Check for updates

RESEARCH ARTICLE

Reconstructing Italy's rural landscape before the Great Acceleration: A geospatial baseline from the *Catasto Agrario* (1929)

[version 1; peer review: 1 approved, 3 approved with reservations]

Filippo Brandolini[1,2]

[1]Center for Sustainability Science and Strategy, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, USA
[2]Università degli Studi di Milano Dipartimento di Scienze della Terra Ardito Desio, Milan, 20133, Italy

v1 **First published:** 27 Jun 2025, **5**:169
https://doi.org/10.12688/openreseurope.20343.1
Latest published: 27 Jun 2025, **5**:169
https://doi.org/10.12688/openreseurope.20343.1

Abstract

This paper presents a geospatial dataset detailing the rural landscape of Italy in the 1920s, based on the Catasto Agrario 1929 survey. The dataset integrates data from the survey into a Geographic Information System (GIS), providing insights into land use and land cover (LULC), demographic characteristics, livestock distribution, crop yields, and precipitation patterns. Historical data have been digitised using Optical Character Recognition (OCR) and organised into a vector format, capturing the administrative boundaries of Italy's provinces as they were in the 1920s. By documenting Italy's rural landscape just before the onset of the Great Acceleration (ca. mid-20th century CE) the dataset offers a critical historical baseline for analysing long-term socio-environmental transformations. The research aims to facilitate future studies on the environmental impacts of Italy's rural transitions, offering an open-access resource that enables comparisons between past and present landscapes. It highlights the role of traditional agricultural practices, such as agroforestry, which were widespread before the shift towards modern monoculture systems. This dataset holds potential for applications in environmental sciences, historical geography, and heritage studies, providing a foundation for exploring sustainable agricultural practices and the enduring effects of rural depopulation and land-use change.

Keywords

Historic Land Use, Agroforestry, Regression Kriging, Demographic dynamics, Rural Heritage

Open Peer Review

Approval Status ? ? ✓ ?

	1	2	3	4
version 1 27 Jun 2025	? view	? view	✓ view	? view

1. **Giuseppe Cillis**, IMAA-CNR, Potenza, Italy
2. **Vassilis Detsis**, Harokopio University, Athens, Greece
3. **Ebrahim Jahanshiri**, Crops For the Future - National Institute of Agricultural Botany, Cambridge, UK
4. **Christos Chalkias**, Harokopio University, Kallithea, Greece

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the Earth and Environmental Sciences gateway.

This article is included in the Marie-Sklodowska-Curie Actions (MSCA) gateway.

This article is included in the Horizon Europe gateway.

Corresponding author: Filippo Brandolini (filbrand@mit.edu)

Author roles: Brandolini F: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project received funding from the European Union's Horizon EU research and innovation programme under the HORIZON-MSCA-2022-PF-01-01 grant agreement No. 101105219, Rural landscape hEritage and CArbon sequeSTration (RhECAST).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Brandolini F. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Brandolini F. **Reconstructing Italy's rural landscape before the Great Acceleration: A geospatial baseline from the *Catasto Agrario* (1929) [version 1; peer review: 1 approved, 3 approved with reservations]** Open Research Europe 2025, **5**:169 https://doi.org/10.12688/openreseurope.20343.1

First published: 27 Jun 2025, **5**:169 https://doi.org/10.12688/openreseurope.20343.1

Introduction

In recent decades, human activities have intensified considerably, resulting in notable changes to the landscape[1]. These changes are associated with significant ecological consequences, including alterations in land use and vegetation, soil erosion, desertification, river dynamics, and more. Land Use and Land Cover (LULC) and their transformations are key variables that substantially affect environmental processes[2]. LULC change datasets are crucial for environmental and resource management[3], allowing for precise monitoring of land use changes such as deforestation[4], urban expansion[5], desertification[6], and agricultural intensification[7,8]. These datasets are essential for sustainable resource management. In environmental modelling, LULC change data supports the prediction of future land use scenarios and the assessment of ecosystem impacts, thereby facilitating biodiversity conservation, water resource management, and mitigation of geomorphological hazard enhanced by climate change[9,10].

Most LULC datasets rely on information gathered from satellite constellations like MODIS[11], Sentinel, and Landsat[12]. Despite their high resolution and spatial coverage[13], a limitation remains their diachronic temporal resolution. The use of satellite data for studying LULC changes began in the early 1970s with the launch of the first Landsat satellite in 1972, marking the start of continuous and systematic Earth observation from space, providing valuable data for LULC studies[14]. However, most significant LULC changes started at least a century earlier, around the mid-19th century, when urbanisation and industrialisation deeply impacted rural landscapes and ecosystems. To understand the impact of pre-1970s changes on climate and environments, it is crucial to integrate data from modern satellites with historical sources like aerial images, cadastral maps and historic cartography[15–18]. Landscape archaeology has the potential to fill this gap in the data by interpreting the spatial and chronological complexity of landscapes through two main tools: Historic Landscape Characterisation (HLC) and Historic Land-Use Assessment (HLA)[19]. These methodologies introduce time-depth in landscape studies by assessing earlier patterns from historic or archaeological mappings, allowing an appreciation of past changes to guide future transformations[20]. The integration of HLC and HLA within GIS (Geographic Information Systems) technologies[21,22] underscores the importance of these landscape archaeological methodologies. They create invaluable datasets about past LULC and facilitate the interplay with other disciplines such as ecology[23,24], geomorphology[25,26], and geospatial statistics[27,28].

The rural landscape of Europe underwent dramatic changes throughout the 20th century, largely driven by the socio-economic transformations associated with the Great Acceleration, a period of rapid industrialisation and agricultural intensification that began after the 1950s[29,30]. These transformations had profound environmental, cultural, and socio-economic impacts, emphasising the need for quantitative analysis to predict future scenarios in the face of ongoing global changes.

Over the last 7 decades, Italy's landscape has seen significant shifts, with urbanisation in the lowlands and forest expansion in the uplands standing out as the most notable trends[31,32]. These changes are closely linked to wider socio-economic factors such as rural depopulation, agricultural intensification, and industrialisation[33]. Particularly after World War II (WWII), socio-economic dynamics triggered a profound transformation of Italy's rural landscape, which had previously been shaped by a long-standing historical rural tradition (Sereni, 1961). Understanding the LULC changes both before and after WWII is crucial for assessing the extent of this transformation in Italy's rural landscape.

Satellite data covers only the period from the 1970s to the present day, while historic aerial images can track LULC changes back to the 1950s. For earlier decades, information about landscape changes can be retrieved from historical cartography, cadastral maps, or land registries. However, these sources have often never been digitised, and even in cases where sporadic digital datasets exist, they often lack geospatial references, making their use in computational analysis challenging. Linking these invaluable sources of information with a spatial dimension facilitates the integration of historical data with LULC datasets developed using aerial images and satellite observations, allowing the reconstruction of human-environment interactions throughout the entire 20th century CE and even before.

The results presented here are based on the *Catasto Agrario 1929*, a survey conducted by the Italian National Institute of Statistics (Istituto Nazionale di Statistica - ISTAT) between 1928 and 1930 and it covers the whole Italian territory. This survey produced a volume for each province of Italy, offering detailed descriptive data on the Italian rural landscape in the 1920s. It serves as a comprehensive inventory, providing data on agricultural and forest areas, lands dedicated to individual crops, average crop yields per hectare plus other invaluable information such as the number of workers employed in agriculture, livestock, monthly precipitation rates, both at municipal and provincial levels[34].

The information retrieved from the Catasto Agrario was organised and georeferenced into a GIS vector dataset, providing an invaluable resource on Italy's economic and environmental conditions in the 1920s. The dataset includes not only land use and land cover (LULC) data, but also information on demography, livestock, yield rates, and precipitation rates at the provincial level. It is intended to serve as a baseline for future research in both environmental and economic sciences, facilitating a deeper understanding of rural transformations in Italy over the past century.

Methods

Data from the *Catasto Agrario* were digitised using a semi-automated workflow and subsequently integrated into a GIS framework to produce vector-based geospatial datasets. A standardised and carefully designed protocol was employed throughout the process to maintain consistency and ensure the reliability of the final dataset.

The initial phase focused on extracting data from the original rural registry. Digital copies of the *Catasto Agrario* volumes were obtained from the official ISTAT repository[35]. Optical

Character Recognition (OCR) was carried out using FineReader PDF 16[36]. This enabled the semi-automated extraction of tabular information, which was subsequently converted into spreadsheet format (.csv)[37]. The second phase involved constructing an empty geospatial dataset representing the configuration of Italy's provinces as they existed in 1929. Over the past 70 years, numerous changes have affected Italy's administrative boundaries; whereas today the country is divided into 107 provinces, in 1929 there were only 91, with several subsequent subdivisions accounting for the increase. It was therefore essential to identify and adjust areas that no longer corresponded to the historical territorial divisions. Each volume of the *Catasto Agrario* included a map depicting the administrative boundaries at the time of data collection. These maps were georeferenced in QGIS[38] following a 'backdating approach'[39,40] to detect discrepancies between historical and contemporary boundaries and place names. In parallel, modern provincial vector layers were sourced from the National Geodatabase[41] and edited to reconstruct the 1929 territorial framework. The resulting layers were exported in the Geopackage (GPKG) format[42], a standardised, platform-independent format selected to facilitate widespread reuse of the dataset. The third phase entailed transferring the information extracted via OCR from the 1920s rural registry into the empty GPKG files. To facilitate this process, a Python script[43] was developed to automatically retrieve data from the OCR-generated spreadsheets, create the necessary table fields, and systematically organise the information for each province. This script can be run through the QGIS Python interface to ensure consistency in database development. It is based on two widely used Python libraries for data manipulation and geospatial data management: `Pandas` and `GeoPandas`. `Pandas` is a powerful Python library primarily used for data manipulation and analysis. It provides data structures like DataFrames and Series, which allow for efficient handling of structured data, such as tabular data from spreadsheets or databases. `Pandas` is particularly useful for data cleaning, filtering, aggregation, and transformation tasks, making it a go-to tool for data scientists and analysts working with large datasets[44]. `GeoPandas` extends the capabilities of `Pandas` to handle geospatial data. It allows for the manipulation of geometric objects, such as points, lines, and polygons, and integrates spatial operations into the familiar `Pandas` framework. With `GeoPandas`, spatial joins, projections, and geospatial analyses can be performed directly on geospatial data, making it a valuable tool for tasks involving mapping, GIS, and spatial data science[45]. The resulting output is an updated GPKG file detailing the rural setting of Italy in the 1920s. As data is added to or removed from the spreadsheets, the Python script automatically updates the GPKG file. The final step involves dataset validation. To minimise topological errors, the QGIS `Topology Checker` plugin was used to identify potential spatial inconsistencies, including overlaps, slivers, and duplicates. These issues were subsequently corrected using the `v.clean` module from the GRASS software suite[46].

The EU CORINE Land Cover (CLC) nomenclature[47] was adopted to classify the LULC types recorded in the *Catasto Agrario*. The CLC system offers a hierarchical framework of classes, accompanied by detailed descriptions and methodologies for visual interpretation, thereby promoting consistency across mapping activities. The LULC categories applied in this dataset include agroforestry (CORINE class 2.1.1), fruit tree plantations (2.2.2), pasture (2.3.1), arable land (2.4.4), grassland (3.2.1), managed forest (3.1.1), and mixed forest (3.1.3). In addition, two supplementary classes were incorporated: uncultivated productive area and unproductive area. The former corresponds to areas identified in the *Catasto Agrario* as *Incolti produttivi*, referring to land that was potentially cultivable but not under active use at the time of survey. The latter encompasses all remaining provincial areas not explicitly classified, such as geologically unproductive zones (e.g., rock outcrops), water bodies (e.g., lakes, rivers), and infrastructure (e.g., buildings, roads, railways). The adoption of these categories was intended to closely mirror the original structure of the *Catasto Agrario*. For example, woodland areas were typically recorded without specifying dominant tree species (i.e., broadleaf or coniferous), which justified their classification as mixed forest. Conversely, chestnut groves, when explicitly identified, were categorised separately as managed forest.

Mean precipitation rates for 1929, recorded in the *Catasto Agrario* for most Italian provinces, were organised in tabular form, with monthly precipitation values documented at weather stations within provincial territories[37]. This valuable rainfall data from the 1920s was digitised into a GIS using a systematic approach. For each province, precipitation rates from individual stations were extracted using the same OCR method mentioned earlier and compiled into a dedicated spreadsheet. Since the exact locations of the weather stations are unknown, and the *Catasto Agrario* provides no coordinates, the centroid of each place name was used to create a vector point layer of meteorological stations in GPKG format. This 'empty' point layer was then populated with data from the spreadsheet using a Python script like the one used for the LULC GPKG dataset previously described.

Geostatistical analysis of the 1920s precipitation data was conducted using regression kriging[43], a widely employed method for rainfall prediction[48–50]. Key factors such as distance from the coastline, elevation, slope, and aspect were used in the analysis because they directly influence local climatic conditions and precipitation patterns. These variables help enhance the spatial accuracy of precipitation estimates by accounting for crucial geographic and topographic factors. As these factors have remained largely unchanged over the past 70 years, they provide a reliable basis for interpolating the 1920s precipitation rates recorded in the *Catasto Agrario*.

Distance from the coastline plays a crucial role in precipitation patterns, as proximity to large bodies of water increases local humidity and rainfall, particularly when moist air masses move inland. As the distance from the coast increases, moisture availability decreases, altering precipitation patterns[50]. A raster file of the Euclidean distance from the Italian coastline was developed using the GRASS module `r.grow.dist`.

Elevation significantly influences precipitation due to orographic effects[48]. As moist air rises over elevated terrain, it cools and condenses, leading to higher precipitation on windward slopes. In contrast, leeward slopes in rain shadows receive less precipitation. Including elevation in the model captures these terrain-driven precipitation patterns, particularly in areas with significant elevation changes[51]. The DEM provided by the Italian Istituto Nazionale di Geofisica e Vulcanologia[52] was downscaled to 200 m and used as the elevation covariate in the regression kriging.

Slope plays a crucial role in how air masses interact with the landscape. Steeper slopes intensify orographic uplift, resulting in increased precipitation, particularly in regions where winds push air upward. In contrast, gentler slopes have a more limited effect on air movement and precipitation. Incorporating slope into the model helps better capture the relationship between terrain and atmospheric conditions[49].

Aspect, the direction a slope faces, also significantly influences local climate. Slopes that face prevailing winds or the sun can experience different microclimates compared to those facing away. For example, south-facing slopes receive more solar radiation, which impacts temperature and air movement, ultimately affecting precipitation patterns. Windward slopes, exposed to prevailing winds, tend to receive more precipitation, while leeward slopes, sheltered from the wind, receive less. Including aspect in the model allows for a more accurate representation of these directional effects on precipitation distribution[50].

Slope and aspect covariates were computed in Python, with slope in degrees and aspect on a 0–360° scale, based on DEM gradients. Terrain attributes—distance from the coastline, elevation, slope, and aspect—underwent imputation for missing data and standardisation to ensure consistency across features. The core analysis employed regression kriging via the `PyKrige` library[53]. A multiple linear regression model was trained with terrain features as independent variables and precipitation values as the dependent variable. A grid matching the DEM resolution was generated for spatial prediction, and the regression kriging model was applied to estimate precipitation values across the region of interest. These predictions were output as a GeoTIFF file, providing a spatial representation of the 1920s precipitation distribution.

Results

The geospatial dataset resulting from the above mentioned operations consists of a multipolygon vector layer representing the 91 provinces of Italy in the 1920s. The vector layer is linked to an attribute table containing 106 entries, reflecting the key characteristics of the Italian historic rural landscape prior to WWII (Table 1 in Extended data). The units of measurement used are hectares (ha) for spatial extent and quintals (q) for weight-based quantities. The attributes were recorded in the dataset according to the original structure of the *Catasto Agrario*, including demographic information (Figure 1), livestock (Figure 1), LULC (Figure 2), detailed crop types, their extent (Figure 3), and yield rates (q/ha) (Table 1 in Extended data).

The geospatial dataset has been analysed using some basic exploratory data analysis tools to summarise its key characteristics. Figure 4 illustrates that, in the 1920s, Italy's rural landscape was predominantly shaped by agropastoral activities, with monoculture arable land being the most widespread agricultural system (Figure 2, A). Agroforestry – the land management system integrating trees and shrubs into agricultural landscapes alongside crops and/or livestock[54] – also played a significant role in the national agricultural economy. The two most common agroforestry practices were *alley cropping*, which involves planting rows of trees in agricultural fields with alleys between the rows for growing crops, and *silvopasture*, where trees, forage, and livestock grazing coexist on the same patch of land. To highlight the relevance of agroforestry in the 1920s national economic systems, Figure 2 displays both situations: the total extent of alley cropping in each province (Figure 2, B) and the total area allocated for agroforestry practices (*alley cropping* + *silvopasture* - Figure 2, C). Conversely, Figure 1 reflects the structure of the *Catasto Agrario*, where silvopasture was classified as a subtype of the pasture class.

When excluding arable land (25.3%) and agroforestry (16.4%), several observations can be made about the remaining categories in Figure 1. Mixed forest, accounting for 17.1% of the total land use, constitutes a significant portion of the landscape. This suggests that forested areas, particularly mixed forests without specific details on tree species, play an important role in the region's land cover. Pasture, including silvopasture (14.5%), also represents a notable land use, indicating the coexistence of livestock farming alongside natural or managed vegetation systems. Unproductive areas (7.7%) and uncultivated productive areas (6.0%) together make up 13.7% of the total land cover. The relatively high percentage of uncultivated but potentially productive land suggests the possibility of alternative land management practices. Unproductive areas, which include rock outcrops or infrastructure, signify parts of the landscape that do not directly contribute to agricultural or forestry activities. Fruit tree plantations (7.5%) reflect the importance of orchards and fruit production in the regions, contributing to agricultural diversity. Grasslands, covering 4.8% of the landscape, represent a smaller portion compared to other categories, likely serving as natural or semi-natural areas for grazing or biodiversity conservation. Lastly, managed forests, comprising only 0.6% of the total, indicate that actively managed woodlands, such as chestnut groves, cover a relatively small extent of the landscape.

Cereal cultivation, especially wheat, was the most common practice. In addition to wheat, the *Catasto Agrario* reported the spatial extent and annual production of rice and corn, while minor cereal production (such as rye, barley, or oats) was generally labelled as '*Other Cereals*'. Additionally, at the national level, average yield rates indicate a slight difference in cereal production between monoculture and agroforestry systems. In many cases, the two systems were integrated within the same landscape, with agroforestry complementing monoculture practices[55]. As displayed in Table 2, there were often no significant differences in production yields, and both systems contributed to maintaining overall agricultural productivity.

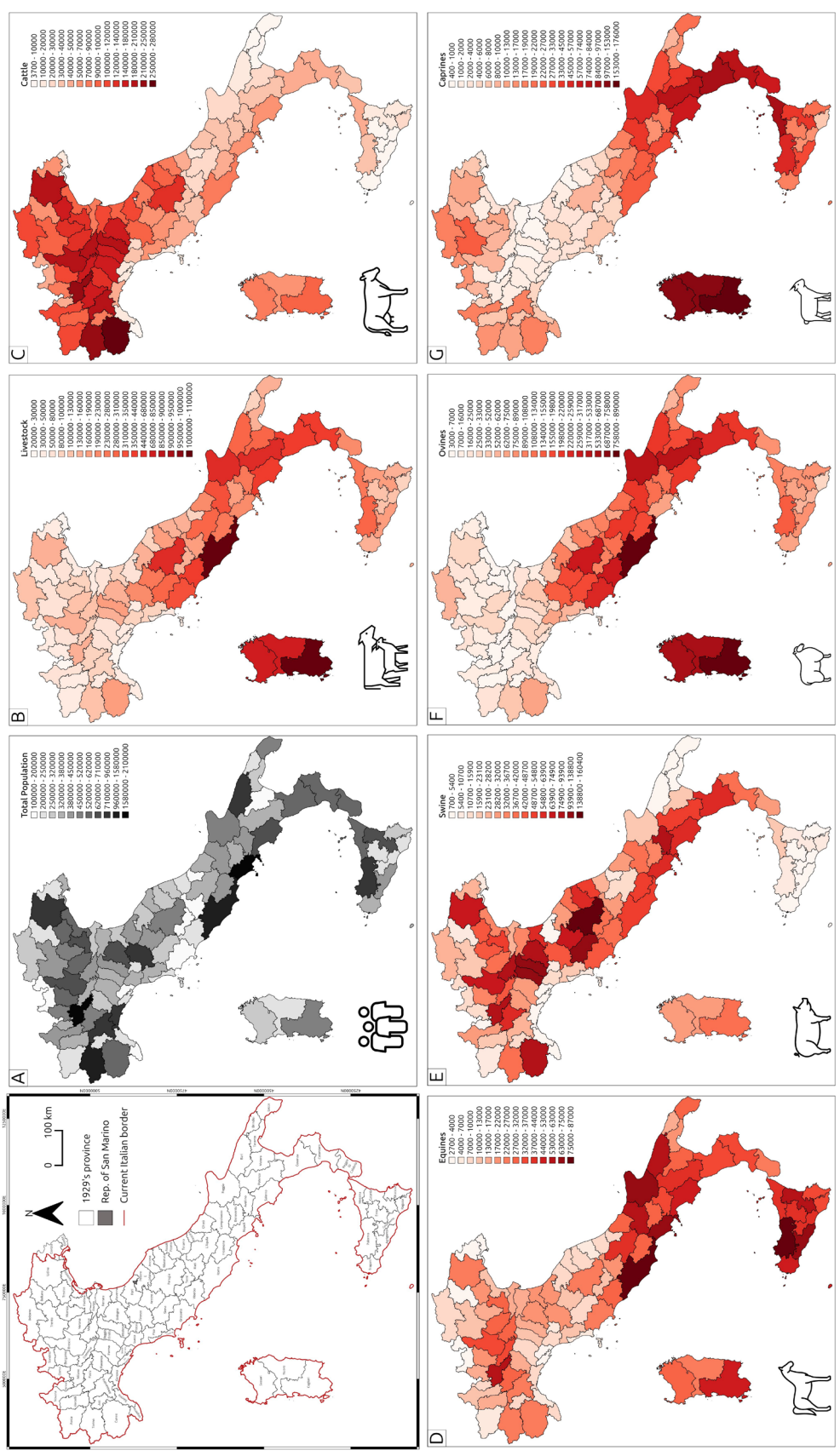

Figure 1. Thematic maps representing demographic and livestock data collected in the Catasto Agrario (1929). A – Total population; **B** – Livestock; **C** – Cattle; **D** – Equines; **E** – Swine; **F** – Ovines; **G** – Caprines.

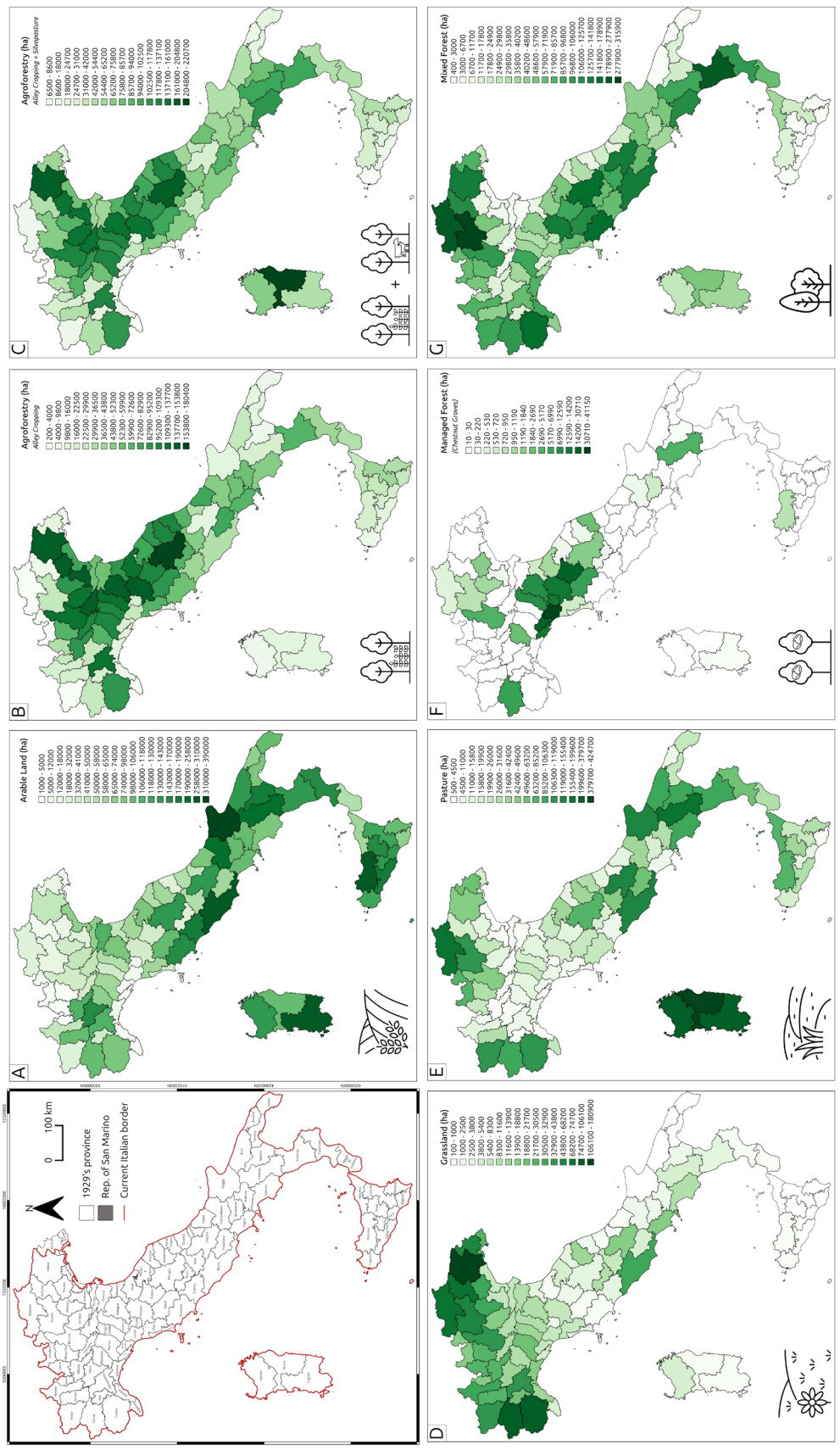

Figure 2. Thematic maps representing land use and land cover (LULC) types derived from the Catasto Agrario (1929). A – Arable land; **B** – Agroforestry (Alley cropping); **C** – Agroforestry (Alley cropping and Silvopasture); **D** – Grassland; **E** – Pasture; **F** – Managed forest; **G** – Mixed forest.

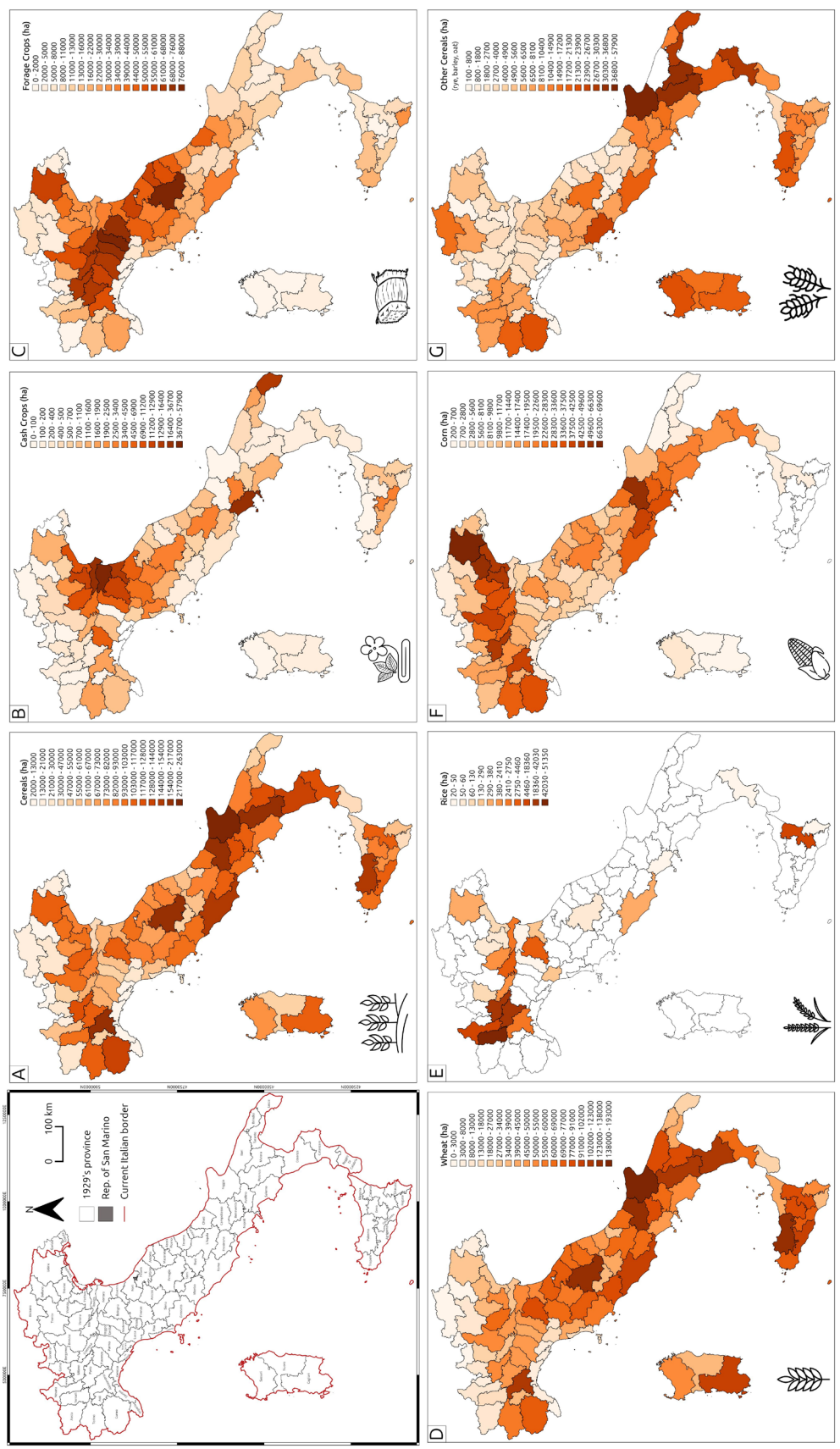

Figure 3. Thematic maps representing crop type distributions recorded in the Catasto Agrario (1929). A – Cereals; **B** – Cash crops; **C** – Forage crops; **D** – Wheat; **E** – Rice; **F** – Maize; **G** – Other cereals.

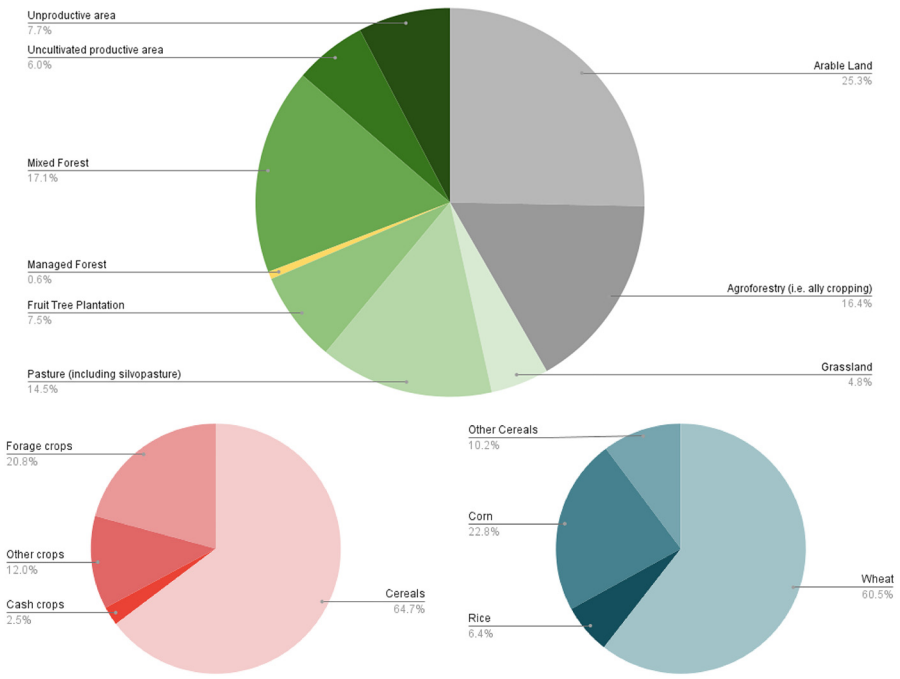

Figure 4. Overview of Italy's historic rural landscape in the 1920s. A – Distribution of LULC types; **B** – Percentage of farmland allocated to each crop type; **C** – Percentage distribution of specific cereal types ('Other cereals' includes rye, barley, and oats).

Table 2. National mean yield rates for each cereal type (wheat, rice, corn, others) in both monoculture (M) and agroforestry (AF) systems.

Wheat Yield (q/ha)		Rice Yield (q/ha)		Corn Yield (q/ha)		Other Cereals Yield (q/ha)	
M	AF	M	AF	M	AF	M	AF
14.48	14.50	41.15	40.29	16.05	16.24	12.48	12.49

As displayed in Figure 2, the primary tree species utilised in agroforestry systems were fruit trees, with regional variations such as apple, pear, apricot, peach, walnut, and cherry. Mulberry was also commonly used, particularly in agroforestry vineyard systems known as *coltura promiscua*, which refers to the traditional combination of trees, vines, and arable crops[55]. The use of citrus trees in polyculture was limited, whereas olive trees were extensively employed in southern Italy in agro-sylvo-pastoral systems, where they were integrated with pastureland, shrubs, and agricultural fields.

The 1929 precipitation rates were collected in a point vector layer associated with an attribute table that reports the number of rainy days and the rainfall (measured in mm) for each month in 1929 at each of the 398 weather stations (Table 3). The result of the regression kriging applied to the weather station dataset is a output representing the precipitation rates in Italy during the 1920s (Figure 6). To evaluate the reliability of the kriging results, several validation methods were employed. Cross-Validation Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 score were used to assess the performance of the kriging model in predicting precipitation rates (Table 4). MAE measures the average absolute difference between predicted and actual values, providing a straightforward indicator of model accuracy. MSE, which squares the errors, penalises larger deviations more heavily, offering insight into the presence of significant outliers. The R^2 score, or coefficient of determination, measures the proportion of variance in the data that the model can explain, giving an overall indication of model fit[56]. The MAE, with a value of 366.64, represents the average absolute difference between predicted and actual precipitation values, offering a clear measure of accuracy. The MSE, calculated as 205,246.11, squares the errors to penalise larger deviations, highlighting the influence of outliers. The R^2 score is 0.21, indicating that the model explains 21% of the variance in the precipitation data. These methods were selected to validate the kriging model's ability to predict spatially variable precipitation with precision.

Discussion and conclusions

The dataset and maps presented here provide a visual overview of Italy's rural landscape during the 1920s. The demographic data (Figure 1, A) illustrates the population distribution across the Italian regions. Similarly, the livestock data offers a

detailed view of the spatial distribution of various livestock types throughout Italy during the 1920s (Figure 1, B–G).

In the context of LULC setting (Figure 2), the prominent role of agroforestry during this period is particularly significant, especially considering its gradual abandonment in favour of modern industrial monoculture agriculture from the 1950s onwards[57]. In provinces of northern Italy, agroforestry was integral to the rural economy of the 1920s (Figure 4). The most common polyculture system, known as '*coltura promiscua*', involved the integration of crops with permanent mulberry trees, which also supported vines for grape cultivation. The exceptionally high occurrence of mulberry trees (Figure 5) was closely linked to the silk industry (prominent since the 1920s), as their

Table 3. Example of the attribute table associated with the point vector layer of precipitation data for 1929. The weather stations of the Umbria Region (UMB) are shown. For each month, the total precipitation (mm) and the number of rainy days (d) are reported, along with the overall amount of rainfall and the total number of rainy days.

Name	Reg	Jan (mm)	Jan (d)	Feb (mm)	Feb (d)	Mar (mm)	Mar (d)	Apr (mm)	Apr (d)	May (mm)	May (d)	Jun (mm)	Jun (d)
TORGIANO	UMB	45	5.3	62	7.3	74	9.4	81	9.2	53	7.8	60	5.7
VALFABBRICA	UMB	55	6.6	64	7.4	62	7.5	90	7.6	76	8.8	56	4.7
CASTIGLIONE DEL LAGO	UMB	57	6.9	55	6.4	73	7	70	7.3	52	7.2	47	4.3
MASSA MARTANA	UMB	79	5.2	72	5	93	6.3	110	8	98	7	66	4
GUALDO TADINO	UMB	96	8.8	121	9.6	110	12	157	12.2	114	11.4	100	7.7
Name	**Reg**	**Jul (mm)**	**Jul (d)**	**Aug (mm)**	**Aug (d)**	**Sep (mm)**	**Sep (d)**	**Oct (mm)**	**Oct (d)**	**Nov (mm)**	**Nov (d)**	**Dec (mm)**	**Dec (d)**
TORGIANO	UMB	24	2.4	26	3.1	68	5.6	71	6.7	76	9.9	46	6.8
VALFABBRICA	UMB	16	2.4	23	2.2	41	5.7	59	6.6	85	9.2	71	7.6
CASTIGLIONE DEL LAGO	UMB	26	2.1	31	2.6	61	3.6	87	5.9	86	8.3	71	7.3
MASSA MARTANA	UMB	33	2	53	2.8	76	4.6	119	5.4	131	7.4	78	6.8
GUALDO TADINO	UMB	37	3.7	65	5	132	8.3	114	9	163	11.7	176	11.2

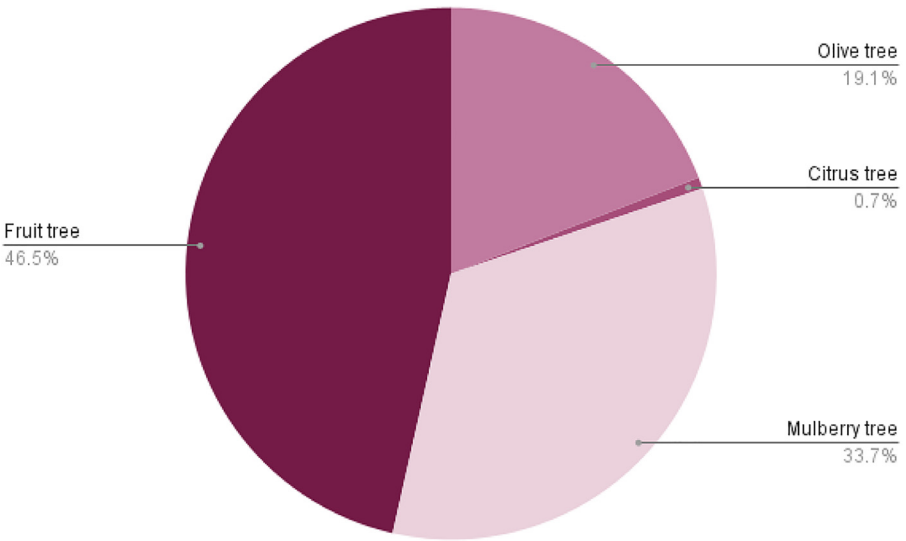

Figure 5. Tree species employed in agroforestry systems, highlighting regional variation in species used for alley cropping, silvopasture, and mixed cropping systems.

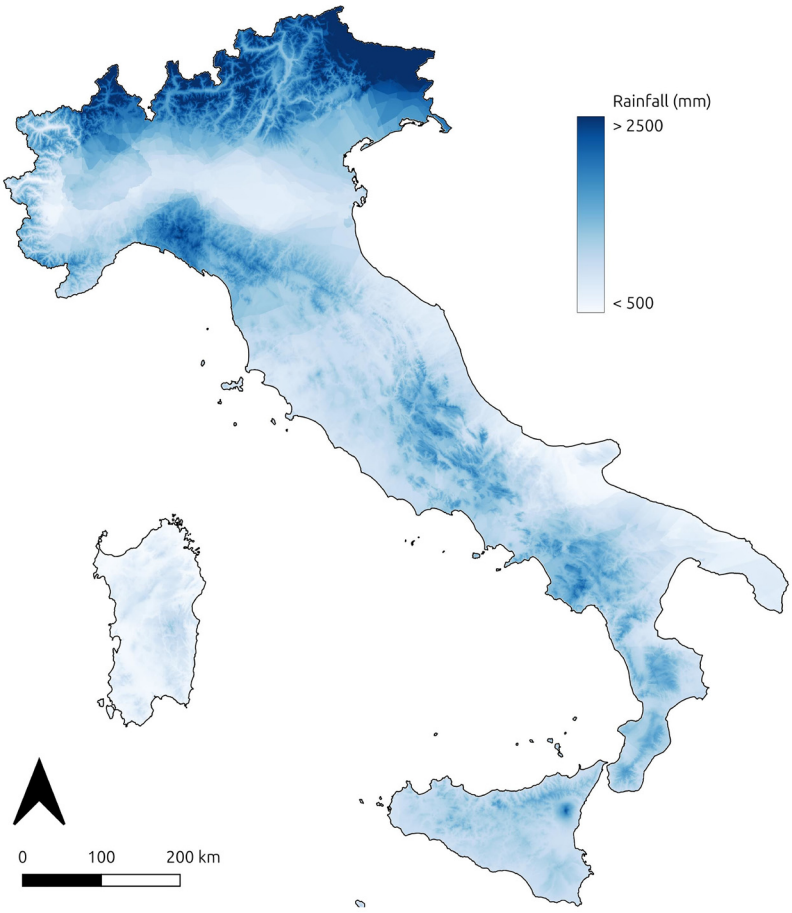

Figure 6. Regression kriging results showing spatial distribution of 1920s precipitation rates across Italy, based on weather station data digitised from the Catasto Agrario (1929).

leaves were used to feed silkworms. The decline of agroforestry after WWII was driven by the advent of mechanisation and the rise of industrial monoculture in national agriculture, influenced by evolving socio-economic dynamics[58]. This radical shift in the national rural LULC setting had potential environmental consequences, as agroforestry is regarded as one of the most sustainable agricultural strategies for supporting biodiversity[59], enhancing carbon sequestration[60], mitigating soil erosion[26], improving the water balance, reducing wildfire risks, and preserving traditional agricultural landscapes and rural knowledge[61]. The widespread abandonment of agroforestry in Italy also had detrimental effects on the cultural values of the landscape. In addition to environmental degradation, the progressive loss of farmers' expertise in these rural practices led to substantial physical changes in the regional landscape heritage[62].

Regarding Figure 6, the validation method for the geostatistical analysis of the 1920s precipitation data highlights uncertainties within this model (Table 4). The R^2 score, along with the relatively high MAE and MSE, suggests that the model is not performing well and has limited predictive power. Since the primary goal of this research was to develop a dataset associated with an output to help visualising precipitation distribution in Italy during the 1920s, the kriging performance was not explored further. For future reuse of the precipitation rate data in climatic or environmental modelling, it will be essential to improve the analysis by incorporating more predictive variables and exploring alternative approaches.

Table 4. Performance metrics for kriging model validation, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 score.

Kriging performance validation method	Value
Mean Absolute Error	363.41
Mean Squared Error	203320.59
R^2 score	0.2236

Finally, the dataset provided here holds significant potential for reuse in future studies, offering valuable insights into a range of research areas. For studies on LULC, these historical data can serve as a baseline for understanding how rural

landscapes have evolved over the past century, especially in response to agricultural industrialization and urbanisation. The demographic pressure that began in the 1950s played a pivotal role in driving both urban expansion and the intensification of agricultural practices. In response to increased food demand, agricultural systems transitioned from extensive models, such as agroforestry, to more intensive methods of cultivation. This shift was further catalysed by the agrarian reforms of the 1940s and 1950s, which, beyond addressing socio-economic demands, were also politically motivated to increase productivity. These reforms transformed Italy's agricultural landscape, particularly in rural areas, by promoting intensification and modernising agricultural structures[63]. By comparing past and present land use, researchers can identify trends in landscape transformation and their environmental, economic, and social impacts. In the context of sustainable agricultural practices, the data can inform research on the effectiveness of pre-industrial traditional systems which are recognized for their benefits in supporting biodiversity, enhancing carbon sequestration, and maintaining soil health[64]. Understanding the factors that led to the abandonment of such systems may also help guide contemporary efforts to integrate sustainability into modern agricultural strategies.

From a heritage preservation perspective, the data can play a crucial role in understanding the cultural significance of traditional rural practices and their impact on regional identities. By documenting landscape changes and the loss of pre-industrial systems, this dataset may support efforts to preserve cultural landscapes, rural knowledge, and agricultural heritage that have been gradually eroded over time[28].

Additionally, demographic data can help researchers track population changes over the last century. Since the 1950s, the depopulation of certain regions and the internal migration of people from remote areas to more industrialized and urbanized zones have been notable phenomena on the Italian peninsula[65]. The geospatial dataset presented here serves as a valuable resource for modeling demographic dynamics in Italy throughout the 20th century. Similarly, the spatial distribution of livestock types can provide deeper insights into the shift from more sustainable, traditional livestock farming methods to modern intensive systems, which developed in response to socio-economic demands following WWII. Cattle farming, particularly for dairy and beef production, is a major contributor to greenhouse gas emissions in Italy[66]. A diachronic analysis comparing the quantitative spatial distribution of livestock in the 1920s with contemporary data could aid in developing new sustainable practices.

In terms of policy development, the insights gained from this dataset can be instrumental in shaping initiatives aimed at revitalising traditional rural practices. Such policies could not only help to restore degraded landscapes but also foster rural development, support biodiversity, and enhance climate resilience. By drawing on lessons from the past, policymakers could create frameworks that balance economic viability with ecological sustainability, promoting a more harmonious relationship between agriculture and the environment[67]. Overall, the dataset provides a rich resource for understanding historical land use dynamics and has wide-ranging applications for addressing contemporary challenges related to sustainability, rural development, and environmental conservation.

Software

The OCR software ABBYY FineReader PDF 16 (A copyright license for the use of this proprietary software was obtained by the author). was used to extract information from scanned PDF files of the ISTAT *Catasto Agrario 1929*. The thematic maps (Figure 1–Figure 3 and Figure 6) and the related geo-database were edited using QGIS 3.42 Munster (free and open-source software). Script codes for building the dataset and performing kriging interpolation were developed using the Python 3.12.3 programming language.

Ethics and consent

Ethical approval and consent were not required.

Data availability

Underlying data

Zenodo: Rhecast_ORE_2025_software data, https://doi.org/10.5281/zenodo.15526385[37]

This project contains the following underlying data:

ORE_2025_Dataset.zip

Data is openly available under the Creative Commons Attribution 4.0 International License (CC BY 4.0)

Extended data

Zenodo: Rhecast_ORE_2025_software data, https://doi.org/10.5281/zenodo.15526385[37]

This project contains the following underlying data:

Extended_data.docx

Data is openly available under the Creative Commons Attribution 4.0 International License (CC BY 4.0)

Software availability

The source code developed for this research is openly available under the terms of the MIT License, an OSI-approved open-source license designed specifically for software. This license permits reuse, modification, and distribution with minimal restriction, provided that the original copyright and license terms are preserved.

- Source code available from: https://github.com/Filbra/ORE_2025
- Archived source code: https://doi.org/10.5281/zenodo.15389080
- License: MIT License

Author contributions

Conceptualization: FB; Data Curation: FB; Formal Analysis: FB; Investigation: FB; Methodology: FB; Supervision: FB; Visualization: FB; Writing – Original Draft Preparation: FB; Writing – Review & Editing: FB.

References

1. Stephens L, Fuller D, Boivin N, *et al.*: **Archaeological assessment reveals Earth's early transformation through land use.** *Science.* 2019; **365**(6456): 897–902.
PubMed Abstract | **Publisher Full Text**
2. Yaghoobi M, Vafaeenejad A, Moradi H, *et al.*: **Analysis of landscape composition and configuration based on LULC change modeling.** *Sustainability.* 2022; **14**(20): 13070.
Publisher Full Text
3. Munsi M, Malaviya S, Oinam G, *et al.*: **A landscape approach for quantifying land-use and land-cover change (1976–2006) in middle Himalaya.** *Reg Environ Change.* 2010; **10**(2): 145–155.
Publisher Full Text
4. Chávez Michaelsen A, Huamani Briceño L, Fernandez Menis R, *et al.*: **Regional deforestation trends within local realities: land-cover change in Southeastern Peru 1996–2011.** *Land.* 2013; **2**(2): 131–157.
Publisher Full Text
5. Zerboni A, Brandolini F, Mariani GS, *et al.*: **The Khartoum-Omdurman conurbation: a growing megacity at the confluence of the Blue and White Nile Rivers.** *J Maps.* 2021; **17**(4): 227–240.
Publisher Full Text
6. Rayne L, Brandolini F, Makovics JL, *et al.*: **Detecting desertification in the ancient oases of southern Morocco.** *Sci Rep.* 2023; **13**(1): 19424.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
7. Göpel J, Schüngel J, Stuch B, *et al.*: **Assessing the effects of agricultural intensification on natural habitats and biodiversity in Southern Amazonia.** *PLoS One.* 2020; **15**(11): e0225914.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
8. Gurgel AC, Reilly J, Blanc E: **Agriculture and forest land use change in the continental United States: are there tipping points?** *iScience.* 2021; **24**(7): 102772.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
9. Wang J, Bretz M, Dewan MAA, *et al.*: **Machine learning in modelling Land-Use and Land Cover-Change (LULCC): current status, challenges and prospects.** *Sci Total Environ.* 2022; **822**: 153559.
PubMed Abstract | **Publisher Full Text**
10. Ahmad MN, Shao Z, Javed A: **Modelling Land Use/Land Cover (LULC) change dynamics, future prospects, and its environmental impacts based on geospatial data models and remote sensing data.** *Environ Sci Pollut Res Int.* 2023; **30**(12): 32985–33001.
PubMed Abstract | **Publisher Full Text**
11. Yin H, Pflugmacher D, Kennedy RE, *et al.*: **Mapping annual Land Use and Land Cover changes using MODIS time series.** *IEEE J Sel Top Appl Earth Obs Remote Sens.* 2014; **7**(8): 3421–3427.
Publisher Full Text
12. Chaves MED, Picoli MCA, Sanches ID: **Recent applications of Landsat 8/OLI and Sentinel-2/MSI for Land Use and Land Cover mapping: a systematic review.** *Remote Sens.* 2020; **12**(18): 3062.
Publisher Full Text
13. Xu P, Tsendbazar NE, Herold M, *et al.*: **Comparative validation of recent 10 m-resolution global land cover maps.** *Remote Sens Environ.* 2024; **311**: 114316.
Publisher Full Text
14. Wulder MA, Roy DP, Radeloff VC, *et al.*: **Fifty years of Landsat science and impacts.** *Remote Sens Environ.* 2022; **280**: 113195.
Publisher Full Text
15. Roeoesli C, Egli M: **Using historical data to access the surface subsidence in the vegetable belt of the Three Lakes Region, Switzerland.** *Swiss J Geosci.* 2024; **117**: 9.
Publisher Full Text
16. Fuchs R, Verburg PH, Clevers JGPW, *et al.*: **The potential of old maps and encyclopaedias for reconstructing historic European Land Cover/Use Change.** *Appl Geogr.* 2015; **59**: 43–55.
Publisher Full Text
17. Grabska-Szwagrzyk E, Jakiel M, Keeton W, *et al.*: **Historical maps improve the identification of forests with potentially high conservation value.** *Conserv Lett.* 2024; **17**(5): e13043.
Publisher Full Text
18. Ridding LE, Newton AC, Redhead JW, *et al.*: **Modelling historical landscape changes.** *Landsc Ecol.* 2020; **35**: 2695–2712.
Publisher Full Text
19. Herring PC: **Framing Perceptions of the Historic Landscape: Historic Landscape Characterisation (HLC) and Historic Land-Use Assessment (HLA).** *Scott Geogr J.* 2009; **125**(1): 61–77.
Publisher Full Text
20. Turner S, Kinnaird T, Koparal E, *et al.*: **Landscape archaeology, sustainability and the necessity of change.** *World Archaeol.* 2020; **52**(4): 589–606.
Publisher Full Text
21. Bender O: **The concept of a historic landscape analysis using GIS with focus on Central Europe.** Geoinformation Technologies for Geo-Cultural Landscapes. CRC Press, 2008; 129–144.
Reference Source
22. Dabaut N, Carrer F: **Historic landscape characterisation: technical approaches beyond theory.** *Landscapes.* 2020; **21**(2): 152–167.
Publisher Full Text
23. Scharf EA: **Deep time: the emerging role of archaeology in landscape ecology.** *Landsc Ecol.* 2014; **29**: 563–569.
Publisher Full Text
24. Arıkan B, Mohr F, Bürgi M: **Exploring the common ground of landscape ecology and landscape archaeology through a case study from Eastern Anatolia, Turkey.** *Landsc Ecol.* 2021; **36**: 2295–2315.
Publisher Full Text
25. Brandolini F, Kinnaird TC, Srivastava A, *et al.*: **Modelling the impact of historic landscape change on soil erosion and degradation.** *Sci Rep.* 2023; **13**(1): 4949.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
26. Brandolini F, Compostella C, Pelfini M, *et al.*: **The evolution of historic agroforestry landscape in the Northern Apennines (Italy) and its consequences for slope geomorphic processes.** *Land.* 2023; **12**(5): 1054.
Publisher Full Text
27. Carrer F, Kossowski TM, Wilk J, *et al.*: **The application of Local Indicators for Categorical Data (LICD) to explore spatial dependence in archaeological spaces.** *J Archaeol Sci.* 2021; **126**: 105306.
Publisher Full Text
28. Brandolini F, Turner S: **Revealing patterns and connections in the historic landscape of the northern Apennines (Vetto, Italy).** *J Maps.* 2022; **18**(4): 663–673.
Publisher Full Text
29. Steffen W, Broadgate W, Deutsch L, *et al.*: **The trajectory of the Anthropocene: the great acceleration.** *Anthr Rev.* 2015; **2**(1): 81–98.
Publisher Full Text
30. McNeill JR, Engelke P: **The Great Acceleration: an environmental history of the anthropocene since 1945.** London, England: Harvard University Press, 2016.
Reference Source
31. Romano B, Zullo F, Fiorini L, *et al.*: **Land transformation of Italy due to half a century of urbanization.** *Land use policy.* 2017; **67**: 387–400.
Publisher Full Text
32. Paolini F: **Environment and urbanization in modern Italy.** University of Pittsburgh Press, 2020.
Reference Source
33. Smiraglia D, Ceccarelli T, Bajocco S, *et al.*: **Unraveling landscape complexity: Land Use/Land Cover changes and landscape pattern dynamics (1954–2008) in contrasting Peri-urban and Agro-forest regions of northern Italy.** *Environ Manage.* 2015; **56**: 916–932.
Publisher Full Text
34. Zanibelli G, Ricci V: **Wheat production in fascist period. A comparison between high farming, latifundium and sharecropping using the Catasto Agrario of 1929.** *J Siena Acad Sci.* 2022; **11**: 92–102.
Reference Source
35. ISTAT: **Catasto Agrario 1929.** 1929; [cited 2024].
Reference Source
36. ABBYY: **FineReader PDF 16.** 2023.
Reference Source
37. Brandolini F: **Rhecast_ORE_ 2025_software data.** *Zenodo.* 2025.
http://dx.doi.org/10.5281/zenodo.15526385
38. QGIS Development Team: **QGIS Geographic Information System.** Open Source Geospatial Foundation Project, 2025.
Reference Source
39. Lieskovský J, Kaim D, Balázs P, *et al.*: **Historical land use dataset of the Carpathian region (1819–1980).** *J Maps.* 2018; **14**(2): 644–651.
Publisher Full Text
40. Brandolini F, Reynard E, Pelfini M: **Multi-temporal mapping of the Upper Rhone Valley (Valais, Switzerland): fluvial landscape changes at the end of the Little Ice Age (18th-19th centuries).** *J Maps.* 2020; **16**: 212–221.
Publisher Full Text
41. Nazionale G: **Geoportale nazionale.** In: *Geoportale nazionale.* 2024; [cited 2024].
Reference Source
42. Yutzler J, Daise P: **OGC® - GeoPackage encoding standard.** 2021.
Reference Source
43. Brandolini F: **Rhecast_ORE_ 2025_Script_Codes.** *Zenodo.* 2025.
Publisher Full Text
44. The pandas development team: **pandas-dev/pandas: pandas.** *Zenodo.* 2024.
Publisher Full Text
45. Jordahl K, Van den Bossche J, Fleischmann M, *et al.*: **geopandas/geopandas:**

v0.8.1. *Zenodo*. 2020.
Publisher Full Text

46. GRASS Development Team: **Geographic Resources Analysis Support System (GRASS)**. Open Source Geospatial Foundation, 2023.
Reference Source

47. European Environment Agency: **The revised and supplemented CORINE land cover nomenclature: updated CLC illustrated nomenclature guidelines**. Vienna, Austria: European Topic Centre on Urban, Land and Soil Systems (ETC/ULS), 2019.
Reference Source

48. Teng H, Shi Z, Ma Z, *et al.*: **Estimating spatially downscaled rainfall by regression kriging using TRMM precipitation and elevation in Zhejiang Province, Southeast China.** *Int J Remote Sens.* 2014; **35**(22): 7775–7794.
Publisher Full Text

49. Seo Y, Kim S, Singh VP: **Estimating spatial precipitation using Regression Kriging and artificial Neural Network Residual Kriging (RKNNRK) hybrid approach.** *Water Resour Manage.* 2015; **29**: 2189–2204.
Publisher Full Text

50. Bostan PA, Heuvelink GBM, Akyurek SZ: **Comparison of regression and kriging techniques for mapping the average annual precipitation of Turkey.** *Int J Appl Earth Obs Geoinf.* 2012; **19**: 115–126.
Publisher Full Text

51. Smith RB: **The influence of mountains on the atmosphere.** *Advances in Geophysics.* Elsevier, 1979; **21**: 87–230.
Publisher Full Text

52. Tarquini S, Isola I, Favalli M, *et al.*: **TINITALY/01: a new Triangular Irregular Network of Italy.** *Ann Geophys.* 2007; **50**(3): 407–425.
Publisher Full Text

53. Murphy B, Yurchak R, Müller S: **GeoStat-Framework/PyKrige: v1.7.2.** *Zenodo.* 2024.
Publisher Full Text

54. Ramachandran Nair PK, Mohan Kumar B, Nair VD: **An introduction to agroforestry**. Springer International Publishing, 2021.
Publisher Full Text

55. Sereni E: **Storia del paesaggio agrario italiano**. 1961.
Reference Source

56. James G, Witten D, Hastie T, *et al.*: **An introduction to statistical learning: with applications in python.** Springer Nature, 2023.
Publisher Full Text

57. Baessler C, Klotz S: **Effects of changes in agricultural land-use on landscape structure and arable weed vegetation over the last 50 years.** *Agric Ecosyst Environ.* 2006; **115**(1–4): 43–50.
Publisher Full Text

58. Paris P, Camilli F, Rosati A, *et al.*: **What is the future for agroforestry in Italy?** *Agrofor Syst.* 2019; **93**: 2243–2256.
Publisher Full Text

59. Lacerda AEB, Hanisch AL, Nimmo ER: **Leveraging traditional agroforestry practices to support sustainable and agrobiodiverse landscapes in Southern Brazil.** *Land.* 2020; **9**(6): 176.
Publisher Full Text

60. Kay S, Rega C, Moreno G, *et al.*: **Agroforestry creates carbon sinks whilst enhancing the environment in agricultural landscapes in Europe.** *Land use policy.* 2019; **83**: 581–593.
Publisher Full Text

61. Brunori E, Maesano M, Moresi FV, *et al.*: **The hidden land conservation benefits of olive-based (*Olea europaea* L.) landscapes: an agroforestry investigation in the southern Mediterranean (Calabria region, Italy).** *Land Degrad Dev.* 2020; **31**(7): 801–815.
Publisher Full Text

62. Zerbe S: **Restoration of multifunctional cultural landscapes: merging tradition and innovation for a sustainable future.** Springer International Publishing, 2022.
Publisher Full Text

63. Bonanno A: **Theories of the state: the case of land reform in Italy, 1944-1961.** *The Sociological Quarterly.* 1988; **29**(1): 131–147.
Publisher Full Text

64. Singh R, Singh GS: **Traditional agriculture: a climate-smart approach for sustainable food production.** *Energy Ecol Environ.* 2017; **2**(3): 296–316.
Publisher Full Text

65. Reynaud C, Miccoli S, Benassi F, *et al.*: **Unravelling a demographic "Mosaic": spatial patterns and contextual factors of depopulation in Italian municipalities, 1981–2011.** *Ecol Indic.* 2020; **115**: 106356.
Publisher Full Text

66. Fiore M, Spada A, Contò F, *et al.*: **GHG and cattle farming: CO-assessing the emissions and economic performances in Italy.** *J Clean Prod.* 2018; **172**: 3704–3712.
Publisher Full Text

67. Waring BG, Gurgel A, Köberle AC, *et al.*: **Natural climate solutions must embrace multiple perspectives to ensure synergy with sustainable development.** *Front Clim.* 2023; **5**: 1216175.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ? ? ✓ ?

Version 1

Reviewer Report 20 August 2025

https://doi.org/10.21956/openreseurope.22012.r56259

© **2025 Chalkias C.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Christos Chalkias
Harokopio University, Kallithea, Greece

The paper “Reconstructing Italy’s rural landscape before the Great Acceleration” presents a novel geospatial dataset derived from the Catasto Agrario survey (1929) that digitises historical records and integrates them into a GIS framework. It reconstructs land use and land cover (LULC), demography, livestock distribution, crop yields and rainfall patterns in the Italian provinces in the 1920s. This work provides a rare pre-satellite baseline for analysing socio-environmental changes before the mid-20th century and represents an open-access resource for interdisciplinary research in environmental sciences, historical geography and cultural heritage studies. The strengths of the study lie in its innovative contribution, which fills a critical temporal gap in the LULC datasets, and in its methodological rigour. By combining OCR, GIS processing and regression kriging, the authors manage to process complex historical data and generate thematically rich geospatial data. The open accessibility of the dataset further enhances its relevance and potential impact and encourages reuse in future studies on landscape change, sustainable agriculture and rural heritage preservation. However, there are some limitations. The regression kriging used to reconstruct the precipitation data has low predictive power ($R^2 = 0.21$), which is acknowledged by the authors but not addressed in detail. Although the semi-automatic OCR process is efficient, it carries a risk of transcription errors despite the validation steps. In addition, the report could provide a deeper examination of the socio-economic drivers of landscape change beyond general references to industrialisation and rural depopulation. Given the complexity of the transition from raw historical documents to GIS-based thematic maps, a schematic diagram illustrating this data flow — from the Catasto Agrario inputs through digitisation, georeferencing and data processing to the final results — would greatly facilitate the reader's understanding and support reproducibility. It is also recommended that the authors explicitly summarise the limitations and assumptions of their approach in the discussion section. This would help future users of the dataset to better understand possible sources of uncertainty and the scope of applicability of the results. Overall, this is a well-executed and highly valuable contribution that forms the basis for further diachronic studies of Italy’s landscapes. With minor methodological clarifications, the addition of a workflow diagram and a concise summary of the assumptions and limitations in the discussion, it could have a strong impact in both academic and applied contexts.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Applied Geography, Geoinformatics, GIS, Historical GIS

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 18 August 2025

https://doi.org/10.21956/openreseurope.22012.r56668

© 2025 Jahanshiri E. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ebrahim Jahanshiri
Crops For the Future - National Institute of Agricultural Botany, Cambridge, UK

The article presents a geospatial dataset reconstructing Italy's rural landscape in the 1920s, derived from the *Catasto Agrario* 1929 survey conducted by ISTAT. The dataset integrates digitized historical data on land use and land cover (LULC), demographics, livestock, crop yields, and precipitation into a GIS framework, using vector formats aligned with 1929 provincial boundaries. Employing OCR for data extraction, Python scripting for automation, and regression kriging for precipitation interpolation, the work creates an open-access baseline for analyzing socio-environmental changes before the "Great Acceleration" post-1950s. It emphasizes traditional practices like agroforestry, provides thematic maps and exploratory analyses, and discusses applications in environmental sciences, historical geography, and sustainability studies. Data and

code are available on Zenodo and GitHub under open licenses, with limitations in kriging performance openly acknowledged.The work contains a logical structure from introduction to methods, results, and discussion, supported by 67 citations that effectively blend current literature on LULC, GIS, and landscape archaeology (e.g., references to CORINE data and kriging methods). The study design is appropriate and academically sound, as it addresses a critical gap in pre-1970s LULC data by creating a reproducible, open-access dataset that facilitates interdisciplinary comparisons between historical and modern landscaps, particularly in the context of Italy's rural transformations and sustainable agriculture. Methods are detailed sufficiently for replication, including OCR tools, QGIS workflows, python libraries (Pandas, GeoPandas, PyKrige), and validation steps like topology checking, with all scripts and data openly available, ensuring full reproducibility.

The statistical analysis via regression kriging is appropriately applied and interpreted, incorporating relevant covariates (e.g., elevation, slope) and validation metrics (MAE, MSE, R^2), while transparently noting limitations like the low R^2 (0.22), which suggests room for refinement but does not undermine the dataset's utility as a visualization tool. Conclusions are supported by results, highlighting agroforestry's historical dominnce and the dataset's potential for reuse in policy and heritage studies, without overstatement. Overall, I have no major criticisms requiring revision; it exemplifies best practices in open data by providing complete, geo-referenced resources under open licenses, making it highly valuable for future research.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Geospatial sciences, open source software, research, food systems and data science

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 18 August 2025

https://doi.org/10.21956/openreseurope.22012.r56676

© **2025 Detsis V.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Vassilis Detsis
Harokopio University, Athens, Greece

The paper presents a workflow to construct a georeferenced database from historical survey data in printed form. The objective was to make existing data available for computerised analyses. Such data predate the era of satellite images and even country-wide coverage of aerial photographs and are thus the only available country-wide window to view land cover and land use coverage in previous times. The survey data were scanned, read through an OCR application, organised in tables and linked to the province areas of the time of the survey. A number of maps and summary analyses showcase possible ways of utilisation of the data.
The paper is well written and presented and, to the best of my knowledge, the cited literature is appropriate. The design is also appropriate and can be used in subsequent works using different kinds of data with or without spatial dimensions. A number of clarifications are due with respect to the methods and the terminology used that will be detailed below. The data, as well as the code used, are fully available. I have checked the datasets and maps, including plotting indicative attributes at random, and they look fine. I noticed a single point in the rainfall map being way off to the west, probably an error that should be corrected. I have not tried running the scripts since I am not familiar with Python. The Discussion and Conclusions are consistent with the Results.

A single major concern is the fact that the CORINE land cover nomenclature is not used appropriately. Class 211 is Non-irrigated arable land, which might include ligneous crops (<25% of the area in question) but not applicable for agriforestry, which is class 244. However, the wording of the paper in the Results and the Discussion sections suggests that the talk is about fruit trees and not forestry trees, in which case class 211 is correct, although its description as agroforestry is not. If the latter is the case it should be corrected throughout the paper. Similarly, arable land belongs to classes 211 and 212 and not, as incorrectly stated, to 244. Class 311 does not refer to managed forest but to broad-leaved forest, managed or otherwise. Furthermore, absence of information is no argument to classify a forest as mixed (class 313). The author should leave it at the second level of the CORINE nomenclature, namely simply forest and class 31.

The author could consider if it would be more productive to stick to the original categories of the survey instead of using the CORINE nomenclature. Linking the two categorisations involves a number of assumptions and it is necessary only if one intends to compare the historical land use and cover with the current one. If the author sticks to this reclassification a table should be provided clearly stating which of the categories listed in the “Extended_data” table is included in which CORINE category used to create the maps in figures 2 and 3. The assumptions underlying each link should also be presented.

Some minor concerns related to methodological and related terminology issues are listed below,

organised by Section:
Methods
This enabled the semi-automated extraction of tabular information, ...: More information is due here. In which respect was the process not fully automated? The OCR applications are not flawless, furthermore the quality of their output depends on the condition of the printed materials. How was the output checked for errors?
In parallel, modern provincial vector layers were sourced from the National Geodatabase and edited to reconstruct the 1929 territorial framework.: This phrase is unclear. How were modern boundaries used to reconstruct the old ones and why, since the boundaries were already drawn from the survey itself, as described in the phrase immediately preceding this one.
These predictions were output as a GeoTIFF file, providing a spatial representation of the 1920s precipitation distribution.: The cell size should be mentioned.

Results
Define *quintal*, what is the equivalent in S.I. units?
Figure 1. Units are missing. In table 1 of the extended_data the reader is informed that it is the number of animals, but it should be included in the caption or legend. The *total population* category, apparently created by the author from the raw data, could also be numbers or animal units. Please present the units.
Figure 2. The caption would be more comprehensible if the table linking these categories to the original ones is provided (see the major concern above). Alternatively, the author could use the original survey categories.
Figure 3. As above. Furthermore, while most categories are clear there is no indication of what *cash crops* might include.
With respect to the argumentation I have no major concerns, two minor issues are listed below:
In the Introduction it is stated that *In environmental modelling, LULC change data supports the prediction of future land use scenarios and the assessment of ecosystem impacts, ...* The statement is correct, but vague. These datasets are invaluable tools enabling an understanding of past changes. Since land systems exhibit path dependency they can indeed support prediction of future scenarios. On the other hand, the processes at work are different from the past and thus future scenarios cannot be simple projections of past changes into the future. So, while the argument is valid in principle, an elaboration on how data on past changes can support future predictions would be appreciated.

In the Discussion it is stated that *In response to increased food demand, agricultural systems transitioned from extensive models, such as agroforestry, to more intensive methods of cultivation.* This is highly questionable. I am not familiar with the Italian context, however experience from elsewhere suggests that the driving force for intensification was the availability of capital to adopt the new technologies that became available in the international context already since the late 19th century.

All in all, this is a welcome contribution in the field of study of historical land systems. It provides a solid workflow along with necessary scripts and a complete geodatabase of the italian context in 1929. It thus enables the informed quantified study of land cover changes and comparisons with published past land cover maps modeled at the continental scale, enabling an assessment of their accuracy. The analyses undertaken highlight indicative potential uses, however a number of clarifications or modifications are due.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Land systems science

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 18 August 2025

https://doi.org/10.21956/openreseurope.22012.r56667

© 2025 Cillis G. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Giuseppe Cillis
IMAA-CNR, Potenza, Italy

The article is well written and the methodology is quite clear.
The use and availability of this dataset is of fundamental importance for those involved in territorial transformations in Italy.
Obviously, it is an application exclusive to Italy, but the workflow can be applied in other territorial contexts where this type of data is available.
To make it even easier to read, we suggest creating a graphical workflow diagram showing all the steps. In addition, the introduction should include a brief description of other Italian studies that have addressed this topic (studies covering the entire country, not just local or regional ones). It might also be useful to describe the type of agricultural cadastre in more detail. Information on its

creation can be found in the ISTAT publication itself.
In the conclusions, mention what other information can be obtained from the Cadastre and what the future implications are.
In the conclusions, report what other information can be obtained from the register and what the future implications are.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: study of the historical dynamics of agricultural landscape transformation in a GIS environment

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.