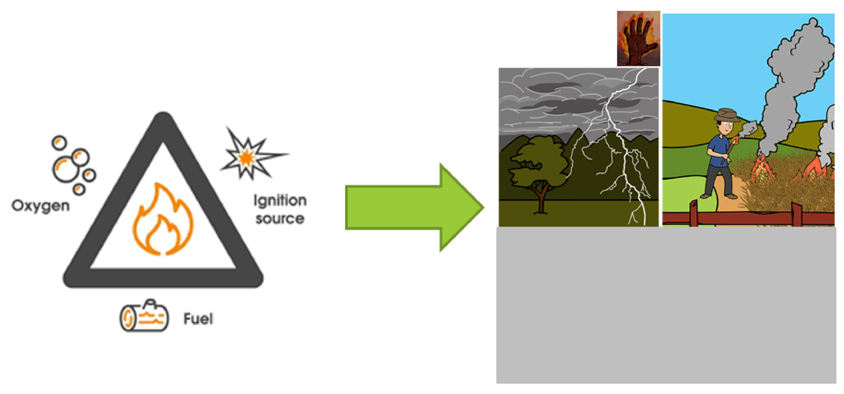
Modelling wildfire ignition causalities using a Machine Leaning approach.

Clara Ochoa, Magí Franquesa, Marcos Rodrigues, Emilio Chuvieco

mclara.ochoa@uah.es

1.Introduccion:

Wildfire ignition requires three fundamental elements—an ignition source, fuel, and oxygen—collectively known as the Fire Triangle. Fires are typically classified as natural (mainly lightning), human-induced (intentional or negligent), or as reignitions of previous fires. However, data gaps remain a major limitation: in Europe, between 40–70% of wildfires are recorded with unknown causes (De Rigo et al., 2017; Prométhée, 2020), hindering accurate statistics and effective prevention strategies. Understanding ignition sources is crucial, as they vary spatially and temporally depending on environmental and human drivers (Curt et al., 2016). This study aims to develop a Random Forest machine learning model to classify wildfire ignition sources of unknown origin in Peninsular Spain and to evaluate the influence of environmental and anthropogenic factors on ignition prediction, contributing to improved fire management and prevention (Tedim et al., 2022).



2. Methods:

2.1 Datasets

EGIF

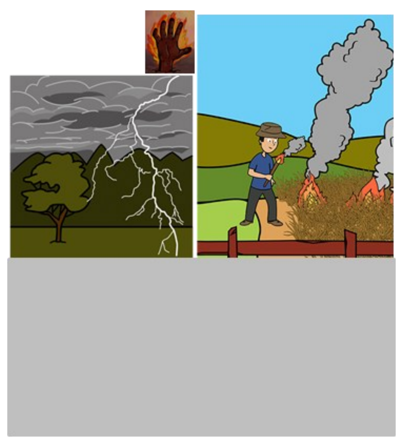
- National database of all forest fires in Spain, reported by regional authorities.
- Attributes: date, cause, burned area and "location".
- 1968-2021
- Pre- 2012 spatial information based on outdated referencing
- Old reference System: locations based on 1:250,000 military map sheets and 10 x 10 km grids.

IberFire

- National scale burned area polygons
- Derived from Landsat.
- 30 m
- 1985-2023
- ≥ 5 ha

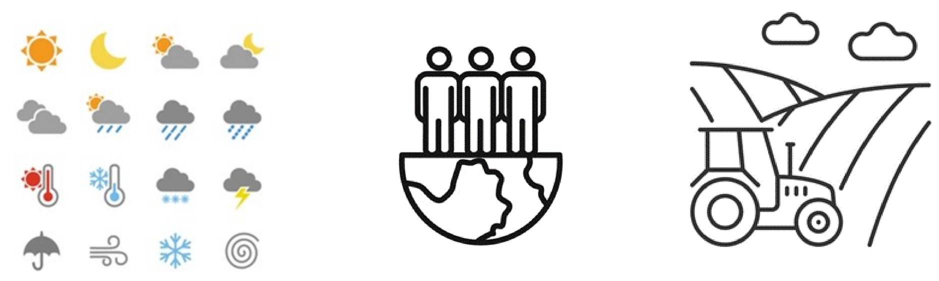
2.2 Dependent Variable:

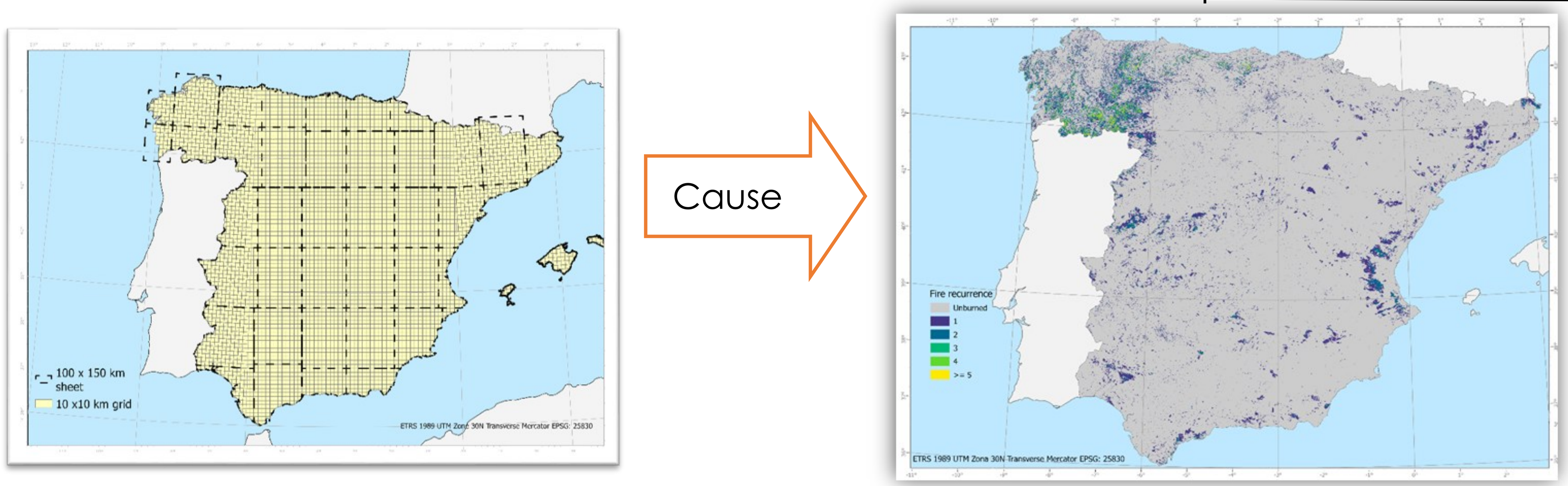
- **Lightning (Natural)** – 2,966 fires
- **Negligence** – 11,809 fires
- **Intentional** – 50,344 fires
- **Unknown –** 54,167 fires
- **Restart of Fire** – 212 fires



2.3 Independent Variables:

- **Demographic:** Population & rural density (WorldPop, Pesaresi et al.)
- **Infrastructure:** Road density (IGN 2024)
- **Land-use Interfaces**: WUI, FAI, WAI, WFI, FUI (CNIG, CORINE, SIOSE)
- **Climatic:** Temperature, humidity, wind, precipitation, evapotranspiration, etc. (1985–2021)
- **Vegetation:** NDVI, SNDVI (Franquesa et al., 2024)
- **Drought Indicators:** SPEI-3 and SPEI-6 (Beguería et al., 2023)
- **Agriculture:** Grazing density (Gilbert et al., 2018)
- **Land Change:** UNFCCC, 2025 dataset (25m resolution)





2.4 Model:

-**Random Forest Classifier** – Predicts fire cause and identifies key explanatory variables.

-**Cross-validation** and Grid Search (5-fold) used to optimize hyperparameters and prevent overfitting

Data Preparation & Balancing

- **Correlation matrix**
- **SMOTE** (Synthetic Minority Oversampling Technique)

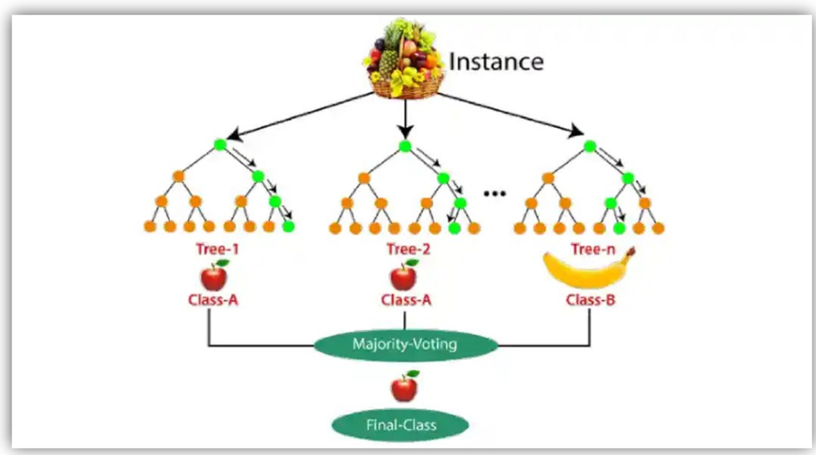
Model Interpretation

- **SHAP** (Shapley Additive Explanations) used to interpret variable influence

- Quantifies each variable's contribution to fire cause classification

-Provides transparency and interpretability for model outputs

- **Spatial SHAP** analysis performed to detect geographic variability in drivers

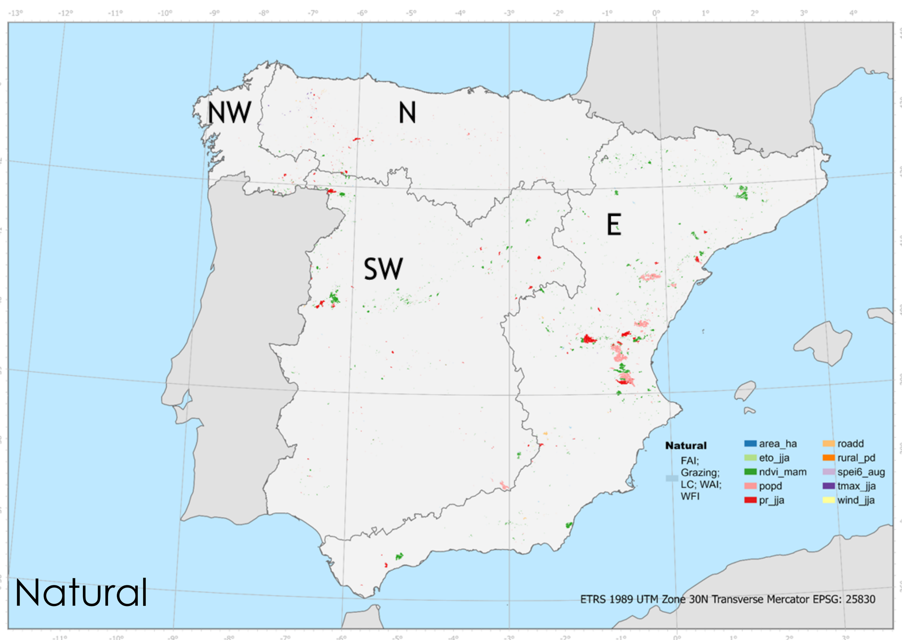


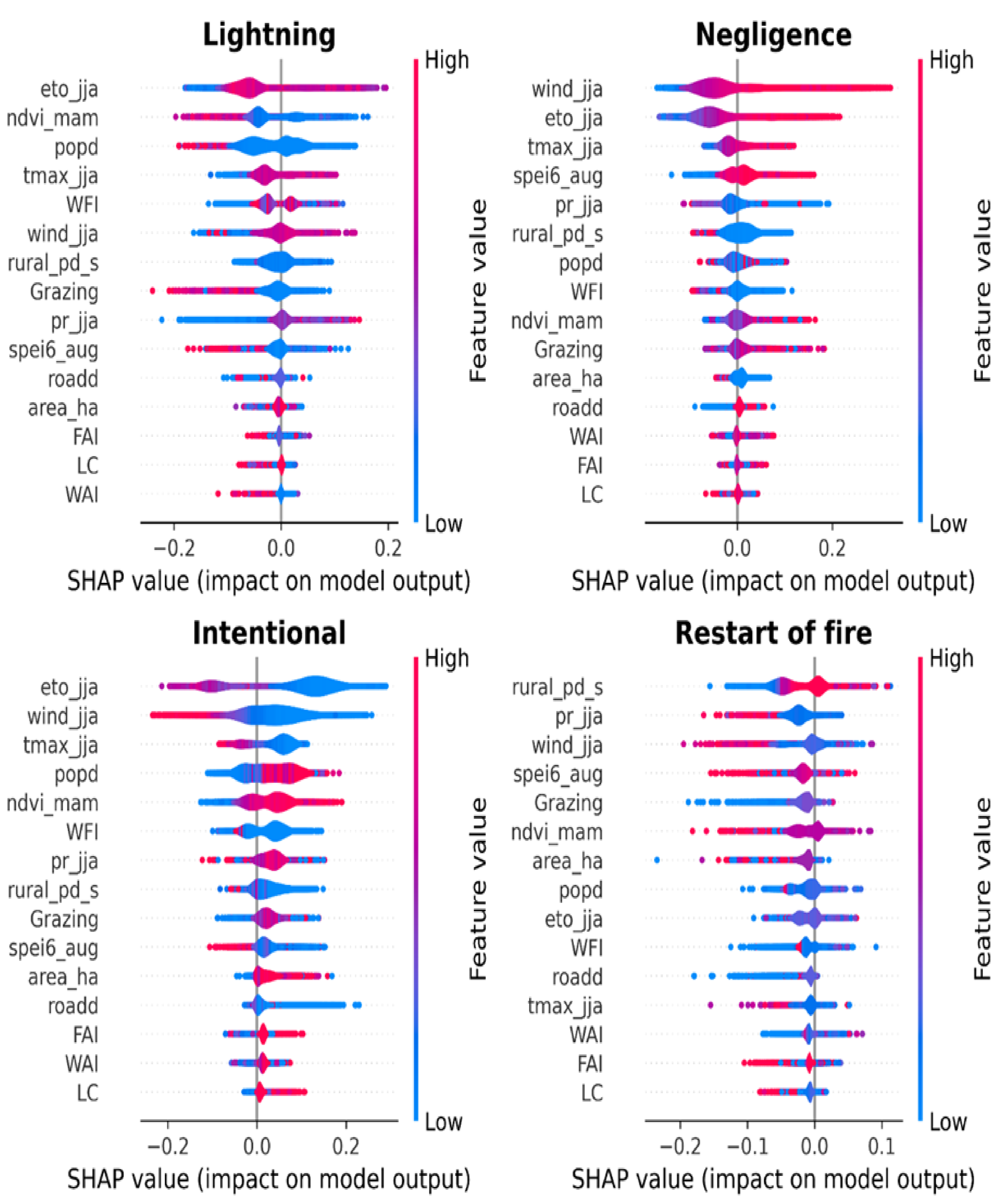
3. Results:

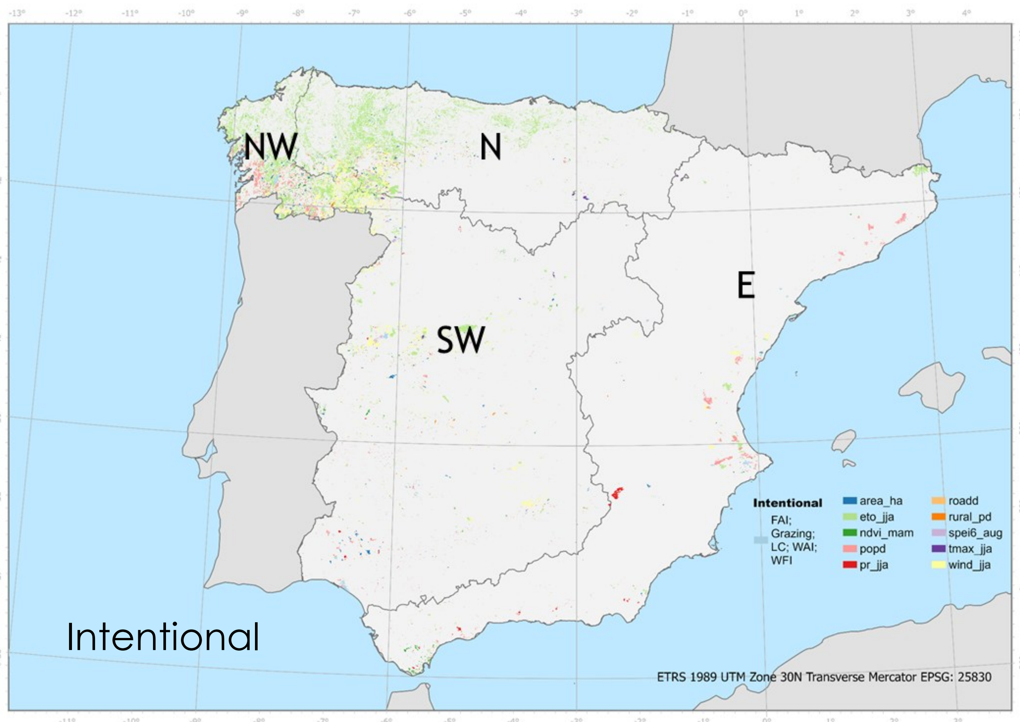
3.1 Accuracy

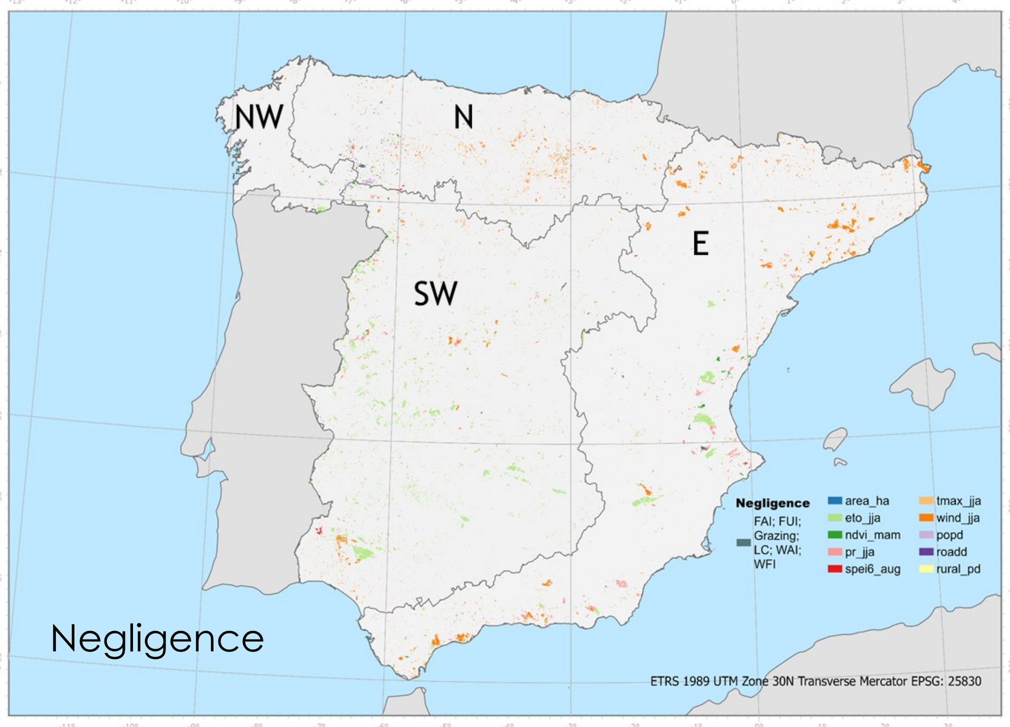
Class	Precision	Recall	F1-Score
Lightning	0.49	0.39	0.43
Negligence	0.57	0.65	0.61
Intentional	0.91	0.89	0.90
Restart of fire	0.15	0.03	0.05
Accuracy global			0.82
F1-score global			0.82

3.2 Global & local Importance









4. Conslusions:

- The Random Forest model, combined with the SHAP interpretability framework, effectively classified wildfire ignition causes in Spain, capturing both environmental and anthropogenic influences.
- Human-related fires (intentional and negligence) displayed contrasting patterns: negligence ignitions occurred under hot, dry, and windy conditions, while intentional fires were associated with cooler, humid, and more densely populated areas.
- Natural (lightning) ignitions were mainly linked to vegetation activity, precipitation, and evapotranspiration, particularly in less populated regions.
- Spatial SHAP analysis revealed strong regional variability, emphasizing the role of the variables in fire ignition dynamics across Spain.
- These findings highlight the importance of incorporating spatially explicit and explainable ML models to improve fire cause prediction and support more targeted prevention strategies as high-quality datasets continue to expand.

5. Reference

(CNIG), C. N. d. I. G. (2005-2020). *Sistema de Información sobre la Ocupación del Suelo en España (SIOSE): series 2005–2020*. https://www.siose.es

Acebes, P., Iglesias-González, Z., & Muñoz-Galvez, F. J. (2021). Do traditional livestock systems fit into contemporary landscapes? Integrating social perceptions and values on landscape change. *Agriculture*, *11*(11), 1107.

Agency, E. E. (1990-2018). *CORINE Land Cover (CLC) Dataset Series: 1990, 2000, 2006, 2012, 2018*. https://land.copernicus.eu/pan-european/corine-land-cover

Beguería, S., Vicente Serrano, S. M., Reig-Gracia, F., & Latorre Garcés, B. (2023). SPEIbase v. 2.9 [Dataset].

Bountzouklis, C., Fox, D. M., & Bernardino, E. D. (2023). Predicting wildfire ignition causes in Southern France using eXplainable Artificial Intelligence (XAI) methods. *Environmental Research Letters*, *18*(4). https://doi.org/10.1088/1748-9326/acc8ee

Costafreda-Aumedes, S., Comas, C., & Vega-Garcia, C. (2017). Human-caused fire occurrence modelling in perspective: a review. *International Journal of Wildland Fire*, *26*(12), 983. https://doi.org/10.1071/wf17026

Franquesa, M., Reig-Gracia, F., & Vicente Serrano, S. M. (2024). Vegetation Indices for the Iberian Peninsula and Balearic Islands (VIIB) Database.

Gilbert, M., Nicolas, G., Cinardi, G., Van Boeckel, T. P., Vanwambeke, S. O., Wint, G., & Robinson, T. P. (2018). Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Scientific data*, *5*(1), 1-11.

IGN. (2024). *Redes de Transporte del CNIG*. https://centrodedescargas.cnig.es/CentroDescargas/redes-transporte

López Escolano, C., Pueyo Campos, Á., Postigo Vidal, R., & Alonso Logroño, M. P. (2016). Valoración y representación cartográfica de la accesibilidad viaria en la España peninsular: 1960-2014.

Ministerio de Agricultura, P. y. A. (2020). *Demografía en la población rural en España* (Análisis y Prospectiva Issue. P. y. A. Ministerio de Agricultura. https://www.mapa.gob.es/dam/mapa/contenido/ministerio/servicios/servicios-de-informacion/analisis-y-prospectiva/ayp-serie-agrinfo/ayp_demografiaenlapoblacionrural2020.pdf

Ochoa, C., Bar-Massada, A., & Chuvieco, E. (2024). A European-scale analysis reveals the complex roles of anthropogenic and climatic factors in driving the initiation of large wildfires. *Science of the total environment*, 170443.

Pesaresi, M., Schiavina, M., Politis, P., Freire, S., Krasnodębska, K., Uhl, J. H., Carioli, A., Corbane, C., Dijkstra, L., & Florio, P. (2024). Advances on the Global Human Settlement Layer by joint assessment of Earth Observation and population survey data. *International Journal of Digital Earth*, *17*(1), 2390454.

Quan, Y., Zhong, X., Feng, W., Chan, J. C.-W., Li, Q., & Xing, M. (2021). SMOTE-based weighted deep rotation forest for the imbalanced hyperspectral data classification. *Remote Sensing*, *13*(3), 464.