Data in brief 27 (2019) 104810

Contents lists available at ScienceDirect

Data in brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Time-series spectral dataset for croplands in France (2006–2017)

Laurence Hubert-Moy*, Jeanne Thibault, Elodie Fabre, Clémence Rozo, Damien Arvor, Thomas Corpetti, Sébastien Rapinel

University of Rennes, LETG, UMR 6554 CNRS, place du recteur Henri Le Moal, 35000 Rennes, France

ARTICLE INFO

Article history:
Received 14 October 2019
Received in revised form 5 November 2019
Accepted 7 November 2019
Available online 15 November 2019

Keywords:
Crop modeling
MODIS
Vegetation index
Time-series analysis
Big data
LPIS

ABSTRACT

Decadal time-series derived from satellite observations are useful for discriminating crops and identifying crop succession at national and regional scales. However, use of these data for crop modeling is challenged by the presence of mixed pixels due to the coarse spatial resolution of these data, which influences model accuracy, and the scarcity of field data over the decadal period necessary to calibrate and validate the model. For this data article, cloud-free satellite "Vegetation Indices 16-Day Global 250 m" Terra (MOD13Q1) and Aqua (MYD13Q1) products derived from the Moderate Resolution Imaging Spectroradiometer (MODIS), as well as the Land Parcel Information System (LPIS) vector field data, were collected throughout France for the 12-year period from 2006 to the end of 2017. A GIS workflow was developed using R software to combine the MOD13Q1 and MYD13Q1 products, and then to select "pure" MODIS pixels located within single-crop parcels over the entire period. As a result, a dataset for 21,129 reference plots (corresponding to "pure" pixels) was generated that contained a spectral time-series (red band, near-infrared band, Normalized Difference Vegetation Index (NDVI), and Enhanced Vegetation Index (EVI)) and the associated annual crop type with an 8-day time step over the period. This dataset can be used to develop new classification methods based on time-series

* Corresponding author.
E-mail address: laurence.hubert@univ-rennes2.fr (L. Hubert-Moy).

https://doi.org/10.1016/j.dib.2019.104810

analysis using deep learning, and to monitor and predict crop succession.

Specifications Table

Subject	Agricultural and Biological Sciences (General)
Specific subject area	Applied remote sensing; crop succession; deep learning.
Type of data	Table and spatial dataset
How data were acquired	Spectral values were generated from MODIS satellite observations; LPIS field data were acquired from farmers' declarations.
Data format	Raw and analyzed.
Parameters for data collection	MODIS pixels strictly included within single-crop parcels from 2006 to the end of 2017.
Description of data collection	Crop type, selected among 19 categories, was assigned to reference plots corresponding to "pure" pixels each year from 2006 to the end of 2017 using the LPIS database. Spectral values (red, near infrared, Normalized Difference Vegetation Index (NDVI), and Enhanced Vegetation Index (EVI)) were collected every 8 days from 2006 to 2017 using MOD13Q1 and MYD13Q1 satellite products.
Data source location	France
Data accessibility	With the article.

Value of the Data

- The dataset, which contains 8-day-interval spectral values and annual crop type (selected among 19 categories) for 21,129 reference plots throughout France for 12 years, provides comprehensive reference information about crop successions.
- The dataset can be used to discriminate crop types for monitoring and predicting crop succession.
- These data may help develop new modeling methods based on long-term time-series analysis using deep learning.

1. Data

The dataset includes a vector GIS shapefile (WGS 84 projection, EPSG 4326) with 21,129 points distributed throughout France (Fig. 1). To each point, representing the centroid of a "pure" MODIS pixel (NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC)), five attribute tables were assigned: one for the annual crop type and four for spectral values in the red, near-infrared, NDVI, and EVI bands in 8-day intervals from 2006 to the end of 2017. The last four table columns are labeled "SpectralBand_Year_DayOfYear". An overview of the spectral time-series dataset including the map of the reference plots and an example of the NDVI profile for one reference plot with the associated crop types is presented Fig. 1.

2. Experimental design, materials, and methods

The dataset was produced using spectral values from MODIS satellite time-series and crop types from the Land Parcel Information System (LPIS). Specifically, the MODIS "Vegetation Indices 16-Day Global 250 m" Terra (MOD13Q1) and Aqua (MYD13Q1) [1] products were collected in WGS84 spatial projection for the whole of France from 2006 to the end of 2017 (by vegetative year, i.e. from November 1 of year n to October 30 of year $n+1$) using the Application for Extracting and Exploring and Extracting Analysis Ready Samples platform (AρρEEARS) [2]. These raster products were already atmospherically and geometrically corrected, each pixel value corresponding to the best value

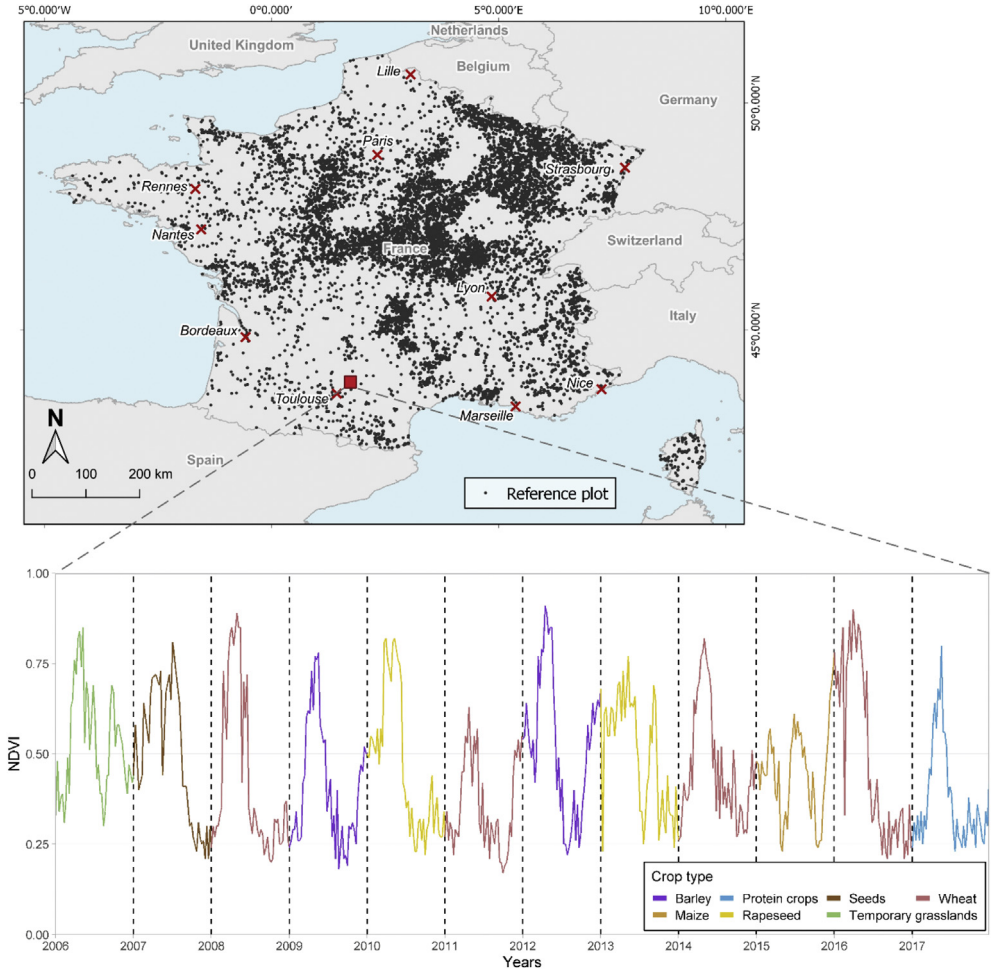

Fig. 1. Overview of the spectral time-series dataset: (top) map of the 21,129 reference plots and (bottom) example of the NDVI profile for one reference plot with the associated crop types.

observed over 16 days [3]. Nonetheless, some pixels may have missing values (na) for certain dates, especially in winter, when atmospheric and light conditions change. The values in the red and near-infrared bands indicate top-of-atmosphere reflectance, while the spectral values of NDVI and EVI range from −1 to 1. The LPIS field data describe the annual crop type for each parcel block (adjacent parcels managed by the same farmer) among 19 categories based on farmers' declarations. The LPIS data for France are available in vector format (www.data.gouv.fr) and were collected each calendar year from 2006 to the end of 2017.

A GIS workflow was developed using R software [4] (Fig. 2). For each year, an attribute query was applied to the LPIS reference data to select only parcel blocks that were covered at least 80% by the same crop type. The other parcel blocks were discarded. In a parallel step, the MOD13Q1 and MYD13Q1 products of each year were combined to generate an annual raster time-series with an 8-day interval. Then, for each year, LPIS data were overlaid on MODIS data to select only “pure” pixels, i.e. strictly included within the single-crop parcel blocks. The centroid of each pure pixel was converted into a vector reference plot with five attribute tables that listed the crop type and spectral values during each

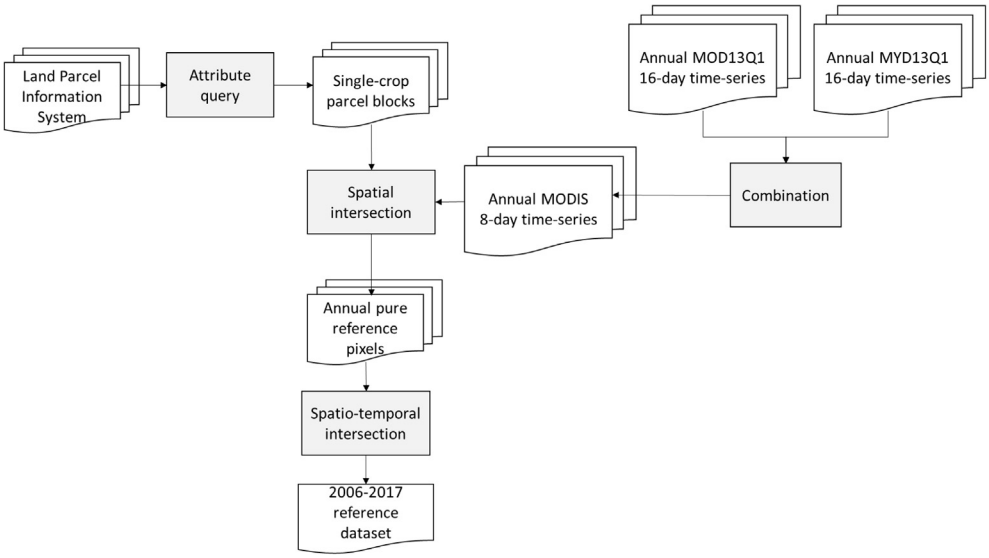

Fig. 2. Methodological workflow for data processing using R software.

year. Reference plots were compiled over the 12-year period, and only those existing for the entire period were retained.

Acknowledgments

This research was funded by the French Ministry of Ecology (grant no. 2101606295). We thank the suppliers for freely providing their data. MODIS images were retrieved from https://lpdaac.usgs.gov, maintained by the NASA EOSDIS LP DAAC at the USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, USA. LPIS data were provided by the French Institut Géographique National.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2019.104810.

References

[1] K. Didan, MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid, NASA EOSDIS LP DAAC, 2015, https://doi.org/10.5067/MODIS/MOD13A1.006. Version 6.
[2] S. Neeley, Analyzing earth data with NASA's AppEEARS tool to improve research efficiency, in: AGU Fall Meeting Abstracts, Washington DC, 2018.
[3] K. Didan, A. Barreto Munoz, R. Solano, A. Huete, MODIS vegetation index user's guide (MOD13 series). https://vip.arizona.edu/documents/MODIS/MODIS_VI_UsersGuide_June_2015_C6.pdf, 2015. (Accessed 5 November 2019).
[4] R. R Core Team, A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2015. http://www.R-project.org/.