This project has received funding from the Interreg Sudoe Programme and the European Regional Development Fund (ERDF) under grant agreement nº SOE3/P3/E0922

# P1.1 Methodology on selection of climate and air pollutant variables and indicators

**https://fecea-viewer.predictia.es**

**Project: New Evaluation Method for Homes of Social, Sustainable and Energy Efficient Interest – Architecture for Climate- in the Sudoe Territory (ARCAS)**

## Technical references

| | |
|---|---|
| **Project Acronym** | ARCAS |
| **Reference** | SOE3/P3/E0922 |
| **Title** | New Evaluation Method for Homes of Social, Sustainable and Energy Efficient Interest – Architecture for Climate- in the Sudoe Territory (ARCAS) |
| **Project coordinator** | Fundacion Estudios Calidad Edificacion Asturias (FECEA) |

| | |
|---|---|
| **Deliverable nº** | D1.1.- Methodology on selection of climate and air pollutant variables and indicators |
| **Dissemination Level** | Internal |
| **Group of Tasks** | GT1 – Selection of indicators: Climate map |
| **GT Leader(s)** | FECEA |
| **Author(s)** | Arturo Gutiérrez de Terán, Emilio Suarez, Manuel Garcia |
| **Approved by coordinator** | 09/02/2023 |
| **Due date of deliverable** | 30/04/2021 |

**Document Version Control**

| Version | Date | Comment | Modified by |
|---|---|---|---|
| v.1 | 02/12/2022 | Initial version in Spanish | Daniel San Martin (Predictia)<br>Markel García (Predictia) |
| v.2 | 18/01/2023 | Initial version in English | FN |
| v.3 | 23/01/2023 | Revisión by ULR | Jérôme Le Dréau |
| v.4 | 07/02/2023 | Corrections | ES, MG |
| v.5 | 09/02/2023 | Final version | ES, MG |

Statement of originality: This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material has been made through appropriate citation, quotation or both. The content of this deliverable reflects the author's views and does not contain any opinion from the management bodies of the Programme Interreg Sudoe.

# INDEX

GLOSSARY

The following acronyms are used within this evaluation report:

API: Application Programming Interface

AEMET: Spanish Meteorological Agency

PNOA: National Air Orthophoto Plan (Plan Nacional de Ortofoto Aérea)

PM: Particulate Matter

RH: Relative Humidity

Td: Dew-Point Temperature

## 1. EXECUTIVE SUMARY

This document describes the index of products included in the ARCAS climate map as well as the main functionalities that are included. The section 2 details the catalogue of products. This catalogue is supported by the description of the data included in the deliverable 1.1 "Climate and pollution indicators report". In addition to the catalogue of products, section 2 includes a description of the methodology used to obtain them. Finally, section 3 includes a description of the main functionalities of the user interface.

## 2. CONTEXT OF THE ARCAS PROJECT

The objective of the ARCAS project is to develop an assessment and design methodology aimed at the renovation of buildings and groups of multifamily housing buildings of social interest, to address energy poverty and promote sustainable renovation, energy efficiency and healthy indoor environments in the SUDOE territory. The project is based on the integration of three research axes:

- AXIS 1 - Energy autonomy - efficiency
- AXIS 2 - Social quality - energy poverty
- AXIS 3 - Air quality - health

As a result of this integration, the work in the project is developed to determine the optimal relationship between the three mentioned axes and obtain the best energy efficiency while maintaining the social quality and well-being of citizens.

ARCAS is based on the use of similar climatology in the South Atlantic region for the development of a tool that allows, through key indicators, the design of building architecture based on maximizing energy efficiency, air quality and thus promoting social welfare, making use of the best available techniques, including renewable energy sources.

This project combines efforts to develop strategies and measures that facilitate the development of policies, at national, regional and local governments scale, for the renovation of multifamily housing buildings with great autonomy and energy efficiency (axis 1), with healthy air quality for building occupants (axis 3) and reducing energy poverty, which is so important in many European countries (axis 2).

ARCAS results and outcomes will be applicable and reproducible in the public and private institutions participating in the project and will be especially useful for professional associations, manufacturers, builders and for national, regional and local public administrations.

The Action Plans that will be developed in an integrated manner on the three axes of the research project by ARCAS beneficiaries, in collaboration with ARCAS associated partners, constitute a key element that will ensure the transfer of knowledge to the entire SUDOE territory, as well as the future sustainability of the ARCAS methodology.
From a methodological point of view, the project is structured in different Work Packages (WP). In the first phase, the indicators that will be used in the ARCAS methodology are defined. These indicators are proposed within the first four Work Packages, as well as the specifications and protocols for their quantification. Those four Work Packages are specifically:

- WP 1 - Climate indicators selection
- WP 2 - Selection of energy efficiency indicators in residential buildings
- WP 3 - Selection of indicators on best technologies available in renewables
- WP 4 - Selection of social quality indicators

In WP 5, the ARCAS methodology will be developed and implemented in a computer tool. Therefore, it is essential that the indicators selected in the previous WPs can be measurable and evaluable, in addition to being compatible with their application to different types of residential buildings and in different countries.

The methodology will be validated in WP 6. For this, a set of demonstration buildings will be selected. As selection criteria, buildings that include a casuistic representative of the three axes considered and the three countries of the consortium will be sought.

WP 7, WP 8 and WP 9 encompass the part of the project that can be considered as the capitalization part. More specifically, in WP 7, the ARCAS certification procedure will be detailed, generating a series of guides for project owners, and other relevant actors that will audit ARCAS projects. This work will be carried out in coordination with the associated partners of the project. As for WP 8, this group of tasks has as its main objective the training of professionals, and to achieve it, a training program will be defined to train professionals in the application and certification of the ARCAS method, and a pilot program will be provided training in professional institutions that belong to the ARCAS project value chain. Finally, in WP 9, strategies will be developed to establish new sustainability, energy efficiency and social quality policies in the renovation of multifamily buildings of social interest. This includes, amongst others, proposals for renovation policies, financing models and criteria to prioritize interventions. For that, the indicators defined in WPs 1 to 4 will be used and will be carried out in coordination with the public administrations and private organizations associated with the ARCAS project.

# 3. PRODUCTS INCLUDED IN THE VIEWER

## 3.1. Weather products

The following table shows the climatic variables, their units, and the dimensions of analysis that have been included in the form of climatologies in the climate map.

**Tabla 1: Variable of the climate map**

| Variable | Units | Data set | Temporary aggregation | Periods | Disaggregation | Magnitudes | ES | FR | PT |
|---|---|---|---|---|---|---|---|---|---|
| Climate Severity Index | Categories | ACAO | Average of annual averages | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Days with snow | Days | ACAO | Average annual counts | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Degree days | ºC | ACAO | Average | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Stormy days | Days | ACAO | Average annual counts | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Foggy days | Days | ACAO | Average annual counts | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Days with hail | Days | ACAO | Average annual counts | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Hours of sun | Hours | ACAO | Average | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Average global radiation on horizontal surface | $kWh/m^2$ day | ACAO | Average | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Minimum temperature | ºC | ACAO | Average | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Medium temperature | ºC | ACAO | Average | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Maximum temperature | ºC | ACAO | Average | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Total rain | mm | ACAO | Average annual sums | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Rain > 0.1mm | Days | ACAO | Average annual counts | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Rain > 1mm | Days | ACAO | Average annual counts | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Rain > 10mm | Days | ACAO | Average annual counts | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Rain > 30mm | Days | ACAO | Average annual counts | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Maximum rain | Days | ACAO | Average of annual maximums | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Full evaporation | mm | ERA5 | Average | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Net solar radiation | $J/m^2$ | ERA5 | Average | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Incident solar radiation | $J/m^2$ | ERA5 | Average | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Dew point temperature | ºC | ERA5 | Average | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Temperature at 2m | ºC | ERA5 | Average | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |
| Minimum temperature | ºC | ERA5 | Average | 1981-2010 | Annual, monthly, seasonal | Value | | ✔ | ✔ |
| Maximum temperature | ºC | ERA5 | Average | 1981-2010 | Annual, monthly, seasonal | Value | | ✔ | ✔ |
| Direction of the wind | ºnorth | ERA5 | Average | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Wind speed | m/s | ERA5 | Average | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | ✔ | ✔ |
| Total rain | mm | ERA5 | Average annual sums | 1981-2010 | Annual, monthly, seasonal | Value | ✔ | ✔ | ✔ |
| PM2.5 | ug/m3 | CAMS | Average | 2003-2020 | Annual, monthly, seasonal | Value | ✔ | ✔ | ✔ |
| PM2.5 > 25 | Days | CAMS | Average annual counts | 2003-2020 | Annual, monthly, seasonal | Value | ✔ | ✔ | ✔ |
| $CO_2$ | ppm | CAMS | Average | 2004-2020 | Annual, monthly, seasonal | Value | ✔ | ✔ | ✔ |
| PM2.5 > 15 | Days | CAMS | Average annual counts | 2003-2020 | Annual, monthly, seasonal | Value | ✔ | ✔ | ✔ |
| CO | ppm | CAMS | Average | 2003-2020 | Annual, monthly, seasonal | Value | ✔ | ✔ | ✔ |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Weather severity in winter | Categories | CORDEX | Average | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, change | ✔ | ✔ | ✔ |
| Weather severity in summer | Categories | CORDEX | Average | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, change | ✔ | ✔ | ✔ |
| Snow > 0.2mm | Days | CORDEX | Average annual counts | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, change | ✔ | ✔ | ✔ |
| Heating degree days | ºC | CORDEX | Average | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, change | ✔ | ✔ | ✔ |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Cooling degree days | ºC | CORDEX | Average | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, change | ✔ | ✔ | ✔ |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Hours of sun | Hours | CORDEX | Average | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, change | ✔ | ✔ | ✔ |
| Descending net radiation | $W/m^2$ | CORDEX | Average | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, change | ✔ | ✔ | ✔ |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Shortwave incident solar radiation | W/m² | CORDEX | Average | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, change | ✔ | ✔ | ✔ |
| Minimum temperature | ºC | CORDEX | Average | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, change | ✔ | ✔ | ✔ |
| Maximum temperature | ºC | CORDEX | Average | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, change | ✔ | ✔ | ✔ |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Total rain | mm | CORDEX | Average annual sums | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, relative change | ✔ | ✔ | ✔ |
| Rain > 0.1mm | Days | CORDEX | Average annual counts | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, change | ✔ | ✔ | ✔ |
| Rain > 1mm | Days | CORDEX | Average annual counts | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, change | ✔ | ✔ | ✔ |
| Rain > 10mm | Days | CORDEX | Average annual counts | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, change | ✔ | ✔ | ✔ |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Rain > 30mm | Days | CORDEX | Average annual counts | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, change | ✔ | ✔ | ✔ |
| Maximum rain | Days | CORDEX | Average of annual maximums | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, relative change | ✔ | ✔ | ✔ |
| Full evaporation | mm | CORDEX | Average | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, change | ✔ | ✔ | ✔ |
| Wind speed | m/s | CORDEX | Average | Historical: 1981-2010<br>Near future: 2016-2035<br>Medium future: 2046-2065<br>Far future: 2081-2100<br>Scenarios: historical, RCP4.5, RCP8.5 | Annual, monthly, seasonal | Value, relative change | ✔ | ✔ | ✔ |

### 3.2. Products in specific locations

For the subset of variables described in the previous table, the map includes, depending on the data availability, statistics of the distribution of the probability of daily values in different climatic periods. These statistics include the average, standard deviation, 5$^{th}$ and 95$^{th}$ percentiles, and histograms for each season of the year.

### 3.3. Process for obtaining products

The data transformation process followed to obtain the climate data included in the map is described below. The sub-processes included in this process are shown in the following figure.

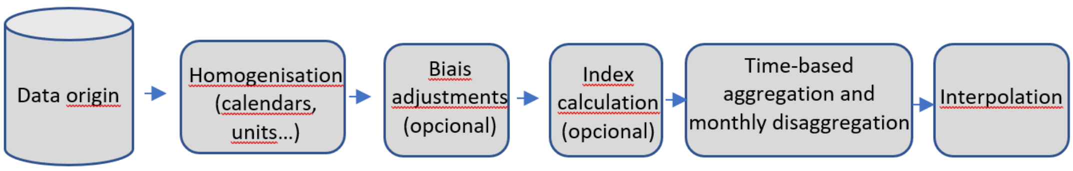



**Figure 1: Schema of the implemented data processing**

First of all, the data is downloaded from the different data sources described in the deliverable D1.1 titled "Climate and pollution indicators report". To this end, different API (Application Programming Interface) that have been enabled by the Copernicus Climate Data Store or Copernicus Atmosphere Data Store services are used.

The downloaded data, on a daily or monthly scale, are processed in order to standardize the units, the geographic coordinates, or the calendars that are used. Correction techniques of standard bias are then applied in the case of climate projections. In particular, the ISIMIP3 technique used in the latest IPCC assessment report has been used. Subsequently, if applicable, the derived index or magnitude is calculated to obtain the different climate data by means of aggregation in climatic time periods. Finally, the data is interpolated to obtain the final products.

## 4. USER INTERFACE

The current version of the application, which is still under development, is available at the following link: [https://fecea-viewer.predictia.es](https://fecea-viewer.predictia.es).

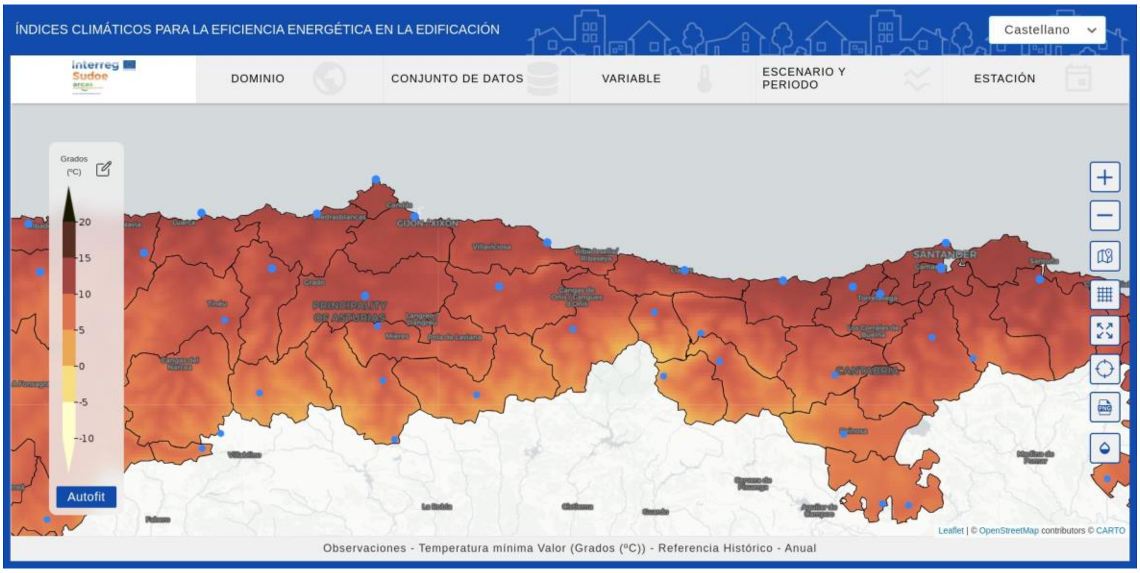



**Figure 2: Screenshot of the current version.**

## 4.1. Architecture and technologies

The architecture of the application is basically based on three components: a *frontend* component based on the React framework, a *backend* written in Java 17 using the Spring library, and a WMS server that expound weather information.

As for the *frontend*, the Leaflet library is used to represent the weather layers. For the representation of the wind direction layer, the Windy-Leaflet library is used. Both libraries are open source. The user interface that has been developed incorporates the following functionalities:

- Export of layers in GeoTIFF format
- Customizing display styles
- Image download
- Implementation of permanent links for sharing on social networks
- Connection with external WMS services such as the PNOA.
- Responsive designs taking into account the accessibility and usability of the user.
- Consultation of specific information

The *backend* component is in charge of transforming the information served by the WMS server and delivering it to the *frontend* in such a way that it is easy to implement and display new data added to the WMS server on the *frontend*. In this sense, a system based on configuration files has been developed that facilitates the loading of new information.

## 4.2. User interface and functionalities

Currently, the web viewer has a dynamic dropdown menu on the top, and the viewer itself at the bottom. The different items of the menu are dynamically generated from the information available in the WMS server, so the implementation of new types of variables is easily scalable. In addition, these elements are interdependent so that the options shown are adapted depending on what the user chooses.

In the first dropdown, there is the geographic domain of the data where the user can select an area (Spain, France or Portugal - Figure 3).


Figure 3: Selection of the geographic area

Following that, the type of data source can be selected. At the moment, there are three types of options: observations, reanalysis and climate projections (Figure 4).


Figure 4: Selection of data source type

The third dropdown lists the variables that are available for that type of source, grouped by type (Figure 5).


**Figure 5: Selection of the variable**

In the fourth dropdown includes options for predictions such as the future period (near, medium, or distant), the magnitude (specific value or anomaly) of the variable, or the type of scenario for the prediction. Currently, only historical information is recorded, so this menu hides the type of scenario and informs about the period and magnitude (Figure 6).


Figure 6: Selection of the time period, scenario, and magnitude

The last dropdown shows a temporary disintegration filter in which different options are collected depending on their availability (Figure 7).

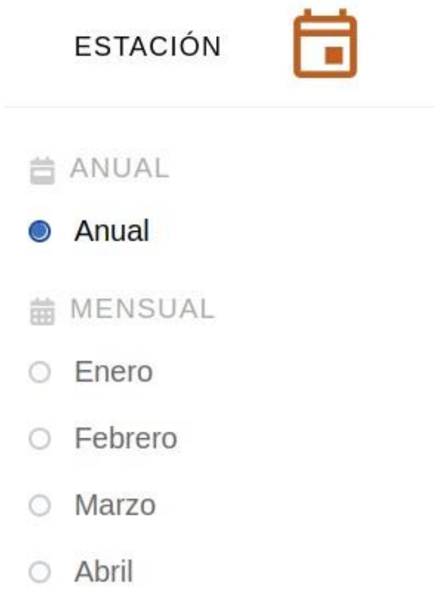


Figure 7: Selection of temporal disaggregation

On the other hand, the viewer itself consists of a map, which shows the information chosen by the user combined with the available stations, some standard controls, and a dynamic legend. Currently, among these controls there is the possibility of zooming, choosing the type of map (such as orthophoto, street, dark, or light), making it full screen, obtaining information on a specific point on the map, downloading the map as an image and changing the opacity of the displayed layer. As for the legend, it provides an editor for the limit of magnitude established by default so that, if necessary for some variable, the colors of the layer can be displayed correctly (Figure 8, left). In addition, the option to auto-adjust the range to the maximum and minimum value displayed is included (Figure 8, right).

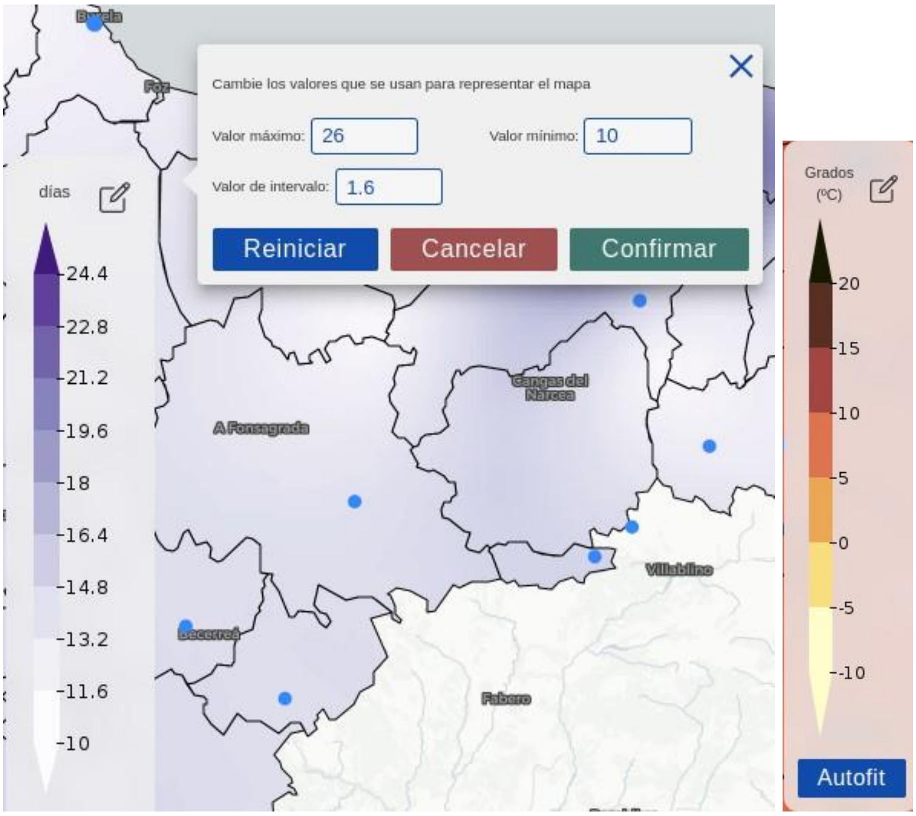


Figure 8: Setting up of the legend limits (left) and auto-fit button (right)

Each variable has been configured with a map of colors and ranges appropriate to its nature and magnitudes. All the variables are represented in the form of choropleths, except for the wind direction variable, which is represented as a vector (Figure 9).

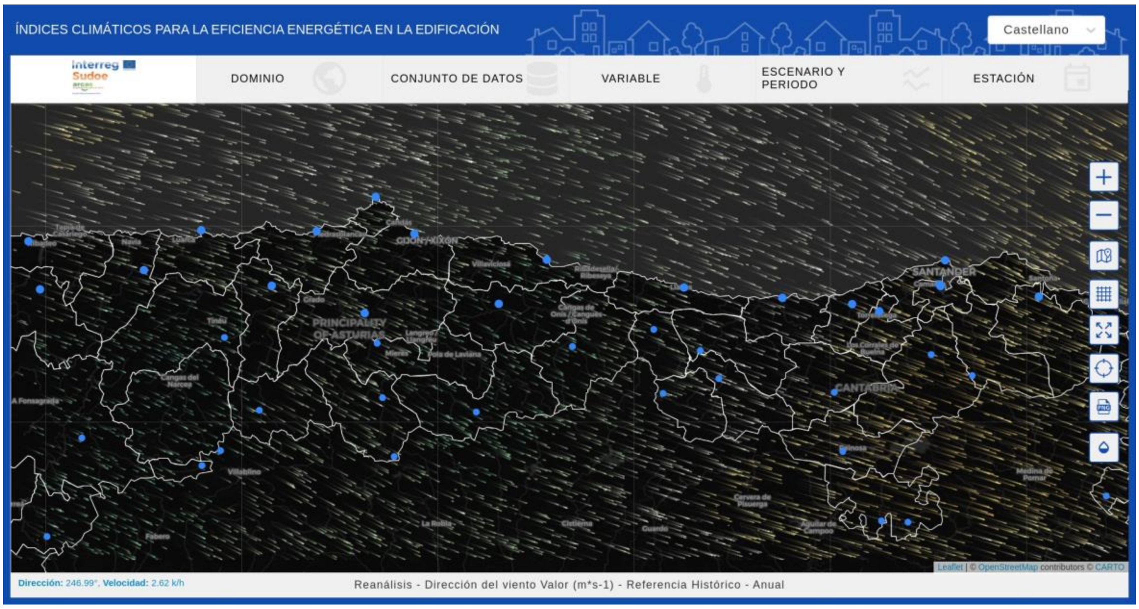



Figure 9: Representation of the wind direction variable

As mentioned previously, apart from including weather in the layer of a map, the application offers detailed information on specific locations. To do this, the blue points represented on the map must be clicked (Figure 10). At that moment, a modal window appears to the user in which information is represented for different greenhouse gas emission scenarios (historical, RCP4.5 and RCP8.5). In each of the tabs, a table is included with the distribution of values of the different climatic variables. This distribution is characterized by some statistics, a box plot, and some seasonal histograms. In addition, the user is allowed to download this information or the files in .MET format generated by AEMET. This screen includes many help texts in order to make it self-explanatory. Equivalent functionalities are offered for France (Figure 11) and Portugal (Figure 12).

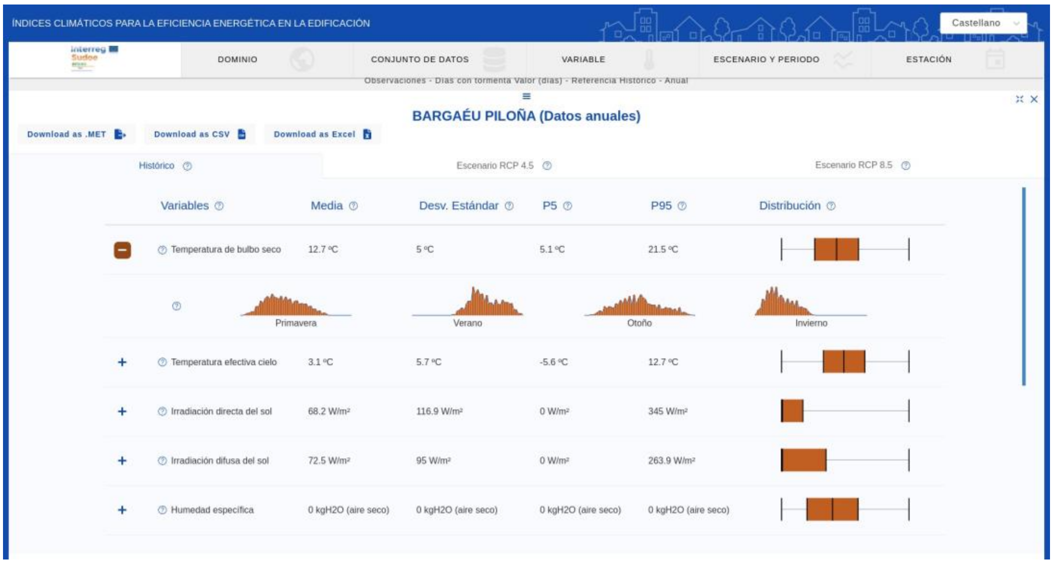



Figure 10: Detailed information on locations

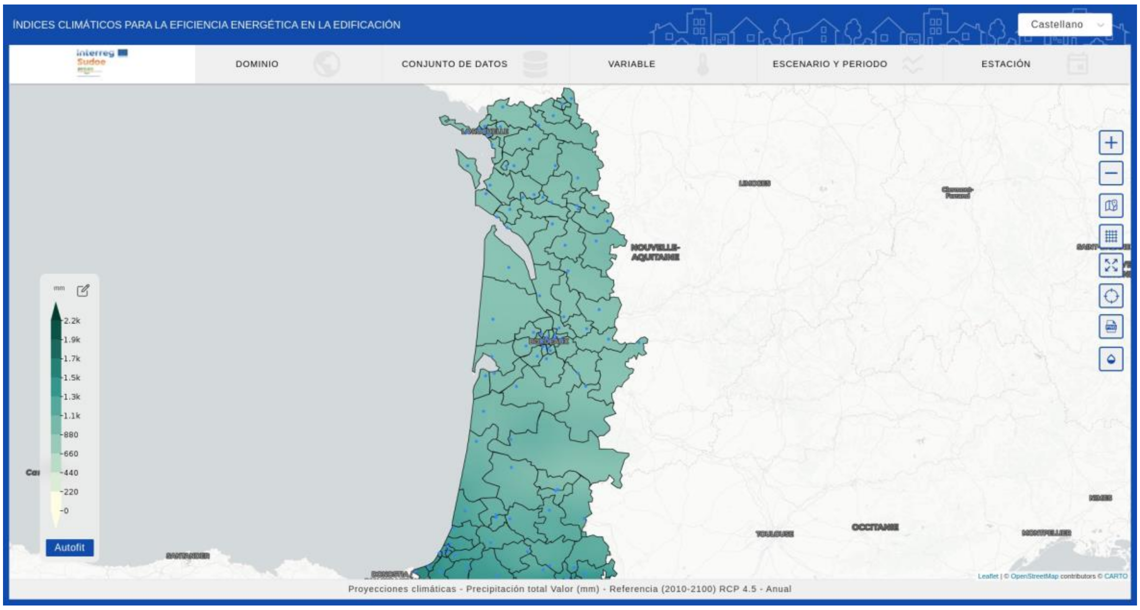



Figure 11: Screenshot of the French area

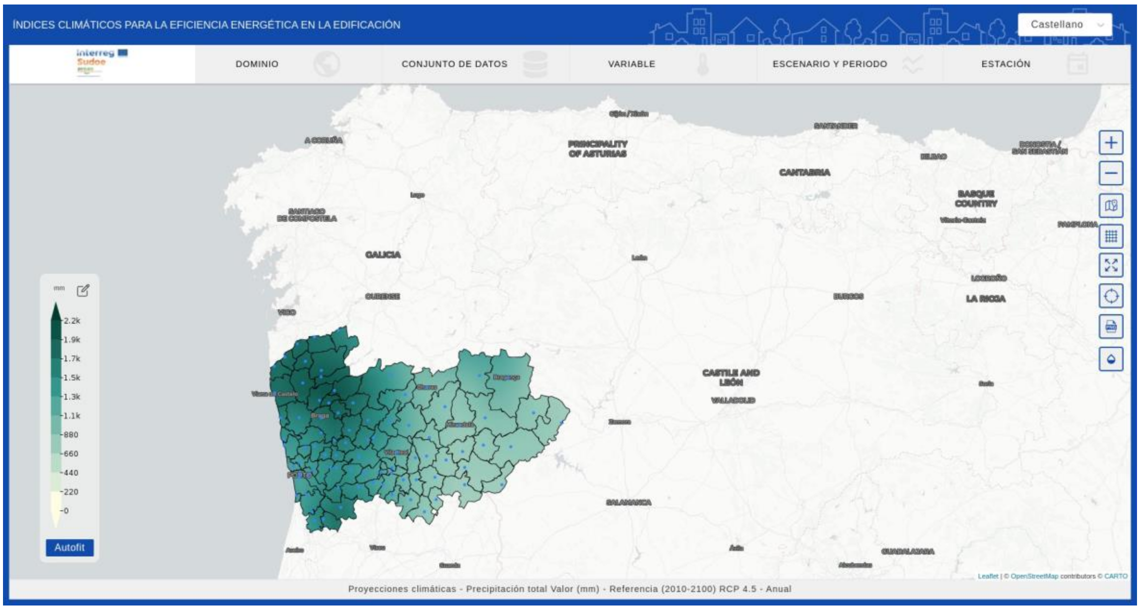


Figure 12: Screenshot of the Portuguese area

The application also includes these features:

- Translation into English and Spanish through a selector located on the right of the top menu
- Homepage with the geographical domains considered and information about the ARCAS project including partnership (Figure 13)
- Specific base layers for each geographic domain (*e.g.*, the National Air Orthophoto Plan-PNOA, for Spain). See Figure 14.

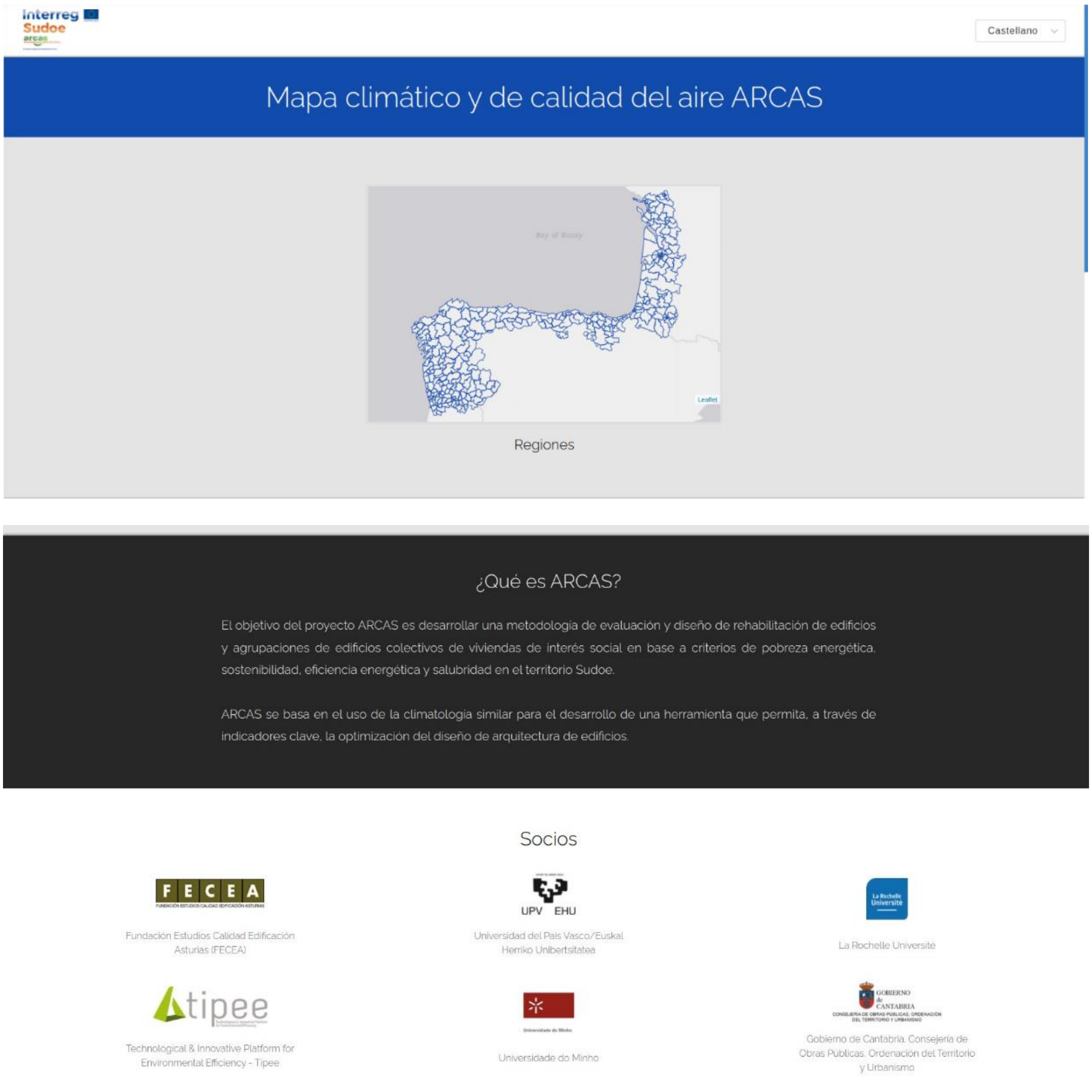
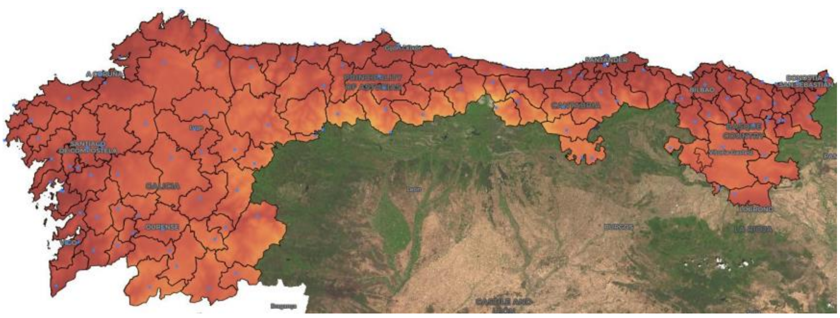



**Figure 13: Screenshot of the homepage**



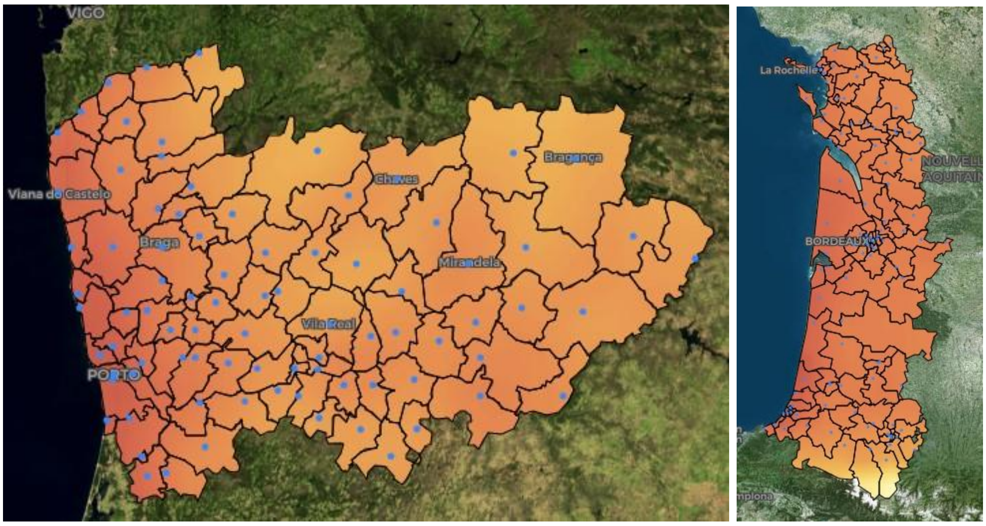



Figure 14: Different base layers per area

# 5. MAPPING AIR QUALITY

## 5.1. Using open source databases

### *5.1.1. Why using databases?*

The study of the information available via the internet to carry out a Diagnosis of the potential risk associated with air quality has provided different lessons. The data were used to remove air quality indicators that allow for rapid vision through maps or color representations.

A part from the fact that the European database is more easily accessible, it can be noted that the air quality index does not express the same level of global pollution. Indeed, between the AQI index and the EAQI index, CO and $PM_{10}$ are not considered in either and the values used to delineate the classes of pollution levels are not identical.

Referring to the socio-economic cost issue, it seems that the consideration of particles is now more impactful than that of carbon monoxide, although the latter is a much more dangerous compound in the very short term.

It is thereby proposed on the one hand to consider the data from the European website but not to use it from the available graphic representation but rather to work on the raw data.

It therefore seems possible to use this data centralizing the pollutants measured in the air to define an overall index applied to ARCAS. The use of this raw data should allow both to provide information on the overall level of pollution but also to provide a set of external pollutant data that can be used to determine whether the level of pollution has a negative impact on housing according to the current air renewal strategy.

To achieve these two objectives, a pre-processing work is required for the integration of information into the ARCAS assessment.

### 5.1.2. Methodology proposed to determine pollutants of interest

The following methodology will be carried out:

- Identification of the most frequently measured pollutants in the SUDOE zone in Spain, France and Portugal. Reconciliation of ambient pollution profiles related to localization, urban planning and human activity.

This type of analysis ensures that pollutants used to calculate indicators are interesting according to time scales and in relation to the mesh of all the stations present in the SUDOE zone.

This analysis sheds light on the lack of sufficient data for certain pollutants such as $PM_{2.5}$.

Finally, if we want to push the analysis by a serial decomposition of Fourier of all the time series (1 per station and pollutant), this can be used to identify typical profiles.

The main objective was to identify and supplement data gaps, enabling the proposed indicators to be obtained at European level. https://www.eea.europa.eu/themes/air

- Choice of pollutants that can allow a relevant assessment of ambient air quality and for which a constructive or system-type adaptation may reduce the risk of impact on occupant health.

According to data published by the European Environment Agency (EEA), air pollution from the three main pollutants ($PM_{2.5}$, $NO_2$ and $O_3$) is the third leading cause of premature death in France with 47,300 deaths annually, behind tobacco (80,000 deaths per year) and alcohol (49,000 deaths per year). Public Health in France estimates that at nine months the average gain in hope deviates to 30 years if all the municipalities of mainland France respected the levels of fine particulate matter $PM_{2.5}$ observed in the 5% of the least polluted municipalities of the same urbanization class.[1]

| | Particules fines | Dioxyde d'azote | Ozone | Cumul |
|---|---|---|---|---|
| **Nombre de décès prématurés** | | | | |
| *Union européenne (28)* | 391 000 | 76 000 | 16 400 | 486 400 |
| *France* | 35 800 | 9 700 | 1 800 | 47 300 |
| **Années de vie perdues** | | | | |
| *Union européenne (28)* | 4 150 000 | 795 000 | 180 000 | NA |
| France | 414 700 | 112 400 | 21 600 | NA |

*Source: AEE, Air quality in Europe in 2018*

**Figure 15: Estimate of the number of premature deaths and years of life lost due to PM2.5, NO2 and O3 pollution**

In view of this information, it seems appropriate to take these three pollutants into account in order to determine a simplified index of Outdoor Air Quality in the context of the ARCAS project. This indicator would be very close to the European index, only Sulphur dioxide would not be integrated. Sulphur dioxide is indeed a pollutant of the outside air but it is not considered at the problem level indoors and does not seem to have remediation solutions related to buildings. By moving away from this parameter, it is possible to avoid a fortuitous deterioration of the overall index to guide the diagnosis of the particular building.

It should also be noted that some data remain missing to achieve a robust global case due to a limited number of measurement stations capable of giving a $PM_{2.5}$ measurement, contrary to the data available for $PM_{10}$. It appears, therefore, that in order to obtain a finer mesh of pollution indicators, an approximation of $PM_{2.5}$ will have to be made.

To address this problem, it will be possible to either use a comparison of the average results given by each station in each country or use a specific average ratio, for example in the following publications : $R_{PM2.5/PM10}$ = 0,53 – 0,59 (https://www.lcsqa.org/system/files/Etude_stat_PM2.5_mars07_v-finale_corr-27-09-07_1.pdf) ; $R_{PM2.5/PM10}$ = 0,65 (https://aaqr.org/articles/aaqr-16-02-oa-0081.pdf) ; $R_{PM2.5/PM10}$ = 0,62 in urban site et 0,68 in rural areas (https://aaqr.org/articles/aaqr-16-09-oa-0406.pdf).

Indicators can be calculated with aggregated data over one or more representative years. The results can be represented as a dynamic map via the TABLEAU software.

The ambient air quality classes for each pollutant may be based on the methodology used by the EAQI index. The annual average will, however, provide a long-term representation of the level of pollution in the future buildings.

### *5.1.3. Database processing methodology to the ARCAS Air Quality Diagnostic Tool*

The public metadata files from https://www.eea.europa.eu/data-and-maps/data/aqereporting-8#tab-european-data have been used to get the key statistics for every measurements stations in the concerned countries and for each measured pollutants. Some of the stations recover data daily, other hourly. For each sampling, the mean and the median are available. A Python code was used to get for every year, station and pollutant these statistics and aggregate them into a single .csv file. This file was the input data used to build the interactive map described in the next sub-section.

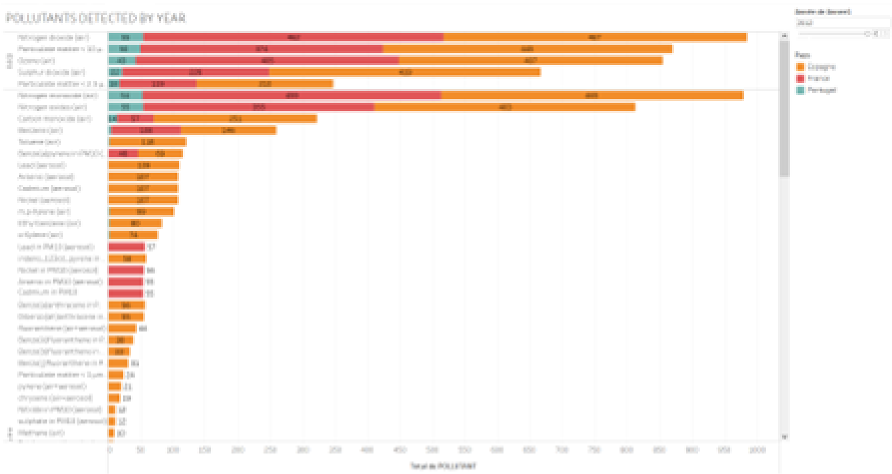



**Figure 16: Compilation of pollutant data collected by year in Europe.**

### *5.1.4. Methodology for exploiting data using the TABLEAU software*

## A. Data wrangling with Tableau Prep

Once the raw data has been extracted by the Python code, the next step is the wrangling data. Data wrangling is the process that allows you to discover, structure, clean, enrich and validate raw data so that you can publish the results in an adapted format to data analysis. To achieve this step, we use ETL (Extract-Transform-Load) tools that are specialized software in data shaping. The ETL tool used in this project is Tableau Prep which is available with Tableau Software.

For this project, the raw data extracted is divided into several files according to their category (characteristics of the measuring stations, data recorded by the measuring stations) and their country (France, Spain, Portugal). We start by merging files from the same category of different countries together and removing unnecessary columns in datasets.

Based on data recorded by the measuring stations, we calculate the average value of each EAQI pollutant per measuring station per year. We then rank these values according to the labels of the EAQI indicator (GOOD, POOR, etc.).

Based on these values, we recover the worst by measuring station per year to derive the value of the EAQI index. This information is then added to the measurement station features dataset.

Then, we merge the list of French municipalities with the dataset of radon potential by municipality. The main problem that appeared during data wrangling is that the EAQI index is linked to a year but not the radon potential. Assuming that the radon potential does not change from year to year, we assign the radon potential of each municipality each year attached to the EAQI index and we add this data in the dataset of the characteristics of the measuring stations by linking them via the municipality's zip code.

At the end of this work, we obtain a file with all the measurement stations and their characteristics (coordinated, EAQI annual values, radon potential) and a file with the data measured by these stations.

## B. Data visualization with Tableau Software

With Tableau Software, we import these two files and connect them through their common columns: measurement station and year.

In a first map view, we create a first layer in map format on which we display the communes with a color corresponding to their radon potential. Then we create a second layer in circle format, over the first layer, which indicates the location of the measuring stations on the map and with the color-coded EAQI index of the circles and the number of EAQI pollutants detected as circle size. We finally add the year as a filter on the map.

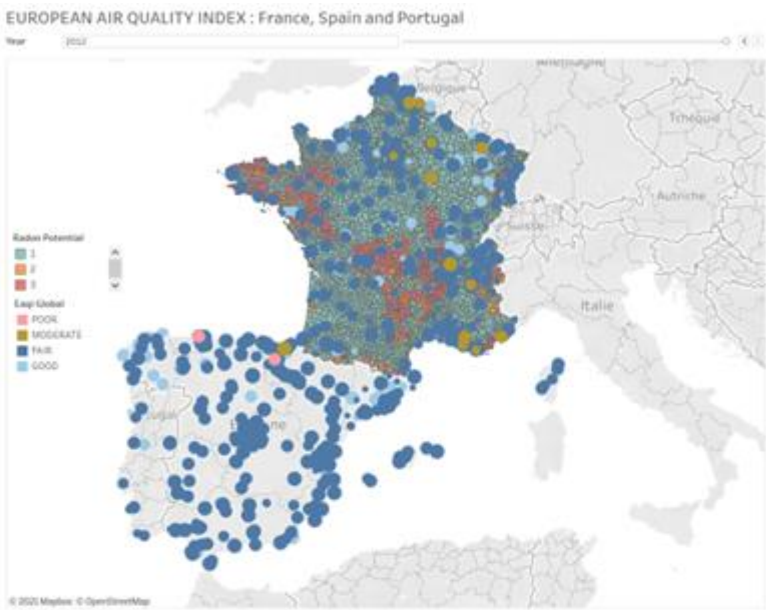
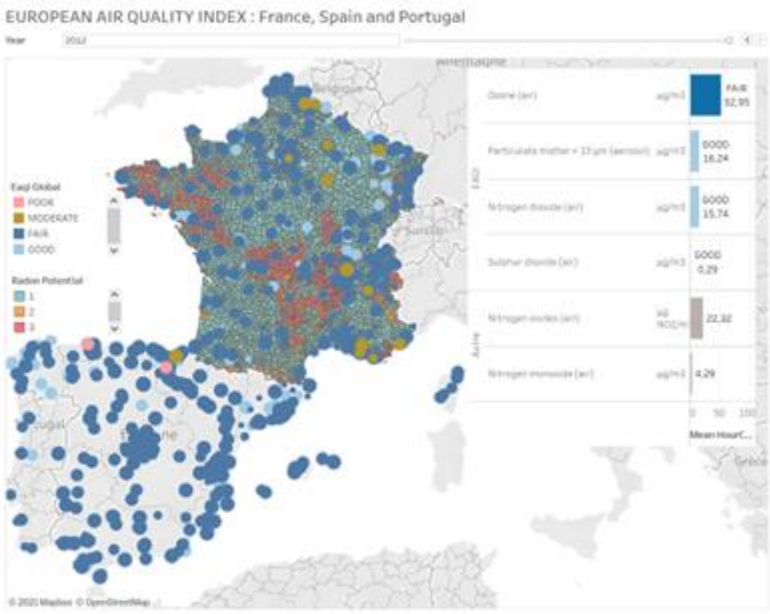



Figure 17: Pollutant map views with TABLEAU.




In a second map view, we use the dataset recorded by the measuring stations to display in a histogram the measured value of each pollutant. For the occasion, we create a family to distinguish pollutants into two groups: the pollutants of the EAQI and the others. The color code of the histogram then uses the same color code based on the EAQI index as on the map.

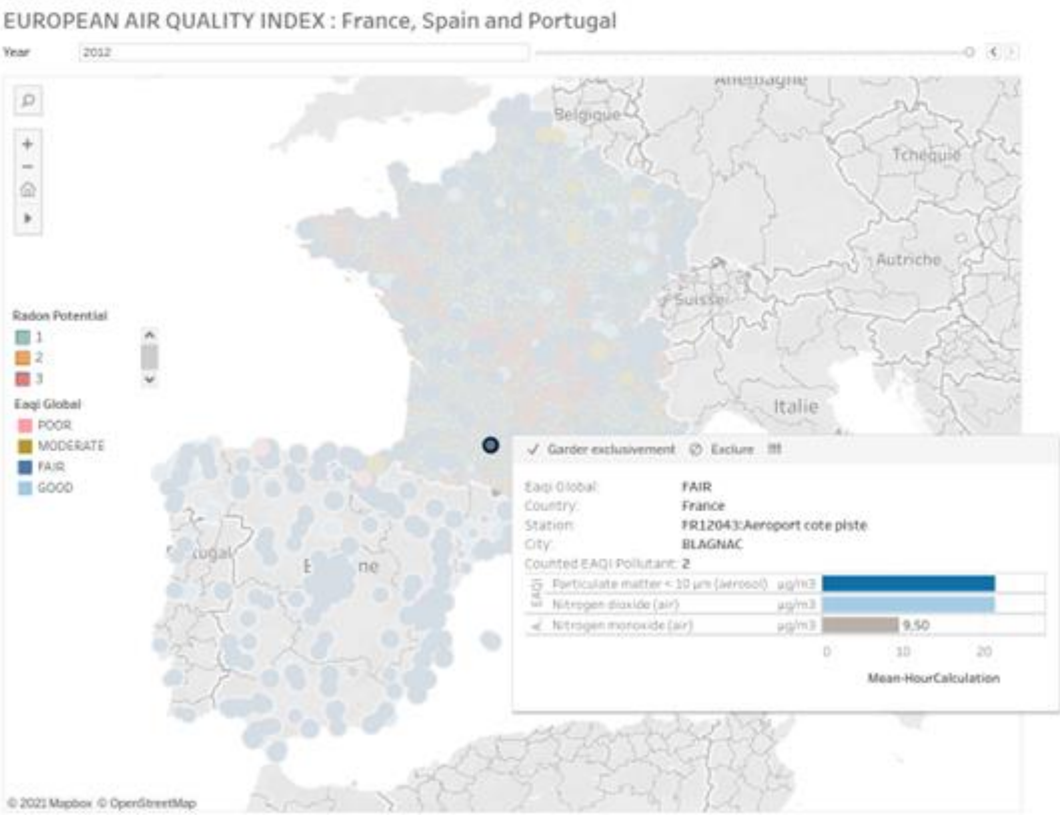



**Figure 18: Pollutant map view with EAQI index bubble.**

In the first view, we set up for the second view to appear in a bubble when we pass our cursor over the measuring stations displayed on the map. Then, the Tableau software automatically performs the same filtering on the second view.

Finally, we created a dashboard that allows the document's title and filters to be shaped. An example could be tested here : [ARCAS.twbx](ARCAS.twbx), using the Tableau Reader (Free), downloadable here: https://www.tableau.com/fr-fr/products/reader.

To use it, first you need to download and install Tableau reader. Then you can open the ARCAS.twbx link.