Environmental and bioclimatic data for epidemiological analysis over French Mediterranean areas

Camille Portes, Dino Ienco, Eric Verdin, Edith Gabriel

▶ **To cite this version:**

Camille Portes, Dino Ienco, Eric Verdin, Edith Gabriel. Environmental and bioclimatic data for epidemiological analysis over French Mediterranean areas. Environmental Data Science, 2025, 3, pp.e31. 10.1017/eds.2024.50 . hal-04881001

HAL Id: hal-04881001
https://hal.science/hal-04881001v1

Submitted on 13 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Environmental Data Science (2024), 3: e31, 1–11
doi:10.1017/eds.2024.50

CAMBRIDGE UNIVERSITY PRESS

DATA PAPER

Environmental and bioclimatic data for epidemiological analysis over French Mediterranean areas

Camille Portes[1], Dino Ienco[2,3], Eric Verdin[4] and Edith Gabriel[1]

[1]BioSP, INRAE, Avignon 84914, France
[2]UMR TETIS, INRAE, Univ. Montpellier, Montpellier 34000, France
[3]INRIA, Univ. Montpellier, Montpellier 34000, France
[4]Pathologie Végétale, INRAE, Avignon 84140, France
Corresponding author: Camille Portes; Email: camille.portes@inrae.fr

Received: 07 March 2024; **Revised:** 08 October 2024; **Accepted:** 16 October 2024

Keywords: climate; epidemiology; France; land type; machine learning

Abstract

Risk-based surveillance is now a well-established paradigm in epidemiology, involving the distribution of sampling efforts differentially in time, space, and within populations, based on multiple risk factors. To assess and map the risk of the presence of the bacterium *Xylella fastidiosa*, we have compiled a dataset that includes factors influencing plant development and thus the spread of such harmful organism. To this end, we have collected, preprocessed, and gathered information and data related to land types, soil compositions, and climatic conditions to predict and assess the probability of risk associated with *X. fastidiosa* in relation to environmental features. This resource can be of interest to researchers conducting analyses on *X. fastidiosa* and, more generally, to researchers working on geospatial modeling of risk related to plant infectious diseases.

Impact Statement

Data collection, transformation, and combination can be time-consuming, especially for scenarios that demand access to official data via complicated administrative procedures. This data paper has the objective to provide a ready to use easy access to data supporting epidemiological analysis in the South of France. Such data can be useful for identifying risk factors, training machine learning models, and providing additional clues to model plant infectious epidemics.

1. Introduction

Epidemiological research in the domain of plant pathology is increasingly reliant on high-resolution spatial information to study spatial pattern and predict breakouts, especially at a large scale (Ojiambo et al., 2017). To support studies related to climate change and the related increase of plant pathogens spread events, the necessity to access ready to use data covering environmental stressors, climatic information, land use statistics and soil composition is increasing. This type of information can support the design of epidemiological surveillance systems, prediction of sensitive areas, and thus for planning strategies to prevent the spread of plant diseases.

Our contribution, here, focuses on a resource containing ready-to-use data to support the prevention of the spread of the *Xylella fastidiosa* disease, a bacteria that obstructs plant vessels,

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/eds.2024.50 Published online by Cambridge University Press

leading to the dehydration and eventual death of the plant. Since 2015, which is the start of the detection campaign in France, about 2500 plants have been tested positive to *X. fastidiosa*, with over 400 plants tested positive in both 2021 and 2022. Currently, positive cases have only been reported in three regions: Corsica, PACA (Provence-Alpes-Côte d'Azur), and Occitania, which motivated our work to focus on these regions. The number of positive cases found in the Occitania region increased in 2022 and 2023. To date, five subspecies of *X. fastidiosa* have been identified: fastidiosa, morus, multiplex, pauca, and sandyi. The multiplex subspecies can be found anywhere in Europe while the subspecies pauca is mainly found around the Mediterranean basin (Trkulja et al., 2022). Most cases in France involve the multiplex subspecies, making it a crucial subspecies to monitor.

With the aim to contribute to a better understanding of the epidemic related to the *X. fastidiosa* bacteria, we have gathered information to create a dataset to understand the correlation between the environment and the presence of the bacteria. To our point of view, this dataset constitutes an important resource to study and model potential breakouts by identifying areas with an high probability of infection. This dataset is a combination of information about climatic conditions, land type, and soil composition, which are information needed assessing the environmental suitability for the spread of this disease. All this multisource information has been standardized to the spatial resolution of (confidential) bacteria observation, namely a 500 m × 500 m grid.

Moreover, the value related to the collected data is not limited to the study of the *X. fastidiosa* bacteria, and it can be employed to study several plant pathogens, such as the Barley yellow dwarf virus. Similar to *X. fastidiosa*, these pathogens are influenced by environmental, soil, and climatic conditions, as shown in Ingwell et al. (2017). In Lemaire et al. (2022), the authors highlight the need of a wide range of biotic and abiotic factors to explain forest dieback in South East of France. Many plant pathogens are susceptible to the same stressors, such as climate, soil type, soil content, and geographical characteristics (altitude and orientation). This is the reason why the collected data can be valuable for supporting analyses related to any other pathogen present on the same geographical area.

2. Study area

We focus on the three administrative French regions, among the 13 that have been sampled in France, where the bacteria is present. The study area thus contains the administrative regions: Occitania, PACA, and Corsica, which are located in the South of France (Figure 1).

The total surface of the study area is 112,446 km^2 with a per region surface of 72,724 km^2, 31,400 km^2, and 8722 km^2 for Occitania, PACA, and Corsica, respectively. The three regions are, as well, characterized by distinct climatic and environmental conditions.

The PACA region is characterized by the two main landscapes: the northern/eastern part belongs to the Alps mountain chain with a peak reaching 4102 m and the southern part is the Mediterranean basin (Techno-Science, 2013). In the western, lies the Rhône valley with a marshy area before reaching sea. The climate of the southern part of the region is characterized by a typical Mediterranean climate with warm temperatures and seasonal precipitation, although influenced by winds blowing down the Rhône valley. The northern/eastern part is a mountainous climate.

Corsica is a mountainous island with a mountain chain stretching less than 200 km in length at the center of the island, featuring high peaks, the highest of which is Monte Cinto at 2710 m (Préfectures-régions, 2015). Much of the island's vegetation consists of *maquis*, which is a specific Mediterranean vegetation made of high evergreen shrubs and bushes. The climate in this region is Mediterranean along the edges and mountainous in the center.

The Occitania region is rather wide compared to the two other regions of the study area and has different types of landscapes: the southern part borders Spain with the Pyrenees mountain chain; the eastern part is along the Mediterranean coast; the northern part consists of foothills of Massif Central; and the remaining western part comprises grasslands, hills, valleys, and major rivers. The region's climate is influenced by its landscape: the southern and northern parts experience a mountain climate, the western

https://doi.org/10.1017/eds.2024.50 Published online by Cambridge University Press

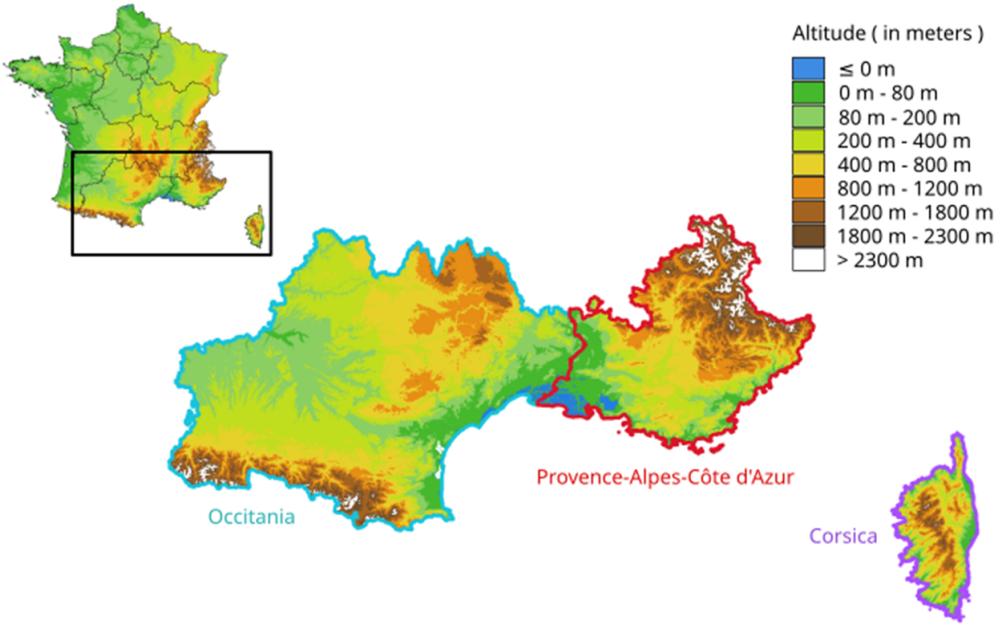


Figure 1. *Relief of the study area from Tile-Grabber (2013).*

part is characterized by a continental climate with Atlantic influences, and the eastern part has a Mediterranean climate (Siges-Occitanie, 2022).

3. Selection of dataset sources and variables

The main idea of this work is to gather together growth promoters and stressors that would influence plant development and plant pathogen epidemiology in general. We chose to combine bioclimatic, soil composition, land type, land orientation, and altitude data.

These variables are natural choices for explaining pathogen spread on plants, as they influence plant growth. The selected information pertains to factors that affect general plant systems, rather than specific species, because our bacteria can infect a wide range of plants, and the output data are intended for use with other pathogens as well.

The climatic variables were chosen in the Météo-France (2020) dataset, which is an updated data source. For our purpose, we needed data up to the end of 2020. The objective is to recover the variables that are correlated with the presence of *X. fastidiosa* in other sources. For instance, Farigoule et al. (2022) found that areas with milder winters and warmer springs and falls are more at risk since it sets suitable conditions for the vectors and thus for the spread of the pathogen. Climatic information is often used in biological and ecological studies for species analysis, as in Roubal et al. (2013), the authors propose a model based on rainfall events, air temperature and duration of relative humidity to predict *Fusicladium oleagineum* presence on olives.

In addition to climatic variables, we chose to use bioclimatic variables (Table A1), as they represent different aspects of temperature and precipitation regimes over time. In Martinetti and Soubeyrand (2019), the authors used bioclimatic variables and found that the presence of *X. fastidiosa* is correlated with climatic conditions, such as precipitation seasonality, minimum temperature in winter, precipitation during the dry season, and solar radiation. These indicators give insights on yearly and seasonal patterns of temperature and humidity and can be calculated with the information at our disposal using the formulas in O'Donnel and Ignizio (2012).

For the land type variables, we chose the Corine Land Cover dataset (C. L. C. Copernicus, 2020), as it provides detailed information within each categories of land use and land cover. We aim to capture the

https://doi.org/10.1017/eds.2024.50 Published online by Cambridge University Press

specific environments and vegetation where *X. fastidiosa* spreads, especially in Corsica, where the multiplex subspecies is predominantly found in maquis (European Food Safety Authority (EFSA) et al., 2022). Landscape types, as discussed in Section 2, differentiate each study region.

We decided to include chemical elements in the dataset because, in Del Coco et al. (2020), correlations have been observed between the presence of the bacteria and the quantity of elements in the soil. The authors established a correlation between a high content of zinc and copper in soil and leaves and the absence of the bacteria. They also observed higher manganese concentrations in less sensitive olive tree species, which could explain the absence of the disease. The chemical elements were taken from InfoSol (2015) as they contain the principal chemical elements, such as copper, zinc, and manganese.

In Ge et al. (2020), the authors investigated the effect of a copper-amended treatment and found that zinc, manganese, and magnesium concentrations increased when the treatment was administrated but did neither slow nor stop the development of the disease. In addition, they observed that a high concentration of copper in sap seemed to increase the virulence of *X. fastidiosa.* Manganese is a chemical element correlated with the plant resistance to the disease (D'Attoma et al., 2019).

The orientation dataset comes from Copernicus (2019), it is given at a high resolution for the entire Europe. The orientation influences the way a plant growth in terms of its associated health status due to possible exposure to climate and weather factors (Auslander et al., 2003). Hence, impact the robustness of the plants toward diseases, such as exposure of the land, wind, and altitude.

The altitude information originally come from *Shuttle Radar Topography Mission* but we accessed to the data through an interface Tile-Grabber (2013) designed to provide easier access to it. We downloaded 10 files that cover the whole France.

4. Data processing

The data processing stage has the objective to combine together the different data sources described in Section 3. The data came in different formats (vector and raster); this prevented us ftrom using the information as they are in a standard process. The purpose was to obtain a spatially consistent data in which all the information are projected onto the same spatial grid, using the spatial grid associated with the *X. fastidiosa* data as the reference grid.

The information presented in Section 3 were provided at various spatial resolutions, thus requiring preprocessing to make them spatially homogeneous. Chemical elements and climatic datasets had a low-resolution scale, 16 km and 8 km, respectively, which represented a lower resolution than the target output. Therefore, we needed to transform them to align with the output's resolution. The Corine Land Cover and the orientation datasets, on the other hand, had high resolutions, 100 and 25 m, respectively, requiring the use of upscaling methods to match the resolution of the *X. fastidiosa* dataset. Figure 2 illustrates the different transformations.

4.1. Data transformation

The data on chemical elements in the soil were obtained from two sources that followed the same sampling protocol. Since the samples were collected from the exact same locations using the same process, we directly merged the two datasets.

For this data source, we only had the information on centroids of cells, which were spaced 16 km apart. To transfer the information from these cells to the target grid, we built a grid that is roughly the same as the original one using the Voronoï method. This method allowed us to recover a similar grid while filling the empty areas, resulting from missing points, with the closest point's information. In addition, it allowed us to fill all areas along the coast, as shown in Figure 3.

We adopt the same strategy to process the climatic dataset, where the points were spaced 8 km apart.

The Corine Land Cover dataset is available for the whole Europe. We spatially intersected the Corine Land Cover information with the geographical extent of our study are. The Corine Land Cover dataset, which has a 100 m resolution, is given in a raster layer format. It comprises 45 distinct values, each

https://doi.org/10.1017/eds.2024.50 Published online by Cambridge University Press

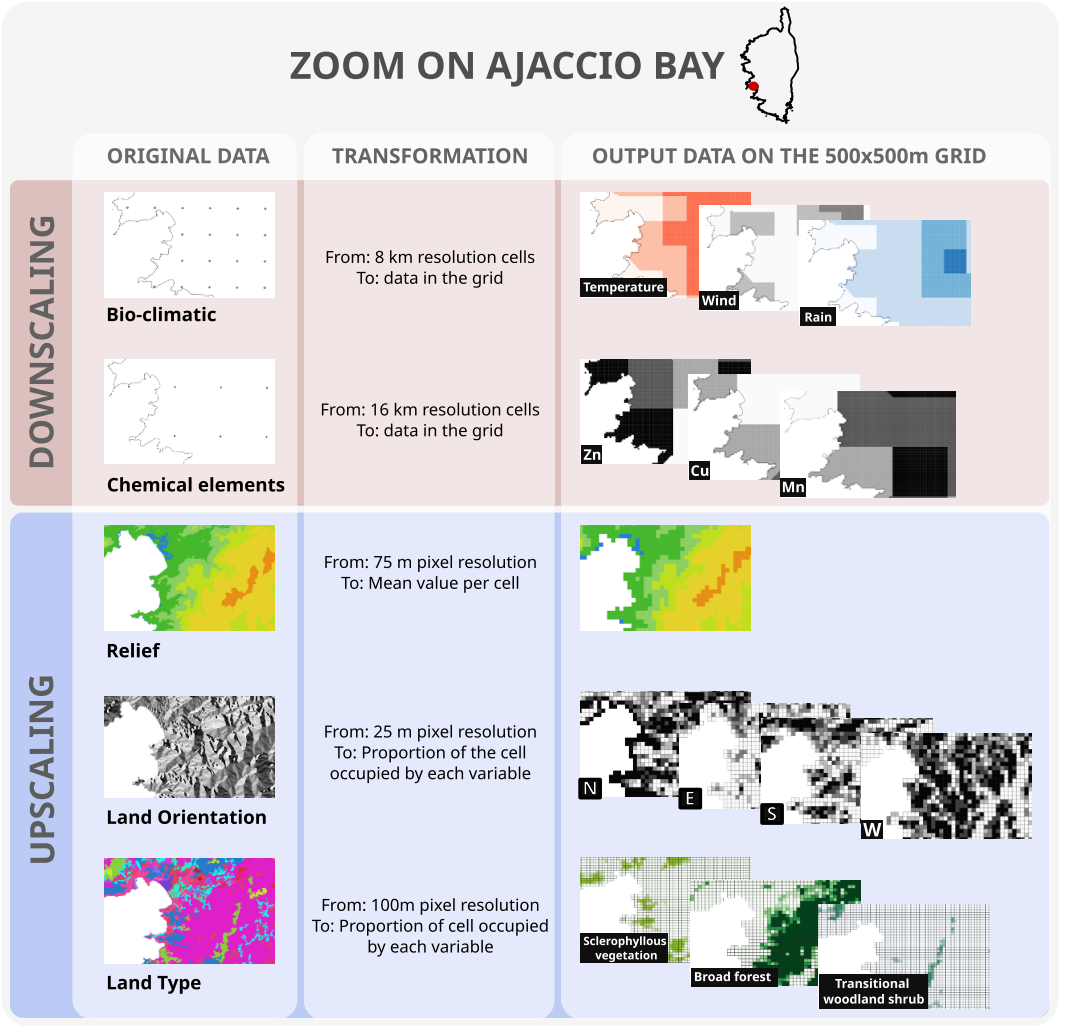
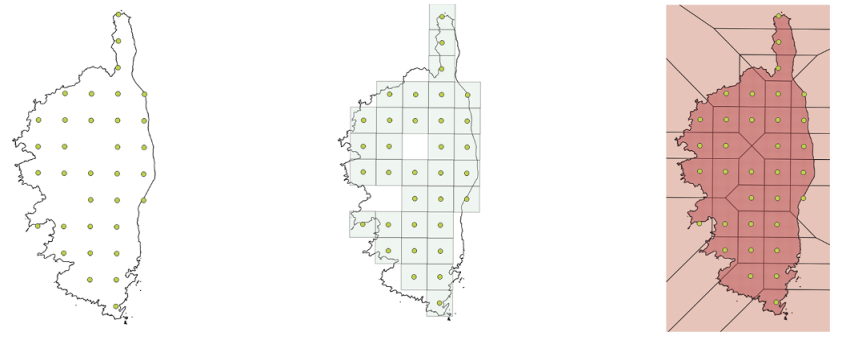


Figure 2. *Data transformation zoom on Ajaccio Bay.*



(a) Original points (b) Points' buffers (c) Voronoï polygons

Figure 3. *Interpolation of data with Voronoï method on chemical elements content in the soil variables.*

https://doi.org/10.1017/eds.2024.50 Published online by Cambridge University Press

representing a specific land type, including artificial surfaces, agricultural surfaces, forest and seminatural areas, wetlands, and water bodies. In order to aggregate the information at the scale of the *X. fastidiosa* grid, we computed the proportion of each soil type within the grid's individual cells, resulting in a new dataset where rows represent the grid cells and columns are the land types.

Similarly, for the Aspect Map information, initially covering the whole Europe, we performed a spatial intersection with the geographical extent of the France territory. The spatial resolution of this data source is of 25 m. A value ranging from 0 to 255 was associated with each grid cell. These values corresponded to land orientation, which we converted into cardinal and intercardinal directions. Using the same approach as with the Corine Land Cover, we incorporated the land orientation data into the *X. fastidiosa* grid structure. Like the Corine Land Cover, the output is a dataset with row corresponding to grid's cell and columns to the cardinal and intercardinal directions.

For altitude information, we performed a slightly different method as we aimed to obtain the mean altitude rather than the proportion of altitude. The original resolution is 75 m × 75 m. We computed the mean altitude per grid cell to acquire information at the grid level.

The *X. fastidiosa* information did not require transformation in either projection or resolution as we used them as the target projection and resolution. The only transformation performed, is on the type of data contained at the cell level; we simplified the level of information by considering binary data (1 if at least one plant tested positive, 0 if all plants tested negative, NA if no plants were tested). It should be noted that there is uncertainty associated with the binary projection. Specifically, cells classified as not infected have not been tested across their entire surface. However, we trust the experts to have sampled the plants most likely to be infected.

The outcome of these processes is a spatial dataset with the multiple information detailed above at the *X. fastidiosa*'s grid resolution.

4.2. Missing data imputation

Some data, especially the one related to chemical elements in the soil had a large amount of missing values, which is due to the characteristics of the variables. Indeed, the samples are made at three depth levels: holorganic layer, surface layer, and subsurface layer. Some of these depth layers did not have enough data to be treated and some have less than 50% of value. We arbitrarily decided to get rid of the data with more than 20% of missing values. The administrative department code of data with missing values was treated to retrieve the code using the majority code of their neighbor's department code. Once all the department codes were available, they were used to impute, department per department, the remaining missing values.

5. Conclusion

In the spirit of open-source and open-science data movement, we are sharing the dataset we presented here along with the related codes as a resource for the community. We gathered data from multiple sources related to the environment and soil characteristics, and performed preprocessing related to spatial projection and spatial resolution in order to integrate them into a unified source. The output of this contribution is twofold: a dataset that can be merged with either the *X. fastidiosa*'s dataset or with other pathogens information and an open-source code that can foster reproducibility of the performed preprocessing stages. The code can also be used as an example for upscaling or downscaling another data source that can be added to the provided dataset.

Some empirical choices have been made depending on the context of our work, as we are providing the code with the data, one can decide to make another choice in the aggregation of the data. These choices include the upscaling method for the Aspect map and Corine Land Cover data, the downscaling method for the climatic and chemical elements data, and in the cleaning process and the removal of incomplete variables. For the upscaling method, we chose to take the proportion of the cell occupied by each variables, which keeps the same amount of information but in a smaller resolution instead of taking the majority class. The advantage of this method is that it maintains the same level of information as the original data, but it

https://doi.org/10.1017/eds.2024.50 Published online by Cambridge University Press

creates a sparse matrix, which could complicate some computations. For the downscaling method, we simply allocate the majority value into the cells. This method is quite straightforward and does not introduce artifacts, but it creates a coarse representation without gradual boundaries.

Open peer review. To view the open peer review materials for this article, please visit http://doi.org/10.1017/eds.2024.50.

Acknowledgments. The authors are grateful for the technical assistance of Loïc Houde for the management of the cluster, which hosted the computation and Davide Martinetti for his advices on the use of **terra** package.

Author contribution. Conceptualization: C.P.; Methodology: C.P., E.G., and D.I.; Data curation: C.P.; Data visualization: C.P.; Writing original draft: C.P., E.G., D.I., and E.V. All authors approved the final submitted draft.

Data availability statement. The dataset Portes et al. (2024) presented here can be found at https://doi.org/10.57745/P7XUII and the code here https://archive.softwareheritage.org/browse/directory/fed68145c4649d1b754ca14f7186f031f263c59c/?origin_url=https://gitlab.paca.inrae.fr/uapv2204620/preparation&revision=f7af17601462b85dee7faadbd0d86d3e0e0d6443&snapshot=dd8bd3901430dd49718754a7b85def2be44a9e73. However, due to confidential restriction related to the *X. fastidiosa* dataset ESV, 2022, researchers must request access to them from "the ESV platform" at https://doi.org/10.15454/RWBIWD and use the script in the Gitlab repository to facilitate data manipulation and make the correspondence with the data we are providing.

Funding statement. This work was supported by a French government grant managed by the Agence Nationale de la Recherche under the "Investissements d'avenir" program, reference ANR-18-EURE-0009 and the BEYOND project, reference ANR-20-PCPA-0002.

Competing interest. The authors declare no competing interests exist.

Ethics statement. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

References

Auslander M, **Nevo E and Inbar M** (2003) The effects of slope orientation on plant growth, developmental instability and susceptibility to herbivores. *Journal of Arid Environments 55*(3), 405–416. https://doi.org/10.1016/S0140-1963(02)00281-1

Contours-régions (2018) Contours des régions françaises sur openstreetmap. Retrieved from https://www.data.gouv.fr/fr/datasets/contours-des-regions-francaises-sur-openstreetmap/ (accessed 1 March 2022).

Copernicus AM (2019) © European union, copernicus land monitoring service 2022, European Environment Agency (EEA), Online. Retrieved from https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1-0-and-derived-products/aspect?tab=metadata (accessed 20 February 2022).

Copernicus CLC (2020) Corine Land Cover (CLC) 2018, version 2020–2021, Online. Retrieved from https://land.copernicus.eu/pan-european/corine-land-cover/clc2018?tab=metadata (accessed 20 February 2022).

D'Attoma G, **Morelli M**, **Saldarelli P**, **Saponari M**, **Giampetruzzi A**, **Boscia D**, **Savino VN**, **De La Fuente L and Cobine PA** (2019) Ionomic differences between susceptible and resistant olive cultivars infected by xylella fastidiosa in the outbreak area of Salento, Italy. *Pathogens 8*(4), 272. https://doi.org/10.3390/pathogens8040272

Del Coco L, **Migoni D**, **Girelli CR**, **Angilè F**, **Scortichini M and Fanizzi FP** (2020) Soil and leaf ionome heterogeneity in X*ylella fastidiosa* Subsp. *Pauca*-infected, non-infected and treated olive groves in Apulia, Italy. *Plants 9*(6). https://doi.org/10.3390/plants9060760

European Food Safety Authority (EFSA), **Delbianco A**, **Gibin D**, **Pasinato L**, **Boscia D, and Morelli M** (2022) Update of the xylella spp. host plant database – Systematic literature search up to 31 December 2021. *EFSA Journal 20*(6), e07356. https://doi.org/10.2903/j.efsa.2022.7356

Farigoule P, **Chartois M**, **Mesmin X**, **Lambert M**, **Rossi J.-P.**, **Rasplus J.-Y., and Cruaud A** (2022). Vectors as sentinels: Rising temperatures increase the risk of xylella fastidiosa outbreaks. *Biology 11*(9). https://doi.org/10.3390/biology11091299

Ge Q, **Cobine PA and De La Fuente L** (2020) Copper supplementation in watering solution reaches the xylem but does not protect tobacco plants against xylella fastidiosa infection. *Plant Disease 104*(3), 724–730. https://doi.org/10.1094/PDIS-08-19-1748-RE

InfoSol I (2015) Cartogrammes des teneurs ponctuelles en manganèse total mesurées sur le rmqs, campagne 1 (2001–2009) & cartogrammes des teneurs ponctuelles en zinc total mesurées sur le rmqs, campagne 1 (2001–2009). Online. Retrieved from https://agroenvgeo.data.inra.fr/geonetwork/srv/fre/catalog.search#/metadata/dae648df-6e5d-5f13-b57c-b79644c6b1ed and https://agroenvgeo.data.inra.fr/geonetwork/srv/fre/catalog.search#/metadata/20fbc45b-6dda-5ea1-bcbc-9147d4a6dd3a (accessed 10 February 2022).

Ingwell LL, **Lacroix C**, **Rhoades PR**, **Karasev AV and Bosque-Pérez NA** (2017) Agroecological and environmental factors influence barley yellow dwarf viruses in grasslands in the US pacific northwest. *Virus Research 241*, 185–195. https://doi.org/10.1016/j.virusres.2017.04.010

Lemaire J, **Vennetier M**, **Prévosto B and Cailleret M** (2022) Interactive effects of abiotic factors and biotic agents on scots pine dieback: A multivariate modeling approach in Southeast France. *Forest Ecology and Management 526*, 120543. https://doi.org/10.1016/j.foreco.2022.120543

Martinetti D and Soubeyrand S (2019) Identifying lookouts for epidemio-surveillance: Application to the emergence of xylella fastidiosa in France. *Phytopathology 109*(2), 265–276. https://doi.org/10.1094/PHYTO-07-18-0237-FI

Météo-France (2020) Siclima. Retrieved from https://agroclim.inrae.fr/siclima/ (accessed 01 March 2022).

O'Donnel MS, and Ignizio DA (2012) Bioclimatic predictors for supporting ecological applications in the conterminous United States. *Data Series 691*. https://doi.org/10.3133/ds691

Ojiambo PS, **Yuen J**, **van den Bosch F and Madden LV** (2017) Epidemiology: Past, present, and future impacts on understanding disease dynamics and improving plant disease management—A summary of focus issue articles. *Phytopathology 107* (10), 1092–1094. https://doi.org/10.1094/PHYTO-07-17-0248-FI

Plateforme ESV Dataverse (2022) Données de surveillance sur végétaux de Xylella fastidiosa. https://doi.org/10.15454/RWBIWD

Portes C, **Gabriel E and Ienco D** (2024) Environmental and Bio-climatic Data over French Mediterranean areas. https://doi.org/10.57745/P7XUII

Préfectures-régions (2015) Geographie de la corse. Retrieved from https://www.prefectures-regions.gouv.fr/corse/Region-et-institutions/Portrait-de-la-region/Geographie/Geographie-de-la-Corse (accessed 18 June 2024).

Roubal C, **Regis S and Nicot P** (2013) Field models for the prediction of leaf infection and latent period of *Fusicladium oleagineum* on olive based on rain, temperature and relative humidity. *Plant Pathology 62*(3), 657–666. https://doi.org/10.1111/j.1365-3059.2012.02666.x

Siges-Occitanie (2022) Climatologie en occitanie. Retrieved from https://sigesocc.brgm.fr/spip.php?article37 (accessed 18 June 2024).

Techno-Science (2013) Géographie de la région paca. Retrieved from https://www.techno-science.net/glossaire-definition/Geographie-de-la-region-Provence-Alpes-Cote-d-Azur.html (accessed 19 June 2024).

Tile-Grabber SRTM (2013) Shuttle radar topography mission tile grabber. Retrieved from http://dwtkns.com/srtm/ (accessed 22 January 2024).

Trkulja V, **Tomić A**, **Iličić R**, **Nožinić M and Milovanović TP** (2022) Xylella fastidiosa in Europe: From the introduction to the current status. *Plant Pathology Journal 38*(6), 551–571. https://doi.org/10.5423/PPJ.RW.09.2022.0127

A. Appendix. Data dictionary

Table A1. *Data dictionary*

Name of variables	Definition
Administrative division	From ESV (2022) and Contours-régions (2018)
department	*Number of the administrative department*
region	*Name of the administrative region*
Chemical elements	From InfoSol (2015)
clay	*Clay content*
fine_silt	*Fine silt content*
coarse_silt	*Coarse silt content*
fine_sand	*Fine sand content*
coarse_sand	*Coarse sand content*
residual_water	*Residual water content*
water_ph	*Water pH*
exchan_aluminum	*Exchangeable aluminum*
total_aluminum	*Total aluminum*
boron_soluble	*Boron soluble in boiling water*
organic_carbon	*Organic carbon*
exchan_calcium	*Exchangeable calcium*
total_calcium	*Total calcium*
total_limestone	*Total limestone*
cation_echange_capacity	*Cation exchange capacity (CEC)*
exchan_iron	*Exchangeable iron*

Continued

https://doi.org/10.1017/eds.2024.50 Published online by Cambridge University Press

Table A1. *Continued*

Name of variables	Definition
free_iron	*Free iron*
free_iron_b	*Free iron*
total_iron	*Total iron*
exchan_potassium	*Exchangeable potassium*
total_potassium	*Total potassium*
organic_matter	*Organic matter*
exchan_magnesium	*Exchangeable magnesium*
total_magnesium	*Total magnesium*
exchan_manganese	*Exchangeable manganese*
total_manganese	*Total manganese*
total_nitogen	*Total nitrogen*
exchan_sodium	*Exchangeable sodium*
total_sodium	*Total sodium*
assimilable_phosphorus	*Assimilable phosphorus*
total_arsenic	*Total arsenic*
extract_cadium	*Extractable cadmium*
total_cadium	*Total cadmium*
total_cobalt	*Total cobalt*
extract_chromium	*Extractable chromium*
extract_copper	*Extractable copper*
total_copper	*Total copper*
total_mercury	*Total mercury*
total_molybdenum	*Total molybdenum*
extract_nickel	*Extractable nickel*
total_nickel	*Total nickel*
extract_lead	*Extractable lead*
total_lead	*Total lead*
total_thallium	*Total thallium*
extract_zinc	*Extractable zinc*
total_zinc	*Total zinc*
Relief	From Tile-Grabber (2013)
altitude	*Mean altitude of the cell*
Climatic variables	From Météo-France (2020)
drainage	*Drainage in mm*
potential_evapo_p_m	*Potential evapotranspiration (formula of Penman-Monteith)*
potential_evapo_siclima	*Potential evapotranspiration computed by SICLIMA*
real_evapo	*Real evapotranspiration*
relative_humidity	*Relative humidity*
specific_humidity	*Specific humidity*
soil_humidity	*Soil humidity index*
effective_rains	*Effective rains*
liquid_rains	*Liquid rains*
solid_rains	*Solid rains*
atmospheric_radiation	*Atmospheric radiation in* J/cm^2
visible_radiation	*Visible radiation in* J/cm^2
runoff	*Runoff in mm*

Continued

https://doi.org/10.1017/eds.2024.50 Published online by Cambridge University Press

Table A1. *Continued*

Name of variables	Definition
temp	*Mean temperature of 24 hours temperatures in* °C
wind	*Wind* ± 10 m *in* m/s
max_temp	*Maximum temperature*
min_temp	*Minimum temperature*
difference_max_min_temp	*Difference between maximum and minimum temperature*
Bioclimatic variable	Using the formulas from O'Donnel and Ignizio (2012) and the data from Météo-France (2020)
BIO1	*Annual mean temperature*
BIO2	*Mean Diurnal range (mean of monthly (max temp − min temp))*
BIO3	*Isothermality (BIO2/BIO7) (×100)*
BIO4	*Temperature seasonality (standard deviation ×100)*
BIO5	*Max temperature of warmest month*
BIO6	*Min temperature of coldest month*
BIO7	*Temperature annual range (BIO5–BIO6)*
BIO8	*Mean temperature of wettest quarter*
BIO9	*Mean temperature of Driest quarter*
BIO10	*Mean temperature of warmest quarter*
BIO11	*Mean temperature of coldest quarter*
BIO12	*Annual precipitation*
BIO13	*Precipitation of wettest month*
BIO14	*Precipitation of driest month*
BIO15	*Precipitation seasonality (coefficient of variation)*
BIO16	*Precipitation of wettest quarter*
BIO17	*Precipitation of driest quarter*
BIO18	*Precipitation of warmest quarter*
BIO19	*Precipitation of coldest quarter*
Type of land	Proportion of the cell corresponding to the culture from C. L. C. Copernicus (2020)
anual_permanent_crops	*Annual crops associated with permanent crops*
agro_forestry	*Agro-forestry areas*
airports	*Airports*
bare_rocks	*Bare rocks*
beaches_dunes_sand	*Beaches dunes sands*
broad_leaved_forest	*Broad-leaved forest*
burnt_areas	*Burnt areas*
coastal_lagoons	*Coastal lagoons*
complex_cultivation_pattern	*Complex cultivation patterns*
coniferus_forest	*Coniferous forest*
construction_sites	*Construction sites*
continuous_urban	*Continuous urban fabric*
discontinuous_urban	*Discontinuous urban fabric*
dump_sites	*Dump sites*
estuaries	*Dump sites*
fruit_berries_plantation	*Fruit trees and berry plantations*
glacier_perpetual_snow	*Glaciers and perpetual snow*
green_urban	*Green urban areas*

Continued

https://doi.org/10.1017/eds.2024.50 Published online by Cambridge University Press

Table A1. *Continued*

Name of variables	Definition
industrial_areas	*Industrial or commercial units*
inland_marshes	*Inland marshes*
agriculture_with_vegetation	*Land principally occupied by agriculture with significant areas of natural vegetation*
mineral_extraction	*Mineral extraction sites*
mixed_forest	*Mixed forest*
moors_heatland	*Moors and heathland*
natural_grasslands	*Natural grasslands*
non_irrigated_arable	*Nonirrigated arable land*
olive_groves	*Olive groves*
pastures	*Pastures*
irrigated_land	*Permanently irrigated land*
peat_bogs	*Peat bogs*
port_areas	*Port areas*
rice_fields	*Rice fields*
road_rail	*Road and rail networks and associated land*
salines	*Salines*
salt_marshes	*Salt marshes*
sclerophyllous_vege	*Sclerophyllous vegetation*
sea_oceans	*Sea and ocean*
sparsely_vegetated_areas	*Sparsely vegetated areas*
sport_facilities	*Sport and leisure facilities*
transitional_woodland_shrub	*Transitional woodland-shrub*
vineyards	*Vineyards*
water_bodies	*Water bodies*
water_courses	*Water courses*
Orientation	Proportion of the cell corresponding to the cardinal and intercardinal directions from A. M. Copernicus (2019)
N	*North*
NE	*North-East*
E	*East*
SE	*South-East*
S	*South*
SW	*South-West*
W	*West*
NW	*North-West*

Cite this article: Portes C, Ienco D, Verdin E and Gabriel E (2024). Environmental and bioclimatic data for epidemiological analysis over French Mediterranean areas. *Environmental Data Science*, 3: e31. doi:10.1017/eds.2024.50

https://doi.org/10.1017/eds.2024.50 Published online by Cambridge University Press