial and temporal variability, it is appropriate to follow a path of "blind" statistical investigation and to use analyses based on experience only as elements of performance comparison.

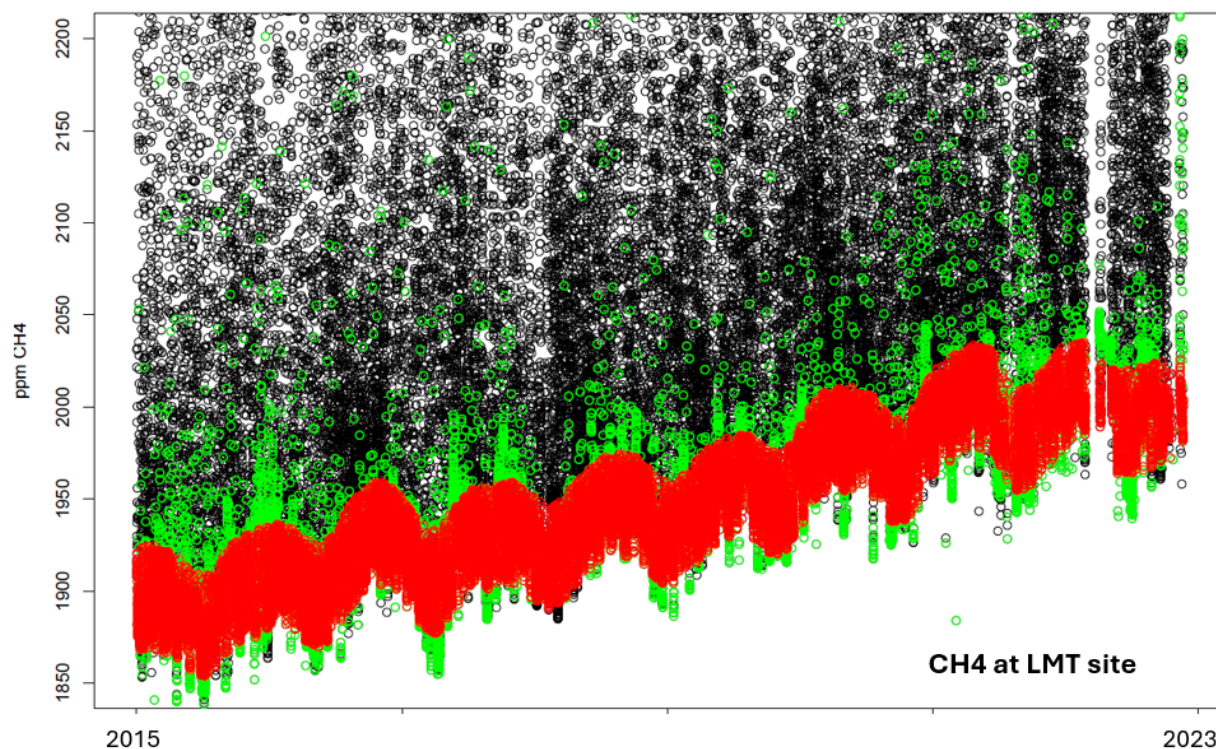


Figure 27: Comparison between the wind selection (green) and the Bads selection (red) for CH4 at LMT site.

4.3 Spatial Data

For the integration of spatial data, it is not necessary to use some particular elaborations, because the general approaches have become particularly robust over the years [31, 32]. Data in raster format are generally directly usable because the system providing them (eg. Sentinel products) has already validated them.

The only type of file that needs special elaborations are those files composing the regional fire register. These files are in vector format, and to be easily used they need a processing phase that transform them into “quadrant” maps. The approach used in this case consisted in moving the data of potential emissions to a system that can aggregate these data into tiles, since the emissive sources are close to a thousand of unity each year (in some they reach 1200 events). It was therefore decided to apply a 5x5 km step grid and reassign to each event falling in a cell the spatial coordinates of the cell’s centroid. In this way, it is also possible to aggregate events occurring simultaneously at short distances, which would not be distinguishable from the observatory data. In this way, the hourly fire register can be easily used for fire localisation.



Figure 28: Transformation of the vectorial data in a gridded data. In a) the complete grid system, in b) the shapes of the fire areas near LMT site.

4. RESULTS AND FUTURE WORKS

The aggregation of all the data made it possible to establish a path of analysis between them, which was useful for identifying correlations between the data and thus developing knowledge of the events through the information obtained.

The scheme of the method is shown below:

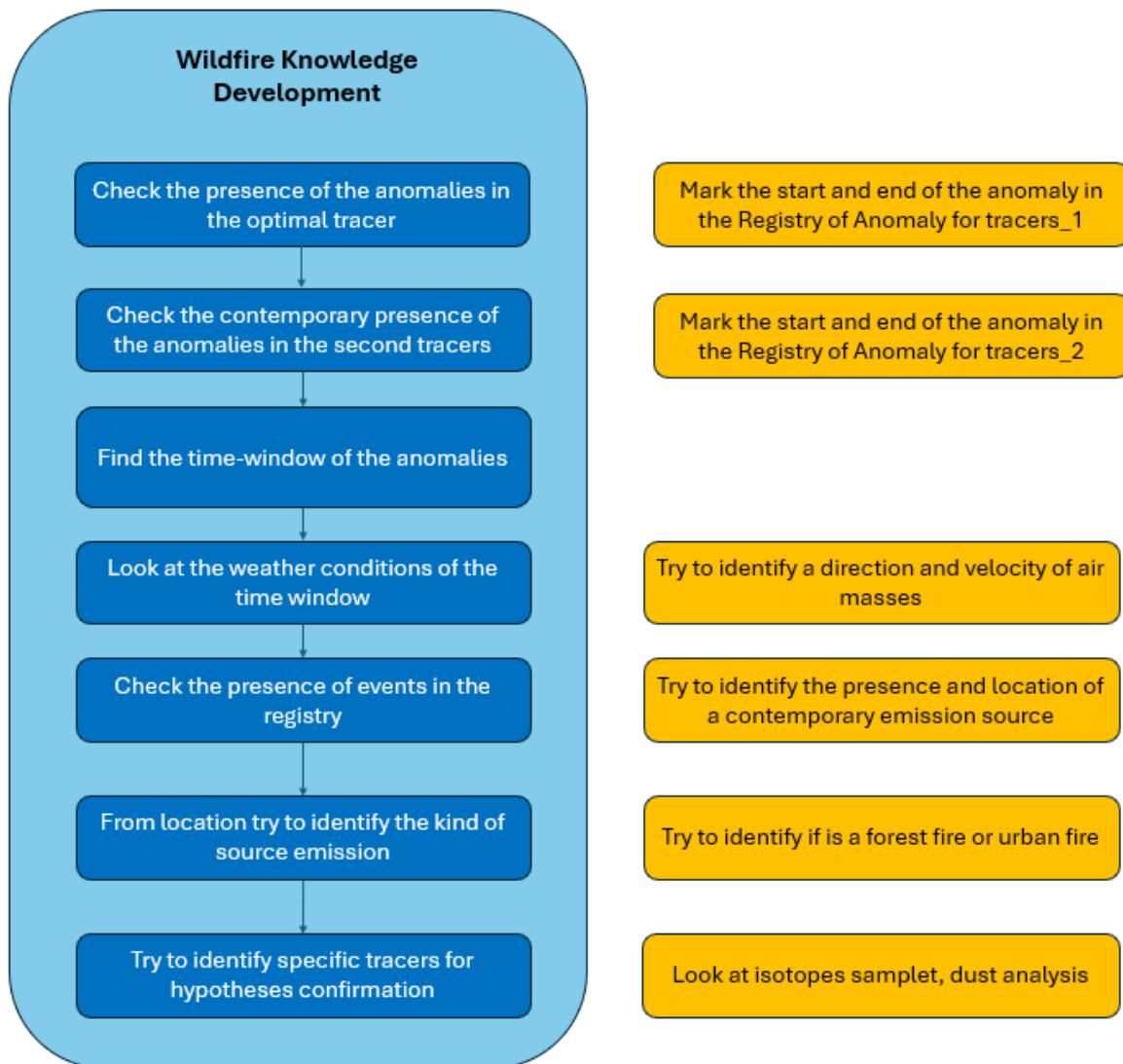


Figure 29: Loop of analysis to mark time, direction, location and type of an emission responsible of instrumental anomalies.

The procedure was designed as follow:

1. Check for the presence of anomalies in the optimal tracers (i.e. CO and eBC);
2. The simultaneous presence of anomalies in the secondary tracers is checked, thus all the other variables (NO_x, O₃, etc.);
3. The event window is identified;

4. Wind conditions during the chosen period are checked, trying to understand the source;
5. Simultaneous and slightly preceding events are identified with respect to the event's window;
6. From steps 4 and 5, an attempt is made to identify the spatial placement of the emissive source;
7. From the spatial placement, an attempt is made to find the possible emission source, whether anthropogenic or natural;
8. From the spatial placement and type of source, an attempt is made to choose which additional pathway needs to be studied (eg. filters, isotopes, etc.) to confirm or deny the source;
9. If the tracing is not considered valid, other emission sources are evaluated, the cycle is started again.

This type of procedure makes it possible to study the data in order to determine the time, direction of propagation, location and type of emitting source that caused the recording of some type of anomaly within the data sequence recorded by the instruments. All this is done by integrating data from different types of observations: in situ, satellite, administrative records, gas, aerosol, meteorological and output model data. Some of the results of these procedures have already been published [1]. The scheme proposed in this scientific study was developed as follow in Figure 5.2

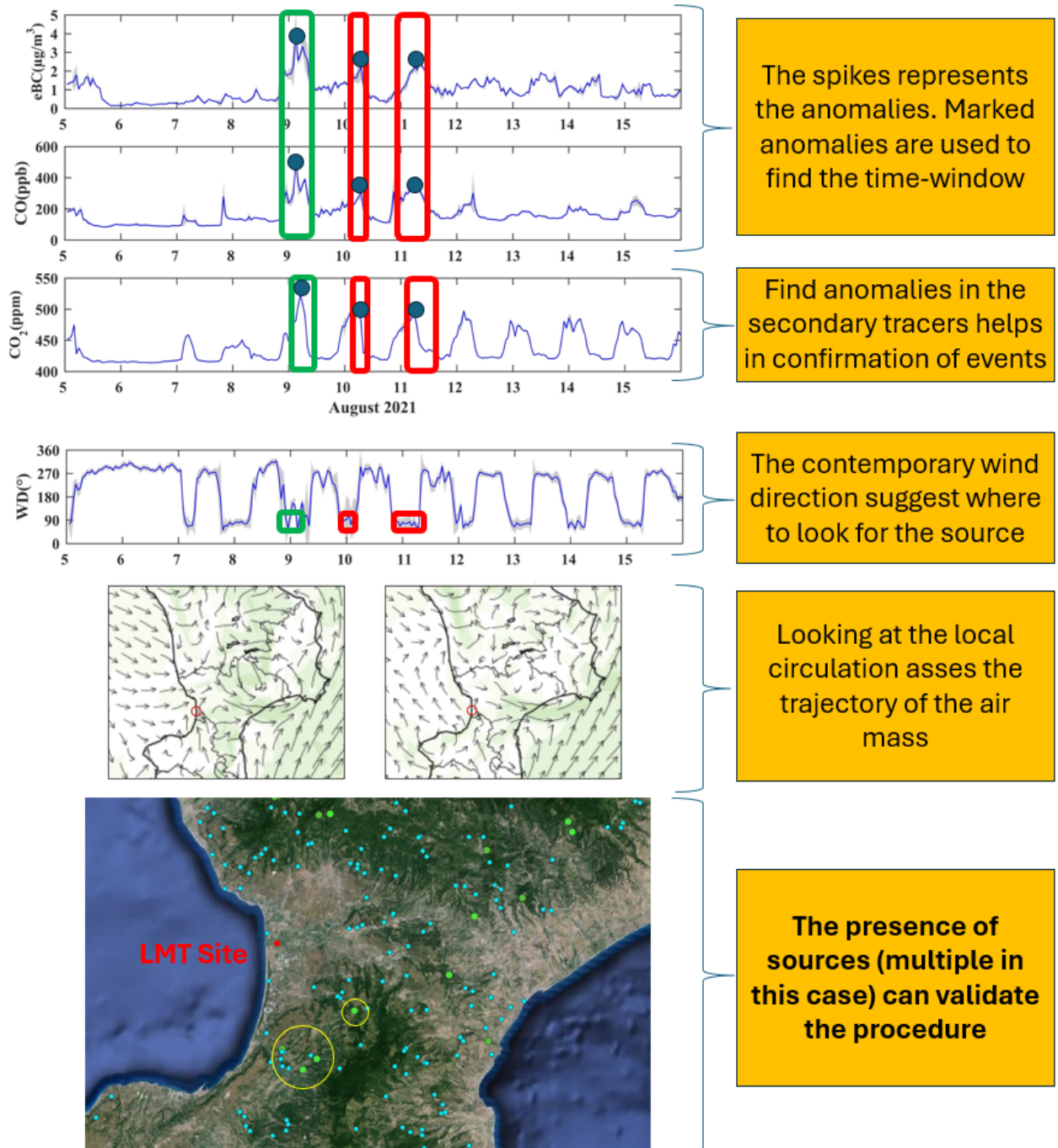


Figure 30: Example of application of the method to track the open fires.

Finally, the new equipment being acquired, such as the C13 isotope analyser in CO₂, will make it possible to distinguish between fossil and non-fossil origins, and therefore between forest fires and other causes.

Post-analysis of the presence of levoglucosan or the EC/OC fraction of carbon in the collected filters using impactors, at the daily sampling stage, will also allow the detection of other tracers confirming the presence of biomass burning, useful for identifying forest fire occurrences and assessing the amount of burnt particles transported.

Aerosol lidar profiles will allow the detection of possible aerosol presence as a stratification of combustion products at higher altitudes compared to in situ measurements performed on the ground.

The intensive campaign planned for the summer of 2024 will allow more data to be collected from different instruments and will be a new source of data to integrate and understand their optimal use for improving tracer identification methods.

BIBLIOGRAPHY

- 1) Malacaria, L.; Parise, D.; Lo Feudo, T.; Avolio, E.; Ammoscato, I.; Gulli, D.; Sinopoli, S.; Cristofanelli, P.; De Pino, M.; D'Amico, F.; et al. Multiparameter Detection of Summer Open Fire Emissions: The Case Study of GAW Regional Observatory of Lamezia Terme (Southern Italy). *Fire* 2024, 7, 198.
- 2) Calidonna, C. R., Avolio, E., Gulli, D., Ammoscato, I., De Pino, M., Donato, A., & Lo Feudo, T. (2020). Five years of dust episodes at the Southern Italy GAW Regional Coastal Mediterranean Observatory: Multisensors and modeling analysis. *Atmosphere*, 11(5), 456.
- 3) Lo Feudo, T., Calidonna, C. R., Avolio, E., & Sempreviva, A. M. (2020). Study of the vertical structure of the coastal boundary layer integrating surface measurements and ground-based remote sensing. *Sensors*, 20(22), 6516.
- 4) Ilyas, I. F., & Chu, X. (2019). *Data cleaning*. Morgan & Claypool.
- 5) Schabenberger, O., & Gotway, C. A. (2017). *Statistical methods for spatial data analysis*. Chapman and Hall/CRC.
- 6) Ruckstuhl, A. F.; Henne, S.; Reimann, S.; Steinbacher, M.; Vollmer, M. K.; O'Doherty, S.; Buchmann, B.; Hueglin, C. Robust extraction of baseline signal of atmospheric trace species using local regression. *Atmos. Meas. Technol.* 2012, 5, 2613–2624. <https://doi.org/10.5194/amt-5-2613-2012>
- 7) Avolio, E., & Miglietta, M. M. (2023). A Comparative Analysis of Two Mediterranean Tornado Hotspots. *Atmosphere*, 14(1), 189.
- 8) Doan, A., Halevy, A., & Ives, Z. (2012). *Principles of data integration*. Elsevier.
- 9) Calvanese, D., De Giacomo, G., Lenzerini, M., Nardi, D., & Rosati, R. (2001). Data integration in data warehousing. *International Journal of Cooperative Information Systems*, 10(03), 237-271.
- 10) Zipkin, E. F., Zylstra, E. R., Wright, A. D., Saunders, S. P., Finley, A. O., Dietze, M. C., & Tingley, M. W. (2021). Addressing data integration challenges to link ecological processes across scales. *Frontiers in Ecology and the Environment*, 19(1), 30-38.
- 11) Doan, A., Halevy, A., & Ives, Z. (2012). *Principles of data integration*. Elsevier.
- 12) Ruffault, J., Curt, T., Moron, V., Trigo, R. M., Mouillot, F., Koutsias, N., ... & Belhadj-Khedher, C. (2020). Increased likelihood of heat-induced large wildfires in the Mediterranean Basin. *Scientific reports*, 10(1), 13790.
- 13) Turco, M., Rosa-Cánovas, J. J., Bedia, J., Jerez, S., Montávez, J. P., Llasat, M. C., & Provenzale, A. (2018). Exacerbated fires in Mediterranean Europe due to anthropogenic

- warming projected with non-stationary climate-fire models. *Nature communications*, 9(1), 3821.
- 14) Zheng, B., Ciais, P., Chevallier, F., Chuvieco, E., Chen, Y., & Yang, H. (2021). Increasing forest fire emissions despite the decline in global burned area. *Science advances*, 7(39), eabh2646.
 - 15) Gharatkar, S., Ingle, A., Naik, T., & Save, A. (2017, March). Review preprocessing using data cleaning and stemming technique. In 2017 international conference on innovations in information, embedded and communication systems (iciiecs) (pp. 1-4). IEEE.
 - 16) Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull.*, 23(4), 3-13.
 - 17) Ganti, V., & Sarma, A. D. (2013). *Data cleaning: A practical perspective*. Morgan & Claypool Publishers.
 - 18) Clark, W. A., & Avery, K. L. (1976). The effects of data aggregation in statistical analysis. *Geographical Analysis*, 8(4), 428-438.
 - 19) Zhou, P., Fan, L. W., & Zhou, D. Q. (2010). Data aggregation in constructing composite indicators: A perspective of information loss. *Expert Systems with Applications*, 37(1), 360-365.
 - 20) Gomes-Gonçalves, E., Gzyl, H., & Mayoral Blaya, S. (2016). A maximum entropy approach to the loss data aggregation problem. *Journal of Operational Risk*, 11(1).
 - 21) NOAA Technical memorandum; The statistical treatment of CO2 data records, (NOAA, 1989) Available online: <https://www.arl.noaa.gov/documents/reports/arl-173.pdf>
 - 22) Ruckstuhl, A.F.; Henne, S.; Reimann, S.; Steinbacher, M.; Vollmer, M.K.; O'Doherty, S.; Buchmann, B.; Hueglin, C. Robust extraction of baseline signal of atmospheric trace species using local regression. *Atmos. Meas. Technol.* 2012, 5, 2613–2624. <https://doi.org/10.5194/amt-5-2613-2012>
 - 23) Artuso, F.; Chamard, P.; Piacentino, S.; Sferlazzo, D.M.; De Silvestri, L.; di Sarra, A.; Meloni, D.; Monteleone, F. Influence of transport and trends in atmospheric CO2 at Lampedusa. *Atmos. Environ.* 2009, 43, 3044–3051. <https://doi.org/10.1016/j.atmosenv.2009.03.027>
 - 24) Integrated Carbon Observation System. Available online: <https://www.icos-ri.eu/icos-national-networks>
 - 25) Gao, S., Yang, W., Zhang, H., Sun, Y., Mao, J., Ma, Z., ... & Bai, Z. (2018). Estimating representative background PM2.5 concentration in heavily polluted areas using baseline separation technique and chemical mass balance model. *Atmospheric Environment*, 174, 180-187.
 - 26) Menichini, E., Iacovella, N., Monfredini, F., & Turrio-Baldassarri, L. (2007). Atmospheric pollution by PAHs, PCDD/Fs and PCBs simultaneously collected at a regional background site in central Italy and at an urban site in Rome. *Chemosphere*, 69(3), 422-434.
 - 27) Aksoy, H., Kurt, I., & Eris, E. (2009). Filtered smoothed minima baseflow separation method. *Journal of Hydrology*, 372(1-4), 94-101.
 - 28) Apadula, F., Cassardo, C., Ferrarese, S., Heltai, D., & Lanza, A. (2019). Thirty years of atmospheric CO2 observations at the Plateau Rosa Station, Italy. *Atmosphere*, 10(7), 418.
 - 29) Ferrarese, S., Apadula, F., Bertiglia, F., Cassardo, C., Ferrero, A., Fialdini, L., ... & Vannini, C. (2015). Inspection of high-concentration CO2 events at the Plateau Rosa Alpine station. *Atmospheric Pollution Research*, 6(3), 415-427.
 - 30) <https://gml.noaa.gov/ccgg/trends/>
 - 31) Hasani, S., Sadeghi-Niaraki, A., & Jelokhani-Niaraki, M. (2015). Spatial data integration using ontology-based approach. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40, 293-296.

- 32) Devogele, T., Parent, C., & Spaccapietra, S. (1998). On spatial database integration. *International Journal of Geographical Information Science*, 12(4), 335-352.