Detecting Temporal Land-Use Regimes in Mediterranean Agricultural Lowlands Using Landsat Time Series and Unsupervised Learning: Evidence from Albania

VALENTINA SINAJ[1], ELIRA HOXHA[1], MALVINA XHABAFTI[1], ALFRED KUME[2]
[1]University of Tirana, Economics Faculty,
Str. Arben Broci, Tirana,
ALBANIA

[2]University of Kent,
UNITED KINGDOM

Abstract: - Mediterranean agricultural lowlands are undergoing intense land-use transformations driven by climatic variability, agricultural intensification, as well as urban pressure. Despite the availability of long satellite archives, long-term seasonal land-use dynamics and their functional heterogeneity remain insufficiently captured by conventional linear trend analyses. This study proposes a geo-information–driven analytical framework that integrates long-term satellite time series with unsupervised machine learning in order to identify functional land-use regimes beyond linear assumptions. Monthly Landsat time series spanning 25 years (2000–2025), derived from Landsat 5 TM, 7 ETM+, 8 OLI/TIRS, and 9 using the Google Earth Engine platform, were analysed within the R environment in order to ensure methodological transparency and reproducibility. Spectral indices revealed pronounced seasonality and increasing vegetation activity. K-means clustering, it identified three temporal spectral regimes—intensive agriculture, transitional urban/sparsely vegetated areas, and mixed agriculture–vegetation systems—delivering a transferable framework for land-use monitoring and spatial planning.The proposed framework improves land-use analysis and also supports spatial planning.

Key-Words: - Geo-information science; Spatial time-series analysis; Landsat; Machine learning; Unsupervised clustering; Land-use dynamics; NDVI; Mediterranean agricultural landscapes.

Received: March 23, 2026. Revised: May 25, 2026. Accepted: September 11, 2026. Published: November 20, 2026.

1 Introduction

Mediterranean agricultural lowlands represent some of the most dynamic land-use landscapes in Europe, characterized by the complex interaction between climatic conditions, agricultural intensification, and increasing urban pressure [1], [2]. Within these systems, agriculture remains an essential territorial function, while at the same time being exposed to structural transformations that directly affect seasonal vegetation patterns and the long-term sustainability of land use [3], [4]. Understanding these dynamics is particularly important in fragmented Mediterranean regions, where land-use changes do not occur in a linear manner but instead through distinct phases and functional regimes driven by intertwined socio-economic and environmental factors [5], [6].

Despite the increasing availability of long-term satellite data and the widespread use of spectral indices, existing studies often rely on linear trend analysis or static land-use classifications for interpreting land-use dynamics. While these approaches provide useful insights into average changes and long-term trends, they still remain limited in their ability to sieze the functional heterogeneity and dynamic behavior of land-use systems, especially in regions characterized by strong seasonality and structural transformations [7], [8], [9]. Linear models, in particular, tend to simplify complex environmental processes and fail to take into account nonlinear interactions between vegetation, hydrological conditions, and anthropogenic pressures [10]. Similarly, traditional land-cover classifications focus on discrete spatial categories and often overlook

temporal variability and transient states within heterogeneous landscapes [5]. In this context, there is a growing need for analytical approaches that go beyond average trends and static classifications, enabling the identification of functional land-use regimes that reflect the underlying processes governing land-use dynamics [9], [11]. These approaches are particularly relevant in Mediterranean agricultural systems, where seasonal variability, irrigation practices and urban expansion interact to produce complex and nonlinear land-use patterns. In this study, functional land-use regimes are understood as recurring spectral-temporal configurations extracted from multi-index satellite time series, which reflect underlying land-use processes, such as intensive agriculture, transitional phases, or mixed land-use dynamics. Unlike traditional land use classifications, that put their focus focus on static spatial categories, this concept emphasizes temporal behavior, system dynamics, as well as the interaction between biophysical and anthropogenic factors gradually [11], [12], [13].

The Fier region represents one of the most important economic and geographical regions of Albania, historically known as the "agricultural granary" of the country as a result of its high production potential and its contribution to national food security [14]. The region is largely located within the Myzeqe Plain and includes areas of high agricultural intensity such as Lushnjë and Divjakë. In these areas, the cultivation of cereals, vegetables, and forage crops, together with the expansion of intensive greenhouse agriculture, plays a key role in shaping the economic structure and spatial organization of this territory.

In the last two decades, the agricultural landscape of this region has undergone remarkable transformations due to the interaction of socio-economic, climatic, and structural factors. Fragmentation of agricultural land following property reforms, rural migration, urbanization pressure, and climate change has directly affected the way land is used and the intensity of agricultural activity [15], [16]. In parallel, a gradual shift towards more intensive forms of production has been observed, particularly through the expansion of greenhouse areas in areas such as Divjaka and Lushnja, significantly changing seasonal cycles and the structure of land use. In Mediterranean regions, hydrological variability represents one of the key factors mediating seasonal stability and long-term productivity of agricultural vegetation. In this context, long-term and objective monitoring of agricultural land use is essential to understand the spatial and temporal dynamics of this transformation. The development of Earth observation technologies and the availability of satellite data from Landsat missions have provided a solid basis for the analysis of vegetation changes on a monthly and multi-year scale. [17] previous studies have shown that analyses based on long-term time series allow for more accurate identification of structural trends and gradual changes in land use, overcoming the limitations of time-isolated analyses [18]. The use of a 25-year time series (2000–2025) enables the identification of seasonal patterns, structural changes and long-term trends in agricultural land use, overcoming the limitations of traditional annual analyses. Time series analysis approaches have also been widely used in climate variable forecasting studies in Albania, confirming the importance of modeling seasonal and long-term dynamics for agricultural and planning applications. These studies show that the variability of meteorological conditions (especially precipitation) has a direct impact on agricultural planning and that the combination of statistical methods with Machine Learning approaches increases the value of data-driven analyses for decision-making [19].

Within this framework, the study employs a set of spectral indices to analyze agricultural land-use dynamics. Vegetation monitoring is carried out through the NDVI (Normalized Difference Vegetation Index) and EVI (Enhanced Vegetation Index), which are widely used to assess the vegetative state, density, and productivity of agricultural vegetation [20], [21]. The dynamics of water and humidity are analyzed through the NDWI index, which serves as a key indicator for the hydrological status of vegetation and irrigation needs in the intensive agricultural areas of Fier and Lushnja [22]. Furthermore, the study examines urbanization pressure and structural changes through the NDBI index (Normalized Difference Built-Up Index) and the assessment of the valid agricultural area (valid_area_ha), analyzing the impact of the expansion of built-up areas and infrastructure on arable land [23]. Spectral analysis is complemented by the use of NIR (Near Infrared) and SWIR 1 & 2 bands, which enable accurate differentiation between agricultural crops, bare soil and urban areas, as well as the assessment of plant stress and soil moisture content [24]. Empirical evidence from the literature shows that combining vegetation and moisture indices

improves the interpretation of biophysical processes related to agricultural intensity and seasonal variability in heterogeneous landscapes [12]. This highlights the need for exploratory approaches that go beyond the assessment of average effects and enable the identification of functional land-use regimes. Overall, the target of this study is to contribute to geographical land-use research by providing a long-term monthly analysis of agricultural landscape dynamics in a representative Mediterranean lowland. Through combining remote sensing data with statistical and exploratory machine learning methods, the study intents to identify functional land-use regimes as well as to deliver transferable insights relevant to spatial planning and sustainable land management in fragmented and intensively cultivated regions of Southern and Southeastern Europe. From a geo-information perspective, land-use dynamics represent spatially structured processes that demand advanced spatial analytics able to capture both temporal variability and spatial heterogeneity. It is important to clarify that, in this study, functional land-use regimes do not correspond to discrete spatial classes of land use, but rather recurring temporal spectral configurations derived from aggregated satellite time series. This approach is consistent with recent developments in time-series remote sensing, that emphasize temporal trajectories and system dynamics over static classifications [8], [25]. These regimes capture the dynamic behavior of land-use systems, reflecting processes such as agricultural intensification, transitional phases, and interactions between vegetation, water availability, and urban pressure. Therefore, the proposed framework focuses on identifying temporal states of the system rather than producing spatial land-cover maps, aligning with data-driven approaches that interpret land-surface dynamics through spectral-temporal patterns [26]. The purpose of this study is to identify and interpret functional land use designations using satellite time series data sets and unsupervised clustering. In contrast to predicting future values, the intent of this study is to understand the long-term (historical) nature of land use structures and dynamics.

2 Study Area

This study directs its focus on the Fier Region, located in the central-southwestern part of Albania, which is considered one of the most important agricultural and economic regions of the country [27]. The region lies approximately between 40.4°–41.1° north latitude and 19.2°–19.9° east longitude, with a total area of about 1,900 km², and includes the municipalities of Fier, Lushnje, Divjakë, Roskovec, Patos, and Mallakastër [14]. The geographical location of the Fier Region within Albania is depicted in Figure 1.

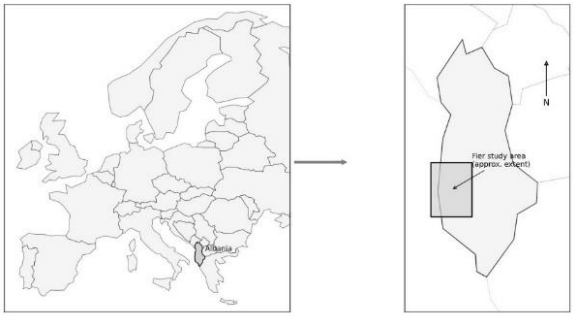


Fig. 1: Fier Region in Albania.
Source: Natural Earth; adapted by the author

From a physical-geographical point of view, the territory of the region is dominated by the Myzeqe Plain, which constitutes the largest lowland plain in Albania and one of the main agricultural areas of the country, as a result od favorable climatic conditions, flat relief, and alluvial soils with high fertility [28]. The elevation above sea level ranges from 0 m in the coastal areas of Divjaka to about 300–400 m in the hills of Mallakastra, representing a gradual transition from lowland plains to hilly relief.

The climate of the Fier Region is typically Mediterranean, with hot, dry summers and relatively mild winters, with precipitation concentrated mainly during the cold season. The average annual temperature ranges from approximately 14°C to 17°C, while the average annual precipitation varying around 900–1,100 mm, with spatial and annual variations which are influenced by the proximity to the Adriatic Sea and the local relief [15]. These climatic conditions, combined with the presence of a relatively developed irrigation and drainage network, create opportunities for intensive agriculture and multiple cropping cycles throughout the year, especially in the lowland areas [29].

The Fier region, and in particular the municipalities of Lushnje and Divjakë, constitutes one of the main branches of agricultural production in Albania, due to its high contribution to the production of cereals, vegetables, fodder crops, as well as the development of intensive agriculture in greenhouses [14]. For this reason, these areas are often described in literature and documents as the "food basket of the

country", playing an essential role in the security of food and the national rural economy [30].

However, over the last two decades, the territory of Fier District has experienced significant changes in land use and characteristics of the earth's surface, driven by socio-economic and structural factors, including the fragmentation of ownership, urbanization processes, the expansion of intensive agriculture, and the partial abandonment of agricultural lands in certain areas [15]. These processes have directly influenced the seasonal and long-term patterns of the structure of the earth's surface**,** as well as the intensity and spatial distribution of agricultural activity.

Due to its high agricultural intensity, pronounced seasonal variability and evolving structural dynamics, the Fier Region constitutes a particularly suitable study area for the long-term analysis of agricultural land use. In this context, the use of monthly time series from Landsat satellite Figures, combined with advanced data analysis methods, including machine learning-based approaches, provides a powerful analytical framework for assessing seasonal and structural changes in the structure of the Earth's surface.

2.1 Research Questions

To address the identified research gap and investigate land use dynamics in Mediterranean agricultural lowlands, this study focuses on three main research questions:

- **RQ1. Vegetation dynamics and temporal patterns**

How have vegetation activity and seasonal patterns evolved in the Fier Region during the period 2000–2025, as reflected by long-term trends and intra-annual variability of NDVI and EVI?

- **RQ2. Role of hydrological conditions and urban pressure**

To what extent do hydrological conditions (NDWI) and urbanization processes (NDBI) influence vegetation dynamics and agricultural stability over time?

- **RQ3. Identification of functional land-use regimes**

Can exploratory machine learning techniques identify distinct functional land use regimes from multi-index satellite time series, and how do these regimes complement traditional statistical analyses?

To support the temporal analysis, the study period (2000–2025) is divided into three sub-periods reflecting the main socio-economic and institutional phases in Albania: (i) 2000–2007 (post-transition period), (ii) 2008–2013 (economic restructuring period), and (iii) 2014–2025 (agricultural intensification phase). This segmentation allows for the assessment of structural changes in land use dynamics over time.

2.2 Study Hypotheses

Based on the research objectives and the identified driving factors of land use dynamics, this study formulates the following hypotheses:

- ***H_1.*** *Increasing vegetation intensity over time*

Vegetation activity, measured by NDVI and EVI, has increased during the study period (2000–2025), reflecting the intensification of agricultural practices.

- ***H_2.*** *Influence of hydrological conditions and urban pressure*

Vegetation dynamics (NDVI/EVI) are significantly affected by hydrological conditions (NDWI) and negatively affected by urban expansion (NDBI).

- ***H_3.*** *Structural changes and non-linear dynamics*

The relationships between vegetation, hydrological conditions, and urban pressure vary over time periods, indicating structural changes and nonlinear dynamics in the agricultural system.

3 Data and Methodology

To facilitate a clear and structured understanding of the methodological design, Figure 2 presents a summary of the complete analytical workflow followed in this study.

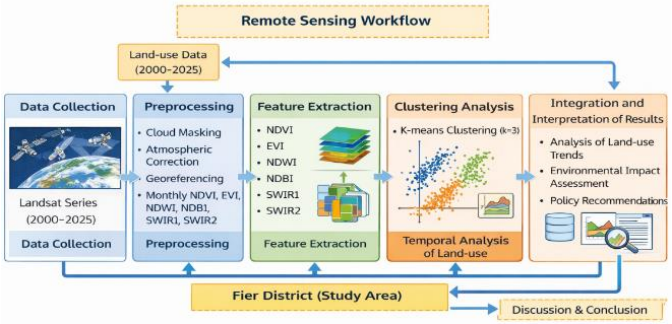


Fig. 2: Workflow of the remote sensing and machine learning framework for land-use dynamics analysis in the Fier region (2000–2025)
Source: Author's calculation

The diagram breaks down the main stages of the analysis, including: (i) Landsat satellite data collection, (ii) pre-processing and quality control, (iii) spectral feature extraction, (iv) identification of functional land use regimes through clustering techniques, and (v) temporal trend analysis.

This workflow integrates remote sensing approaches with statistical methods and machine learning algorithms, enabling advanced analysis of land use dynamics. Each component of this methodological framework is discussed in detail in the following subsections.

Data Source and Collection

This study is based on long-term time series of satellite data, extracted from the archives of the Landsat 5 (TM), Landsat 7 (ETM+), Landsat 8 (OLI/TIRS) and Landsat 9 missions, covering the period 2000–2025. These datasets provide multispectral observations with 30 m spatial resolution and are widely used for monitoring vegetation dynamics and land use change. All satellite data were accessed and preprocessed using the Google Earth Engine (GEE) platform, which provides atmospherically corrected Landsat surface reflectance products and ensures consistency across sensors [17], [31]. Specifically, GEE was used for data extraction, cloud and shadow masking, and monthly composite image generation. Post-processing of the data, statistical analysis, and machine learning modeling were performed in the R environment. This involves computing spectral indices, building time series datasets, regression modeling, and cluster analysis. The separation between data extraction (GEE) and analytical processing (R) ensures both computational efficiency and methodological transparency.

Cloud and Shadow Masking

Clouds, cloud shadows, and low-quality pixels were removed using the Quality Assessment (QA) bands provided with the Landsat surface reflectance products, following established masking procedures [32]. This step is essential for reducing spectral noise and improving the reliability of the derived indices, especially in analyses based on long-term time series.

Time Consistency and Monthly Series Construction

To ensure consistency across all Landsat missions (Landsat 5, 7, 8, and 9), the analysis relies on the harmonized surface reflectance products available in Google Earth Engine, which provide radiometrically consistent observations across all sensors [17], [30].

The monthly composite images are generated by aggregating all valid observations within each month using the arithmetic mean. Formally, for each spectral index, the monthly value is calculated as:

$$X_t = \frac{1}{n_i}\sum_{i=1}^{n_t} x_{it} \quad (1)$$

where x_{it} represents the value of pixel i in month t, while n_t is the number of valid observations. This approach reduces the impact of outliers and residual atmospheric effects and is commonly used in satellite time series analysis [25].

Months with insufficient valid observations (due to cloud contamination or limited data availability) were excluded from the analysis. This filtering ensures that the resulting time series is based on reliable and comparable observations and avoids biases in both trend estimation and clustering results [18], [33].

Calculation of Spectral Indicators

For each month of the study period, summary statistics were calculated for a set of spectral indicators, widely used in the remote sensing literature for the analysis of vegetation, hydrological conditions, and urbanization. These indicators allow the characterization of the biophysical dynamics of the land surface over time and are essential for the analysis of time series.

Vegetation Indicators

NDVI (Normalized Difference Vegetation Index), calculated according to the formula:

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (2)$$

NDVI is one of the most widely used indicators for monitoring vegetation and serves to assess its density, vitality, and photosynthetic activity [21] The values of this index are theoretically limited to the interval [-1,1], where values close to 1 indicate dense and active vegetation, values close to 0 are associated with bare surfaces (soil, sand), and negative values usually indicate the presence of water, clouds, or non-vegetated surfaces. In practice, for agricultural areas, NDVI values usually vary in the interval [0.2,0.8], depending on the density and phenological stage of the vegetation. However, the index presents some

limitations, especially in areas with high biomass, where the saturation effect can occur, reducing the sensitivity to changes in vegetation density [20].

- **EVI (Enhanced Vegetation Index)**

$$EVI = 2.5 \cdot \frac{NIR - RED}{NIR + 6 \cdot RED - 7.5 \cdot BLUE + 1} \quad (3)$$

where BLUE represents the reflectance in the blue spectral band, primarily used to correct atmospheric scattering effects.

EVI was developed in order to reduce atmospheric influences and the saturation effect that characterizes NDVI in areas with high biomass, providing a more consistent assessment of vegetative activity [20]. EVI values normally lie in the interval [-1,1], although in practical applications to vegetation they appear mostly in the [0,1]. Compared to NDVI, EVI is more sensitive to changes in canopy structure and less affected by background soil and atmospheric conditions.

Water and Urbanization Indicators

- **NDWI (Normalized Difference Water Index):**

$$NDWI = \frac{NIR - SWIR}{NIR + SWIR} \quad (4)$$

NDWI is used to assess vegetation water content and water stress, reflecting water availability in agricultural systems [22]. The values of this index are theoretically limited to the interval [-1,1], where higher values indicate higher water content in the vegetation. In Mediterranean regions, the NDWI is particularly important for analyzing the impact of hydrological conditions on agricultural productivity. However, the interpretation of the NDWI can be affected by the characteristics of the land surface and the presence of bare soil, requiring combined analysis with other spectral indices.

- **NDBI (Normalized Difference Built-Up Index):**

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \quad (5)$$

NDBI is used to identify built-up areas and to assess urban pressure on agricultural land [23]. Positive NDBI values are usually associated with urban or built-up areas, while negative values indicate the presence of vegetation or water. However, this index can be affected by the effect of "mixed pixels", especially in transitional areas where urban and agricultural elements coexist.

Exploratory Analysis with Machine Learning

To complement the statistical analysis and identify nonlinear patterns in the data, an exploratory machine learning approach based on the K-means clustering algorithm was applied. This algorithm is widely used in high-dimensional data analysis to identify hidden structures and recurring patterns [34].

Let $X = \{x_1, x_2, \dots, x_n\}$ a set of vectors where each $x_i \epsilon R^p$ represents a monthly observation composed of spectral indices (NDVI, EVI, NDWI, NDBI). The goal of the algorithm is to divide these observations into K groups in order to minimize the variance within the groups:

$$min_C \sum_{k=1}^{K} \sum_{x_{i \epsilon C_k}} \|x_i - \mu_k\|^2 \quad (6)$$

where C_k represents the kth group and μ_k its centroid.

In this study, a functional land use regime is defined as the spectral-temporal state of the land use system at a given time point (t). Mathematically, each monthly observation is assigned a functional regime R_t based on the nearest centroid in the multidimensional feature space:

$$R_t = arg\ min_k \|x_t - \mu_k\|^2 \quad (7)$$

where x_t represents the vector of spectral indices at time (t), while μ_k represents the cluster centroid (k). According to this formulation, functional regimes are interpreted as recurring spectral-temporal states identified from satellite time series rather than as static land cover classes. This approach is in line with contemporary developments in remote sensing time series analysis, which emphasize the importance of temporal trajectories and functional states of the system over traditional static classifications [25], [26].

Selecting the optimal number of clusters

The optimal number of clusters (K) was determined using a combination of the Elbow method and silhouette analysis [35]. The elbow method is based on the analysis of the within-cluster variance as a function of K, where the "elbow" point identifies the optimal trade-off between model complexity and improved variance explanation [35], [36]. In this study, the results showed a marginal reduction in variance beyond K=3, suggesting a stabilization of the

cluster structure. Furthermore, silhouette analysis, which assesses the quality of classification based on the distances within and between clusters, confirmed that *K*=3 provides an optimal balance between internal homogeneity and separation between clusters. The silhouette coefficient is defined as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (8)$$

where $a(i)$ is the average distance of an observation from the elements of its group and $b(i)$ is the minimum average distance from the nearest group.

The silhouette results confirmed that *K=3* provides an optimal balance between internal homogeneity and inter-cluster separation. From an interpretative perspective, three clusters allow for an impactful representation of the main functional states observed in the study area: (i) intensive vegetation activity, (ii) transient or mixed conditions, and (iii) lower vegetation or built-up impact. Increasing the number of clusters did not lead to significantly clearer or more interpretable patterns [34]. As a result, the susequent clusters represent temporal spectral regimes rather than spatially explicit land-use classes. This distinction is consistent with recent approaches in time-series remote sensing, where clustering techniques are used to identify recurring temporal patterns and system states rather than fixed spatial categories [26], [37].

These regimes are interpreted as functional states of the land-use system, reflecting recurring combinations of vegetation, moisture, and signals built up over time. Such interpretations are consistent with data-driven and exploratory frameworks that emphasize the role of spectral-temporal dynamics in understanding land surface processes and transitions [25], [32].This approach allows for the discovery of nonlinear patterns and hidden structures in the data, which cannot be fully captured by traditional parametric models [9], [11].

Trend Analysis and Structural Changes

Long-term trends and seasonal dynamics of the vegetation index (NDVI) were analyzed using linear regression models applied to monthly time series. The basic model specification includes a time trend and monthly dummy variables to capture intra-annual variability, allowing the separation of seasonal effects from long-term trends [38].

The formal model can be presented as:

$$NDVI_t = \beta_0 + \beta_1 t + \sum_{m=2}^{12} \gamma_m D_{m,t} + \epsilon_t \quad (9)$$

where t represents the time trend and $D_{m,t}$ are dummy variables for months.

To analyze structural changes over time, interaction terms and segmentation of the study period are included, allowing the relationships between variables to vary across time periods. This approach is consistent with the literature on the analysis of structural changes in time series [39].

Furthermore, the SWIR1 and SWIR2 bands, in combination with spectral indices, provide complementary information on surface conditions, including soil moisture and surface physical characteristics [22], [24]. These variables support the interpretation of structural changes, particularly in relation to transitions between agricultural and built-up or semi-urban land-use conditions.

To increase the robustness of the cluster analysis, the validation of the results was performed using the silhouette coefficient, the Calinski–Harabasz (CH) [40] index and the Davies–Bouldin (DB) index [41], which evaluate the compactness and separation of clusters. Also, the differences between clusters for the main spectral indices were assessed through analysis of variance (ANOVA), while in cases where the parametric assumptions were not fully met, the non-parametric Kruskal–Wallis test was also considered [42]. Chi-square test of independence was used to evaluate differences in functional regimes across historical periods of study [43].

The first-order transition matrix was also calculated for the determination of temporal stability of the identified functional regimes. Transition probabilities were estimated using the formula:

$$P_{ij} = P(R_{t+1} = j / R_t = i) \quad (10)$$

where R_t the functional regime (cluster) identified at time point (t), and P_{ij} expresses the probability that the system will transition from regime (i) to regime (j) in the subsequent period.

This approach allows for the assessment of the persistence of regimes and the identification of transitions between functional land use states.

Overall, this analytical framework allows for the identification of temporal changes in vegetation dynamics and the evaluation of their relationships

with environmental and anthropogenic factors, providing a solid basis for the interpretation of land use transformations.

4 Data Analysis

4.1 Data Quality Control

Before conducting statistical, econometric and exploratory analyses, a detailed data quality check was performed for the monthly time series for the period 2000–2025. Given that satellite data are sensitive to atmospheric factors and non-uniform coverage over time, special attention was paid to identifying and excluding months with insufficient satellite coverage. These months were identified through a combination of the low number of available Images and the valid area of analyzable pixels, measured through the variable *valid_area_ha,* which serves as an indicator of the spatial quality of the monthly estimates. The entire data processing process, statistical analysis, and machine learning procedures were performed in the R environment [44], using packages such as *dplyr, ggplot2, stats, cluster,* and *factoextra.*

The excluded months constitute a relatively small proportion of the total observations and are distributed non-systematically throughout the study period. This distribution significantly reduces the risk of temporal bias and ensures that the identified patterns are not artifacts of missing data in specific periods [17], [25].

Further analysis of the distribution of spectral variables confirmed that the values of vegetation and building indices (NDVI, EVI, NDWI, and NDBI) lie within the expected physical ranges, as defined by the relevant remote sensing literature [20], [21], [22], [23]. The logical relationships between summary statistics (minimum, mean, and maximum) are consistent over time, suggesting the absence of structural anomalies or systematic errors in data processing. Even though linear models are useful for estimating average effects and long-term trends, comparative studies have shown that they are limited in capturing structural heterogeneity and nonlinear relationships in environmental data [9].

Furthermore, the seasonal distribution of mean NDVI values reflects the phenological cycles characteristic of agriculture in the Myzeqe Plain, with lower values during the winter months and a marked increase during spring and early summer. This seasonal pattern is consistent with findings reported in similar studies on long-term vegetation monitoring in Mediterranean regions and reinforces the reliability of the dataset for long-term trend analysis, seasonality, and exploratory Machine Learning applications [7], [8].

Overall, the quality control results indicate that the constructed database is statistically robust, physically consistent, and methodologically appropriate for addressing research questions RQ1–RQ3 and testing hypotheses H_1 –H_3.

To provide an overview of the main characteristics of the database and to verify the consistency of the variables used in the analysis, descriptive statistics for the spectral indices and auxiliary indicators were calculated for the entire study period (2000–2025), in Table 1. These statistics serve as a preliminary basis for the interpretation of subsequent analyses, allowing the analysis of the average level, variability, and expected physical ranges of the satellite indices. Supplementary figures and tables are presented in the Appendix to shift the analytical focus of the main text.

Table 1. Descriptive statistics of spectral indices (2000–2025)

Variable	Average	Min	MAX	Dev. Std.
NDVI (mean)	0.474	-0.036	0.648	0.087
NDVI (min)	-0.767	-0.999	0.194	0.240
NDVI (max)	0.904	0.044	0.999	0.087
EVI (mean)	0.312	0.150	0.859	—
NDWI (mean)	0.111	-0.079	0.643	0.088
NDBI (mean)	-0.111	-0.643	0.079	0.088
Valid area (ha)	70,184	87	79,444	20,237

Source: Author's calculation

Descriptive statistics show that the spectral indices present sufficient seasonal and interannual variability, reflecting the typical dynamics of a region with intensive agriculture. The values of NDVI and EVI lie within the expected physical ranges and suggest moderate to high vegetative activity, while the distribution of NDWI and NDBI signals, respectively, the role of hydrological conditions and the presence of urban pressure. Overall, these characteristics confirm

the suitability of the data for further econometric and exploratory analyses.

4.2 Correlation Analysis

Figure 3 presents the correlation matrix between the main spectral indices (NDVI, NDWI, NDBI) and infrared bands (SWIR1 and SWIR2), providing an overview of the linear relationships that characterize the dynamics of vegetation, humidity and urban pressure in the Fier Region during the period 2000–2025.

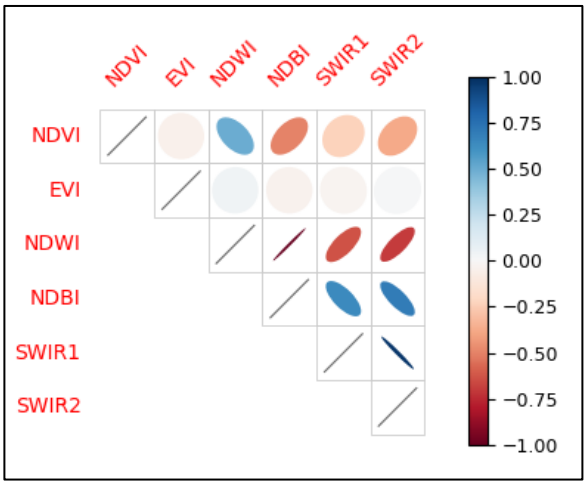


Fig. 3: Correlation matrix between spectral indices (NDVI, EVI, NDWI, NDBI) and SWIR1/ SWIR2 bands derived from Landsat imagery for the period 2000–2025

Source: Author's calculation

The correlation analysis between spectral indices and infrared bands reveals clear relationships related to vegetation dynamics and anthropogenic pressure in the Fier Region during the period 2000–2025. First, a strong positive correlation is observed between NDVI and NDWI ($r \approx 0.75$), which suggests that increased vegetative activity is associated with more favorable hydrological conditions—an expected result for a region with a Mediterranean climate, where water and humidity are determining factors for agricultural productivity. In contrast, NDVI presents a strong negative correlation with NDBI ($r \approx -0.77$), signaling that the increase in the built-up area component is associated with a decrease in green/active vegetation, consistent with the impact of urbanization and infrastructural transformations on agricultural land.

A finding that requires careful interpretation is the almost perfect negative correlation between NDWI and NDBI ($r \approx -1.00$). This result, beyond its empirical meaning, largely reflects the fact that these indices have a close mathematical relationship with common spectral bands (e.g. NIR and SWIR1), so this relationship should not be read as a "causal effect" but as a structural link that can increase collinearity in inferential models.

At the band level, the negative correlations of NDVI with SWIR1 ($r \approx -0.40$) and SWIR2 ($r \approx -0.51$) indicate that increased reflectance in the mid-infrared is associated with reduced vegetation density, being typical of drier surfaces, bare soils or built-up areas. At the same time, the positive correlations of NDVI with SWIR1 ($r \approx 0.61$) and SWIR2 ($r \approx 0.69$) reinforce the role of these bands as indicators of built-up areas and structural changes in land cover.

Overall, the correlation matrix suggests that vegetation dynamics in Fier are shaped by a strong interaction between hydrological conditions (related to NDWI) and urban/structural pressure (related to NDBI and SWIR), supporting the logic of the research questions on the role of water and urbanization in land use changes.

4.3 Regression Models and Interpretation of Results

To analyze the long-term and seasonal dynamics of the vegetation index (NDVI) in the Fier Region, several linear regression models have been estimated with monthly data for the period 2000–2025. All models include monthly dummies to control for seasonality and a linear time trend, enabling the separation of seasonal effects from long-term trends.
Since monthly time series can be characterized by heteroskedasticity, serial autocorrelation, and correlation between explanatory variables, all estimates are reported with heteroskedasticity- and autocorrelation-consistent standard errors. The use of robust standard errors ensures that statistical inference is not distorted by violations of classical OLS assumptions, especially in the presence of moderate collinearity between spectral indices [45], [46].

In this context, the basic models are presented as reference specifications that capture the long-term trend and seasonality of NDVI, upon which subsequent analyses with additional variables and structural interactions are built. The estimation results from the baseline regression models are reported in Table 2.

Table 2. Baseline regression models analysing long-term trends, seasonal effects, and the influence of water availability and urbanization on NDVI dynamics

Variables	(1) NDVI	(2) NDVI	(3) NDVI	(4) NDVI
ndwi			0.15*** (0.06)	
ndbi				-0.17*** (0.06)
year	0.01*** (0.0005)	0.01*** (0.002)	0.01*** (0.001)	0.01*** (0.001)
factor(month)2	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.02 (0.02)
factor(month)3	0.06*** (0.02)	0.06*** (0.02)	0.05*** (0.02)	0.05*** (0.02)
factor(month)4	0.10*** (0.02)	0.10*** (0.02)	0.09*** (0.02)	0.08*** (0.02)
factor(month)5	0.13*** (0.02)	0.13*** (0.02)	0.12*** (0.02)	0.11*** (0.02)
factor(month)6	0.07*** (0.02)	0.07*** (0.02)	0.06*** (0.02)	0.06*** (0.02)
factor(month)7	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)
factor(month)8	-0.01 (0.02)	-0.01 (0.02)	-0.001 (0.02)	0.001 (0.02)
factor(month)9	-0.02 (0.02)	-0.02 (0.02)	-0.01 (0.02)	-0.01 (0.02)
factor(month)10	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)
factor(month)11	0.05** (0.02)	0.05** (0.02)	0.05** (0.02)	0.04** (0.02)
factor(month)12	0.05** (0.02)	0.04** (0.02)	0.04** (0.02)	0.04** (0.02)
post2010	—	3.15 (4.28)	—	—
year:post2010	—	-0.002 (0.002)	—	—
Constant	-10.5*** (0.99)	-11.3*** (3.62)	-9.8*** (1.02)	-9.71*** (1.01)
Observ.	290	290	290	290
R^2	0.50	0.51	0.52	0.52
Adjusted R^2	0.48	0.48	0.49	0.50
Residual Std. Error	0.06 (df=277)	0.06 (df=275)	0.06 (df=276)	0.06 (df=276)
F Statistics	23.50* (12;277)	20.19* (14;275)	22.66* (13;276)	23.08* (13;276)

Notes: Standard errors are disclosed in parentheses. ***, **, and * indicates statistical significance at the 1%, 5%, and 10% levels, respectively.
Source: Author's calculation

The results of the baseline models (Models 1–4 in Table 2) show that NDVI presents a statistically significant long-term increasing trend ($\beta \approx 0.01$, $p < 0.001$), even after controlling for seasonality through monthly dummies. This result suggests a gradual intensification of vegetative activity at the district level during the study period, consistent with the findings of previous studies on agricultural intensification in Southeastern Europe [47].

The coefficients of the monthly dummies confirm the typical phenological cycle of agriculture in the Myzeqe Plain, with NDVI peaks during April–May and lower values during winter and late summer. These seasonal patterns are consistent with the Mediterranean climate and agricultural calendar of the region.

The inclusion of NDWI in the model shows a positive and statistically significant impact on NDVI ($\beta \approx 0.15$, $p < 0.01$), confirming the key role of water conditions in supporting vegetative productivity. Analogously, NDBI shows a significant negative effect on NDVI ($\beta \approx -0.17$, $p < 0.01$), suggesting that the expansion of built-up areas is associated with the contraction of active vegetation. These results are consistent with the literature on urban pressure on high-fertility agricultural lands [48], [49].

Specification with a "structural break" after 2010 did not reveal statistically significant changes in either the level or the slope of the trend in this form of the model. This suggests that changes in NDVI are better described through changing functional relationships with explanatory factors, rather than through a single discrete break in time.

4.4 Models with Different Periods

The literature on long-term land use analysis suggests that inter-periodic divisions help identify structural changes associated with economic and institutional transformations, which are often reflected in spectral indices [19]. Figure 4 presents the monthly dynamics of the average NDVI value for three distinct time periods: 2000–2007 (post-transition period), 2008–2013 (period of crisis and structural changes), and 2014–2025 (period of subsidies and agricultural intensification), while the graph for the entire period is in Figure A1 in the appendix.

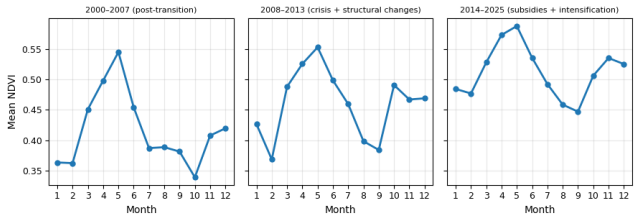


Fig. 4: Monthly dynamics of mean NDVI values across three reference periods, illustrating seasonal variability in vegetation activity
Source: Author's calculation

In all three periods, a clear seasonality of vegetation is observed, with an increase in NDVI during spring, a peak in late spring and early summer, and a decrease during late summer and autumn, in line with the phenological cycles of Mediterranean agriculture. However, significant differences appear in the level and stability of NDVI between periods. The period 2000–2007 is characterized by lower values and higher seasonal variability, reflecting an agricultural system still in a consolidation phase after the transition. During 2008–2013, a moderate improvement in vegetation intensity is observed, but accompanied by more pronounced seasonal fluctuations, which suggests structural uncertainty in land use. In the period 2014–2025, NDVI presents higher and more stable levels throughout the year, especially during the months of peak vegetation, suggesting an intensification of agricultural activity related to supporting policies and investments in the sector.

This visual evidence in Table 3 (Appendix) provides the empirical motivation for formally testing structural changes between periods through the regression models presented below.

Models with period dummies and interaction terms provide stronger evidence for structural changes in NDVI dynamics. Focusing on the period 2000–2007 as a baseline, the dummy model interprets that 2008–2013 is associated with an average increase in NDVI of about 0.04 ($p < 0.01$), while 2014–2025 is associated with an even higher increase of about 0.09 ($p < 0.01$). However, the inclusion of the time-centered trend (year_c) weakens these fixed effects, suggesting that part of the inter-period differences reflects an overall increasing trend rather than discrete shifts in the mean.

The most convincing evidence for structural changes comes from interaction models. The effect of NDWI in the baseline period is small and not statistically significant, but becomes significantly positive in 2008–2013 (ndwi × period2008–2013 ≈ 0.27, $p < 0.05$) and even stronger in 2014–2025 (ndwi × period2014–2025 ≈ 0.48, $p < 0.01$). This result suggests that in the later periods, NDVI is increasingly sensitive to water conditions, reflecting an increasing dependence of agricultural production on water and irrigation management.

Similarly, NDBI does not show a clear effect in the base period, but its interactions with periods are negative and statistically significant: ndbi × period2008–2013 ≈ −0.28 ($p < 0.01$) and ndbi × period2014–2025 ≈ −0.45 ($p < 0.01$). This shows that after 2008, the expansion of built-up areas is associated with an increasingly pronounced decrease in NDVI, suggesting an intensification of urban and infrastructural pressure on lands with active vegetation.

4.5 Exploratory Analysis with Machine Learning and Clustering

Exploratory analysis was performed using the K-means clustering algorithm, which divides observations into homogeneous groups based on the multidimensional distance between them. The spectral indices NDVI, EVI, NDWI, NDBI, SWIR1, and SWIR2 were used as input variables, which represent vegetation condition, surface moisture, urban impact, and soil dryness, respectively.

To ensure that each variable contributes equally to the clustering process, all variables were standardized using the Z-score method (zero mean and one standard deviation), a recommended practice in Machine Learning analyses with satellite data [11].

Two configurations of the algorithm were tested, namely k = 3 and k = 4, while the separation quality was evaluated using the silhouette coefficient.

The quality of clustering was assessed using the silhouette coefficient for different configurations of the K-means algorithm (k = 3 and k = 4). The results showed that the configuration with k = 3 presented higher cohesion and clearer separation of observations, being selected as the main configuration for the interpretation of the results. The details of the clustering quality assessment are presented in Table A1 in the Appendix.

In addition to the silhouette coefficient, the quality of alternative clustering solutions was also assessed through the Calinski–Harabasz (CH) [40] and Davies–Bouldin (DB) [41] indices.

Table 3. Cluster validation metrics for alternative K-means configurations

k	Silhouette	Calinski-Harabasz	Davies-Bouldin
3	0.38	206.37	1.01
4	0.34	191.74	0.98

Source: Author's calculation

The results presented in Table 4, the configuration with k = 3 achieved the highest silhouette coefficient (0.38) and the highest value of the Calinski–Harabasz index (206.37), indicating a better separation of the clusters. Although the Davies–Bouldin index was slightly lower for the configuration with k = 4 (0.98 versus 1.01), the difference was small and did not improve the interpretability of the results. Consequently, taking into account all the validation criteria, the configuration with k = 3 was selected as the optimal solution for identifying functional land use regimes.

Cluster centers (centroids) represent the average values of standardized variables within each group and serve as key elements for interpreting cluster structure and distinguishing land use profiles, summarizing the dominant multidimensional characteristics of the data [34], [50]. The cluster centers (centroids) in standardized values are presented in Table 5 for k=3 and in Table A3 in the Appendix.

Table 4. Cluster centroids (Z-scores) for the K-means solution (k = 3)

Cluster	NDVI	EVI	NDWI	NDBI	SWIR1	SWIR2
1	0.83	1.16	0.80	-0.81	-0.01	-0.13
2	-0.69	-0.63	-0.94	0.94	0.83	0.90
3	0.05	-0.38	0.43	-0.42	-1.12	-1.10

Source: Author's calculation

These results show clear structural differences between clusters, reflecting different combinations of surface biophysical characteristics.

4.5.1 Statistical Validation of Cluster Differences

To estimate whether the identified clusters represent statistically distinct groups, appropriate validation analyses were performed [36]. The results of ANOVA and Kruskal–Wallis are presented in Table 6.

Table 5. Statistical validation of cluster differences using ANOVA and Kruskal–Wallis tests

Variable	ANOVA F	ANOVA p-value	Kruskal-Wallis H	Kruskal-Wallis p-value
NDVI	97.41	< 0.001	158.97	< 0.001
EVI	230.35	< 0.001	181.63	< 0.001
NDWI	216.95	< 0.001	201.96	< 0.001
NDBI	224.91	< 0.001	203.07	< 0.001
SWIR1	247.90	< 0.001	190.44	< 0.001
SWIR2	295.38	< 0.001	203.77	< 0.001

Source: Author's calculation

The results showed statistically significant differences between clusters for all spectral indices analyzed ($p < 0.001$). This confirms that the identified clusters represent different spectral-temporal regimes and not random data distributions.

Land use profiling is achieved by interpreting the combination of spectral indices within each cluster, with the aim of identifying distinct land use types that reflect complex interactions between vegetation, moisture, and urban impact [11], [13]. Interpreting clusters as functional land use profiles, rather than simply as statistical groups, allows for a more processual understanding of the interactions between vegetation, moisture, and anthropogenic pressure [12]. Based on the centroids and mean values of the spectral indices, three distinct land use profiles were identified (Table 5, while for k=4 in Table A2 in the Appendix).

The first cluster represents areas with developed vegetation and high humidity, typical of intensive agriculture. The second cluster is characterized by low values of vegetation and humidity indices, as well as high SWIR values, suggesting bare lands or areas in urban transition. The third cluster presents an intermediate profile, reflecting transitional areas with mixed land uses.

Cluster visualization is performed through principal component analysis (PCA), which reduces the dimensions of multidimensional data while preserving most of the variance, and enables graphical interpretation of the cluster separation in a reduced space [51].

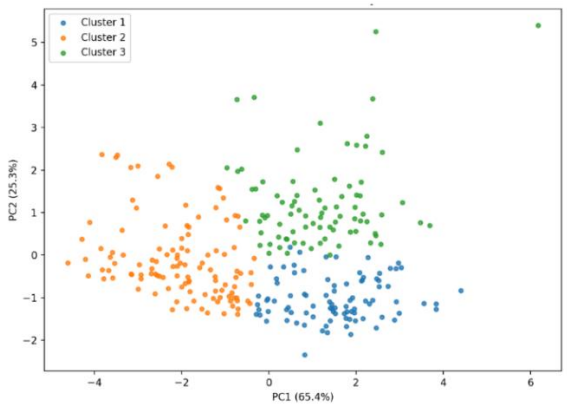


Fig. 5: Principal Component Analysis (PCA) visualization of K-means clustering results (k = 3) based on multi-index Landsat time-series data.
Source: Author's calculation

Figure 5 presents the visualization of clusters identified with the K-means algorithm (k = 3) in the space of the two principal components of the PCA analysis. (for k=4 in Figure A2 in the appendix).

A clear separation between the three clusters is observed, indicating that the observations within each group are relatively homogeneous, while the differences between the groups are pronounced. The first two components explain about 90.7% of the total variance, suggesting that the main structure of the data is preserved even after dimensionality reduction. This visual separation confirms the quality of the clustering and supports the choice of the configuration with k = 3 as the most appropriate solution for interpreting the land use profiles.

Figure 6 depicts the temporal evolution of the functional land-use regimes identified using K-means clustering (k = 3). The segmented figures show the changes in the relative weight of the three land-use profiles across distinct time periods, allowing a clearer interpretation of structural dynamics and the influence of the socio-economic context on land-use evolution.

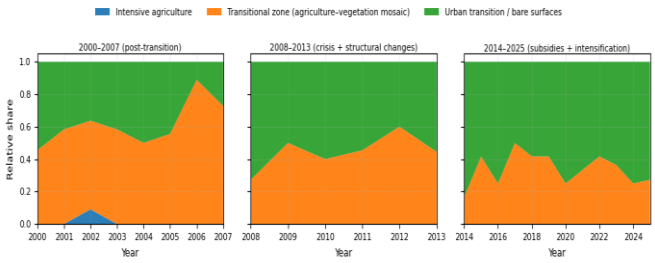


Fig. 6: Stacked area chart illustrating the temporal evolution of the three functional land-use clusters (k = 3) over the study period
Source: Author's calculation

To estimate whether the distribution of functional regimes varies between the historical periods of analysis (2000–2007, 2008–2013, and 2014–2025), the Chi-square test of independence was applied to the period × cluster frequency table. The distribution of the functional land-use regimes across the three historical periods, together with the results of the Chi-square test, is presented in Table 7.

Table 6. Distribution of functional land-use regimes across historical periods and Chi-square test results

Period	Cluster 1	Cluster 2	Cluster 3
2000–2007	19 (22.4%)	46 (54.1%)	20 (23.5%)
2008–2013	17 (27.0%)	26 (41.3%)	20 (31.7%)
2014–2025	54 (38.3%)	43 (30.5%)	44 (31.2%)

Note: Chi-square test: $\chi^2(4) = 13.53$, $p = 0.009$.
Source: Author's calculation

The results showed a statistically significant relationship between time period and cluster affiliation ($\chi^2(4) = 13.53$, $p = 0.009$), suggesting that the structure of functional land use regimes has changed significantly during the study period. Specifically, Cluster 2 dominates in the period 2000–2007 (54.1%), while after 2014, a significant increase in the weight of Cluster 1 (38.3%) is observed, reflecting a gradual shift towards regimes with higher vegetative activity and agricultural intensification.

The segmented figures show that the structure of land use profiles has changed significantly between time periods, reflecting both seasonal factors and long-term socio-economic influences. In the period 2000–2007, transitional profiles and bare areas dominate, while in the period 2008–2013, a gradual increase in intensive agriculture is observed, but with pronounced structural fluctuations. After 2014, the intensive agriculture profile shows a more stable presence, suggesting a process of agricultural intensification linked to supporting policies and sectoral investments. These findings confirm that Machine Learning-based approaches are effective in capturing complex and non-linear land use dynamics, which are often not fully identified by traditional linear statistical methods [11], [13].

In this context, unlike linear regression models, which estimate average effects of explanatory factors on NDVI, clustering analysis identifies distinct land use regimes, where combinations of vegetation, moisture, and urban pressure are manifested in

functional profiles. This indicates that land use changes are not uniform in time and space, but are organized in nonlinear structures that are more clearly captured by exploratory Machine Learning approaches.

To analyze the temporal stability of the identified functional regimes, a first-order transition matrix was calculated. This matrix represents the probability of transition from one functional regime to another between two consecutive time periods (Table 8).

Table 7. First-order transition matrix between functional land-use regimes

From / To	Cluster 1	Cluster 2	Cluster 3
Cluster 1	0.722	0.165	0.113
Cluster 2	0.072	0.687	0.241
Cluster 3	0.289	0.078	0.633

Source: Author's calculation

The results show relatively high levels of persistence for all three functional regimes, with self-transition probabilities of 72.2% for Cluster 1, 68.7% for Cluster 2, and 63.3% for Cluster 3. This suggests that the identified regimes represent relatively stable spectral-temporal states and not random configurations. The most frequent transition between regimes is observed from Cluster 3 towards Cluster 1 (28.9%), indicating the existence of gradual changes between functional land use profiles during the study period.

5 Discussion of Research Questions and Hypotheses

This study offers new insights into the dynamics of agricultural land use systems in the Mediterranean lowlands by integrating long-term satellite time series with statistical and exploratory machine learning approaches. The findings highlight the importance of considering both temporal variability and functional heterogeneity when analyzing land use dynamics.

5.1 Interpretation of Vegetation Dynamics and Driving Factors

The results showcase a clear long-term increase in vegetation activity (NDVI), accompanied by strong seasonal patterns consistent with Mediterranean agricultural systems. This trend can be interpreted as evidence of gradual agricultural intensification, most likely driven by the increasing use of irrigation, technological inputs, and the expansion of greenhouses. Similar patterns have been reported in Mediterranean regions undergoing structural transformation and intensification of agricultural production [7] [8].

The positive relationship between NDVI and NDWI confirms the critical role of water availability in supporting agricultural productivity, especially in Mediterranean climate conditions characterized by summer droughts. This finding is consistent with previous studies highlighting irrigation as an essential determinant of vegetation stability and productivity [22], [29]. The strengthening of this relationship in further periods suggests an increasing dependence of agricultural systems on water management, reflecting a shift towards more intensive and input-dependent systems.

On the contrary, the negative relationship between NDVI and NDBI highlights the impact of urban expansion on agricultural land. This relationship becomes stronger over time, suggesting that urban pressure increasingly contributes to the reduction or fragmentation of productive agricultural areas. Similar dynamics have been documented in studies on land-use change and urban expansion in Mediterranean and transition economies [48], [49].

5.2 Functional Land-Use Regimes and Non-Linear Dynamics

Beyond the average trends, clustering analysis depicts the existence of distinct functional land-use regimes, characterized by different combinations of vegetation activity, moisture conditions, and urban pressure. These regimes reflect the complex and nonlinear nature of land-use systems, where different processes coexist and evolve over time.

It is important to note that the clustering approach is applied to temporal observations, meaning that the regimes identified represent spectral-temporal configurations rather than clear spatial land-use classes. This distinction is crucial for interpretation and addresses a common limitation in remote sensing studies that attempt to directly equate spectral patterns with land-cover categories.

The identification of three dominant regimes – intensive agriculture, transitional/bare land and mixed crop-vegetation systems – suggests that land use change in the study area does not occur as a uniform process, but rather as shifts between functional states. This perspective is consistent with recent advances in earth system science, which emphasize the importance

of dynamic and process-based interpretations of land use [2], [5].

5.3 Contribution of Machine Learning Approaches

The integration of machine learning techniques, especially K-means clustering, offers added value compared to traditional statistical approaches. While regression models capture the average relationships between variables, clustering enables the identification of co-occurring patterns and hidden structures within the data. Statistically significant differences between clusters confirmed that the identified groups represent distinct land use configurations. Also, the results of the Chi-square test showed that the distribution of these regimes has changed between historical periods, while the transition matrix evidenced a high degree of temporal persistence. This combination of approaches allows for a more comprehensive understanding of land use dynamics, capturing both linear effects and nonlinear interactions. Similar findings have been reported in recent remote sensing studies that highlight the advantages of machine learning methods in analyzing complex environmental systems [11], [13].

5.4 Generalization Capability and Limitations

Although this study is presented as a case study for the Fier Region, the proposed geoinformation-based workflow is designed to be transferable to other agricultural lowlands where long-term Earth observation data are available. The approach depends on globally accessible Landsat imagery (30 m resolution) and standardized preprocessing procedures, enabling the construction of consistent monthly time series across different regions. The selected spectral indicators (NDVI, EVI, NDWI, NDBI, SWIR1, SWIR2) are widely used and physically interpreted, supporting the applicability of the framework to different agricultural contexts.

However, generalization should be understood as the transferability of the analytical workflow and not as the direct transfer of interpretations of the clusters. The identified clusters represent temporal spectral regimes derived from aggregated observations, not clear spatial land use classes. Consequently, the meaning of the clusters is context-dependent and reflects local combinations of phenology, irrigation practices, and land-use patterns. As such, similar cluster labels may not directly correspond to identical land-use regimes in other regions. In addition, the optimal number of clusters is not universal. While k = 3 provided the best separation in this study, regions with more complex land-use systems or environmental gradients may require different clustering configurations. Similarly, the use of monthly composites may be appropriate for Mediterranean environments, but may need adjustment in regions with different climatic or phenological characteristics.

A major limitation of the study is the lack of clear spatial mapping of the identified regimes and the lack of independent ground-truth validation. As a result, the findings should be interpreted as indicative of temporal functional patterns rather than as definitive land-use classifications. Future applications should include robust spatial analyses and validations using independent datasets, such as CORINE Land Cover, high-resolution imagery (e.g., Sentinel-2), or field-based observations.

Although the results of this study are interpreted and statistically strong, future studies could explore different clustering methods. These can include Gaussian Mixture Models, Spectral Clustering, Dynamic Time Warping Clustering, or other machine learning techniques. In addition, methods like state-space models, Functional Data Analysis, and graph-based representations could help improve the modeling of the temporal and spatial dependencies in land use dynamics.

6 Conclusions

This study provides a long-term monthly analysis of agricultural land-use dynamics in a representative Mediterranean lowland by integrating Landsat satellite data with statistical and exploratory Machine Learning methods. The results confirm that the agricultural landscape of the Fier Region has undergone significant structural transformations over the period 2000–2025, accompanied by a gradual intensification of agricultural activity and a pronounced seasonality of vegetation.

Time-series analysis shows that vegetation indices clearly reflect both long-term and seasonal land-use dynamics, while statistically significant relationships between NDVI, surface moisture conditions, and urban pressure highlight the combined role of biophysical and anthropogenic factors in shaping these processes. However, linear regression results also reveal the limitations of linear approaches in capturing the structural heterogeneity and non-linear

relationships that characterize fragmented Mediterranean agricultural landscapes.

In this context, exploratory K-means clustering contributes to the identification of functional land-use regimes, revealing diverse profiles that reflect different combinations of vegetation intensity, hydrological conditions, and urban pressure. Compared with linear models, this approach provides a more comprehensive understanding of non-linear structures and complex land-use dynamics, demonstrating the added value of integrating Machine Learning techniques into the analysis of long-term satellite time-series data.

All in all, the findings suggest that combining monthly time-series analysis with exploratory Machine Learning approaches constitutes a useful framework for investigating land-use dynamics in Mediterranean agricultural regions. Even though the results should be clarified as indicative of functional regimes rather than strict land-cover classifications, they offer valuable insights for spatial planning, sustainable land management, and the assessment of the impacts of agricultural intensification and urbanization in heterogeneous landscapes.

Finally, this methodological framework is transferable and can be applied to other agricultural lowlands in Southern and Southeastern Europe, contributing to a broader geographical understanding of land-use transformations in the context of climate change and socio-economic development.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the author(s) utilized artificial intelligence (AI) technologies solely for the purpose of linguistic refinement and grammatical correction. Following the application of these tools, the author(s) comprehensively reviewed, verified, and edited the resulting text. The author(s) maintain full accountability and responsibility for the accuracy, integrity, and final content of the publication.

References:

[1] Lambin, E. F., & Meyfroidt, P. (2011). Global land use change, economic globalization, and the looming land scarcity. *Proceedings of the National Academy of Sciences, 108*(9), 3465–3472. https://doi.org/10.1073/pnas.1100480108.

[2] Verburg, P. H., Erb, K.-H., Mertz, O., & Espindola, G. (2015). Land system science and sustainable development of the earth system: A global land project perspective. *Anthropocene, 12*, 29–41. https://doi.org/10.1016/j.ancene.2015.09.004.

[3] FAO. (2017). *The future of food and agriculture: Trends and challenges*. Food and Agriculture Organization of the United Nations. ISBN: 978-92-5-109551-5.

[4] Salvati, L. (2014). Towards a polycentric region? The socio-economic trajectory of Rome, an 'eternally Mediterranean' city. *Tijdschrift voor Economische en Sociale Geografie, 105*(3), 268–284. https://doi.org/10.1111/tesg.12054.

[5] Plieninger, T., Draux, H., Fagerholm, N., Bieling, C., Bürgi, M., Kizos, T., Kuemmerle, T., Primdahl, J., & Verburg, P. H. (2016). The driving forces of landscape change in Europe: A systematic review of the evidence. *Land Use Policy, 57*, 204–214. https://doi.org/10.1016/j.landusepol.2016.04.040.

[6] Sylla, M. B., Giorgi, F., Coppola, E., & Mariotti, L. (2013). Uncertainties in daily rainfall over Africa: Assessment of gridded observation products and evaluation of a regional climate model simulation. *International Journal of Climatology*. https://doi.org/10.1002/joc.3551.

[7] Pettorelli, N., et al. (2018). Satellite remote sensing of ecosystem functions: Opportunities, challenges and way forward. *Remote Sensing in Ecology and Conservation, 4*(2), 71–93. https://doi.org/10.1002/rse2.59.

[8] Zhu, Z., Zhang, J., Yang, Z., Aljaddani, A. H., Cohen, W. B., Qiu, S., & Zhou, C. (2020). Continuous monitoring of land disturbance based on Landsat time series. *Remote Sensing of Environment, 238*, 111116. https://doi.org/10.1016/j.rse.2019.03.009.

[9] Xhabafti, M., & Sinaj, V. (2024). Exploring hybrid approaches for weather forecasting domain: A review. In *Proceedings of the 19th International Conference European Integration – Realities and Perspectives (EIRP 2024)*, Galați, Romania, pp. 502–510. Danubius University, [Online]. https://dp.univ-danubius.ro/index.php/EIRP/article/view/498 (Accessed Date: January 15, 2026).

[10] Dormann, C. F., Elith, J., Bacher, S., et al. (2013). Collinearity: A review of methods to

deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x.

[11] Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

[12] Xhabafti, M., & Sinaj, V. (2024). Implementation and comparative analysis of hybrid forecasting models for drought prediction: Case of a specific region in Albania. *In* Proceedings of the 5th International Congress of Applied Statistics (UYIK-2024), pp. 1–10.

[13] Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, *39*(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343,

[14] INSTAT. (2023). Agriculture. Tirana, Albania: Institute of Statistics, [Online]. https://www.instat.gov.al/en/themes/agriculture-and-fishery/agriculture/ (Accessed Date: January 15, 2026).

[15] World Bank. (2020). *Agriculture and rural development in Albania*. Washington, DC: World Bank, [Online]. https://www.worldbank.org/en/topic/agriculture (Accessed Date: January 15, 2026).

[16] Food and Agriculture Organization. (2022). *Land use and agricultural sustainability in South-Eastern Europe*, [Online]. https://www.fao.org/ (Accessed Date: January 15, 2026).

[17] Wulder, M. A., Roy, D. P., Radeloff, V. C., Loveland, T. R., Anderson, M. C., Johnson, D. M., et al. (2019). Current status of Landsat program, science, and applications. *Remote Sensing of Environment*, *225*, 127–147. https://doi.org/10.1016/j.rse.2019.02.015.

[18] Kumari, S., & Muthulakshmi, P. (2024). SARIMA model: An efficient machine learning technique for weather forecasting. *Procedia Computer Science*, *235*, 656–670. https://doi.org/10.1016/j.procs.2024.04.064

[19] Xhabafti, M., Vika, B., & Sinaj, V. (2024). A comparative study of statistical and deep learning model-based weather prediction in Albania. *WSEAS Transactions on Computer Research*, 12, 151–160. https://doi.org/10.37394/232018.2024.12.15.

[20] Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1–2), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2.

[21] Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0.

[22] Gao, B. C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3.

[23] Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, *24*(3), 583–594. https://doi.org/10.1080/01431160304987.

[24] U.S. Geological Survey. (2023). *Landsat collection 2 level-2 science products*, [Online]. https://www.usgs.gov/landsat-missions (Accessed Date: January 15, 2026).

[25] Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D. (2010). Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment*, *114*(1), 106–115. https://doi.org/10.1016/j.rse.2009.08.014.

[26] Kennedy, R. E., Yang, Z., & Cohen, W. B. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms. *Remote Sensing of Environment*, 114(12), 2897–2910. https://doi.org/10.1016/j.rse.2010.07.008.

[27] INSTAT. (2023). *Agriculture and rural development in Albania: Regional statistical indicators*. Tirana, Albania: Institute of Statistics (INSTAT), [Online]. https://www.instat.gov.al/en/themes/general-and-regional-statistics/regions/ (Accessed Date: January 15, 2026).

[28] Çela, G., Laçi, S., & Papathimiu, S. (2025). Drought assessment under climate change using SPI, SPEI, and the hydrothermal coefficient: A case study of the Myzeqe Plain, Albania. *American Journal of Engineering Research, 14*(6), 65–72, [Online].https://www.ajer.org/papers/Vol-14-issue-6/14066572.pdf (Accessed Date: January 15, 2026).

[29] Food and Agriculture Organization of the United Nations (FAO). (2019). *Country Report of Albania: Support to the Implementation of the Regional Initiative on Empowering Smallholders and Family Farms*. Budapest, Hungary: FAO, [Online]. https://www.fao.org/3/ca4791en/ca4791en.pdf (Accessed Date: January 15, 2026).

[30] Ministry of Agriculture and Rural Development. (2021). *Agricultural development and irrigation systems in Albania*. Tirana, Albania: Ministry of Agriculture and Rural Development, [Online]. https://bujqesia.gov.al (Accessed Date: January 15, 2026).

[31] Roy, D. P., Wulder, M. A., Loveland, T. R., et al. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment, 145*, 154–172. https://doi.org/10.1016/j.rse.2014.02.001.

[32] Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment, 118*, 83–94. https://doi.org/10.1016/j.rse.2011.10.028.

[33] Xhabafti, M., & Sinaj, V. (2025). Hybrid forecasting methods across varied domains: A systematic review. *IAES International Journal of Artificial Intelligence, 14*(4), 2601–2612. http://doi.org/10.11591/ijai.v14.i4.pp2601-2612.

[34] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011.

[35] Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

[36] Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons, Hoboken, NJ. ISBN: 978-0-470-31748-8. https://doi.org/10.1002/9780470316801.

[37] Udelhoven, T. S., et al. (2009). Assessment of rainfall and NDVI anomalies using distributed lag models. *International Journal of Remote Sensing, 30*(8), 1961–1976. https://doi.org/10.1080/01431160802546829.

[38] Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). Mason, OH: Cengage Learning. ISBN: 978-1-111-53439-4.

[39] Stock, J. H., & Watson, M. W. (2015). *Introduction to Econometrics* (3rd ed.). Harlow, UK: Pearson. ISBN: 978-1-292-07955-4.

[40] Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics, 3*(1), 1–27. https://doi.org/10.1080/03610927408827101.

[41] Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 1*(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909.

[42] Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association, 47*(260), 583–621. https://doi.org/10.1080/01621459.1952.10483441.

[43] Agresti, A. (2018). *An introduction to categorical data analysis* (3rd ed.). John Wiley & Sons, Hoboken, NJ. ISBN: 978-1-119-40526-9.

[44] R Core Team. (2023). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing, [Online]. https://www.r-project.org/ (Accessed Date: January 15, 2026).

[45] White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*(4), 817–838. https://doi.org/10.2307/1912934.

[46] Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica, 55*(3), 703–708. https://doi.org/10.2307/1913610.

[47] Kuemmerle, T., Levers, C., Erb, K., Estel, S., Jepsen, M. R., Müller, D., Plutzar, C., Stürck, J., Verkerk, P. J., Verburg, P. H., & Reenberg, A.

(2016). Hotspots of land use change in Europe. *Environmental Research Letters, 11*(6), Article 064020. https://doi.org/10.1088/1748-9326/11/6/064020.

[48] Seto, K. C., Güneralp, B., & Hutyra, L. R. (2012). Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences, 109*(40), 16083–16088. https://doi.org/10.1073/pnas.1211658109.

[49] Li, X., Zhou, Y., Eom, J., Yu, S., & Asrar, G. R. (2019). Projecting global urban land expansion and heat island intensification through 2050. *Environmental Research Letters*, *14*(11), 114037. https://doi.org/10.1088/1748-9326/ab4b71.

[50] Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Chichester, UK: John Wiley & Sons. ISBN: 978-0-470-74991-3.

[51] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A*: Mathematical, Physical and Engineering Sciences, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Valentina Sinaj conceived the study, developed the research framework, and performed the main data processing and analysis. The remaining authors contributed equally to methodological refinement, interpretation of results, preparation of the manuscript, and critical revision of the paper. All authors reviewed and approved the final version of the manuscript.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

The research presented in this article was supported by the Albanian-American Development Foundation (AADF) through the READ (Research Expertise from the Academic Diaspora) Program.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US

APPENDIX

Table 8. Regression models including period-specific effects to evaluate changes in NDVI dynamics over time

Variable	(1)	(2)	(3)	(4)
ndwi			-0.03 (0.07)	
ndbi				0.01 (0.07)
year_c	0.004 (0.003)	0.003** (0.001)	0.003** (0.001)	0.003** (0.001)
period 2008–2013	0.04*** (0.01)	-0.02 (0.05)	-0.003 (0.02)	-0.002 (0.02)
period 2014–2025	0.09*** (0.01)	0.05 (0.03)	-0.01 (0.02)	-0.01 (0.02)
Month 2	-0.01 (0.02)	-0.01 (0.02)	-0.02 (0.02)	-0.02 (0.02)
Month 3	0.06*** (0.02)	0.06*** (0.02)	0.05*** (0.02)	0.05** (0.02)
Month 4	0.10*** (0.02)	0.10*** (0.02)	0.08*** (0.02)	0.08*** (0.02)
Month 5	0.13*** (0.02)	0.13*** (0.02)	0.11*** (0.02)	0.11*** (0.02)
Month 6	0.07*** (0.02)	0.07*** (0.02)	0.06*** (0.02)	0.06*** (0.02)
Month 7	0.02 (0.02)	0.02 (0.02)	0.03 (0.02)	0.03* (0.02)
Month 8	-0.01 (0.02)	-0.01 (0.02)	0.01 (0.02)	0.01 (0.02)
Month 9	-0.02 (0.02)	-0.02 (0.02)	0.001 (0.02)	0.004 (0.02)
Month 10	0.02 (0.02)	0.02 (0.02)	0.03 (0.02)	0.03 (0.02)
Month 11	0.05** (0.02)	0.05** (0.02)	0.04** (0.02)	0.04** (0.02)
Month 12	0.05** (0.02)	0.05** (0.02)	0.04** (0.02)	0.04** (0.02)
year_c × period 2008–2013		0.003 (0.01)		
year_c × period 2014–2025		-0.001 (0.003)		
second × period 2008–2013			0.27** (0.11)	
ndwi × period 2014–2025			0.48*** (0.10)	
ndbi × period 2008–2013				-0.28*** (0.11)
ndbi × period 2014–2025				-0.45*** (0.09)
constant	0.38*** (0.01)	0.37*** (0.02)	0.38*** (0.02)	0.37*** (0.01)
observation	290	290	290	290
R^2	0.50	0.51	0.56	0.56
AdjustedR^2	0.48	0.48	0.53	0.54

Notes: Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Source: Author's calculation

Table A1. Clustering quality for different K-means configurations

Number of clusters (k)	**Average silhouette score**	**Evaluation**
3	0.38	Better and more stable partitioning
4	0.34	Weaker partitioning

Source: Author's calculation

Table A2. Cluster summary (k=4)

cluster_k4	ndvi	evi	ndwi	ndbi	swir1	swir2	n
1	0.406998	0.267262	0.020125	-0.0098	0.249804	0.162135	100
2	0.557549	0.377658	0.222226	-0.2237	0.180621	0.105812	45
3	0.465696	0.274104	0.12538	-0.1176	0.177258	0.106621	71
4	0.524682	0.375356	0.150149	-0.1466	0.228359	0.139919	73

Source: Author's calculation

Table A3. Cluster centroids(Z-score),k=4

cluster_k4	ndvi	evi	ndwi	ndbi	swir1	swir2
1	-0.78462	-0.71651	-1.01342	1.015919	0.85442	0.947428
2	1.033452	1.007794	1.237787	-1.2705	-0.90548	-1.00101
3	-0.13246	-0.59978	0.248348	-0.21304	-1.05147	-1.01067
4	0.650842	1.038313	0.515726	-0.52093	0.333937	0.190939

Source: Author's calculation

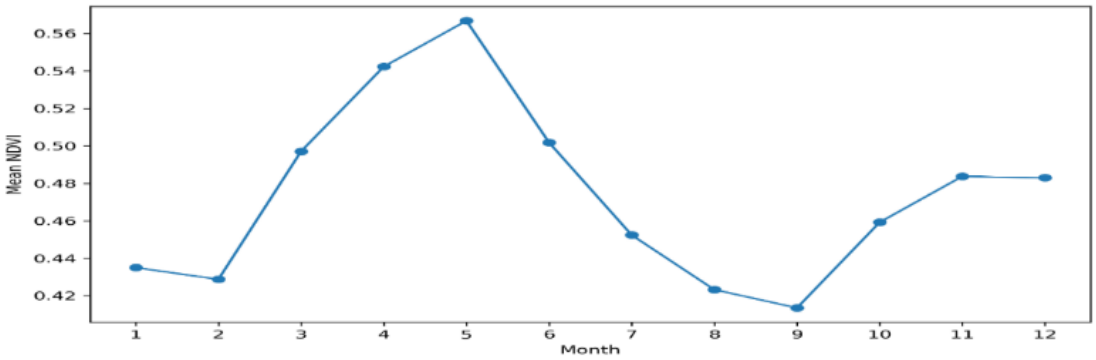


Fig. A1: Monthly NDVI dynamics in the Fier region (2000–2025)
Source: Author's calculation

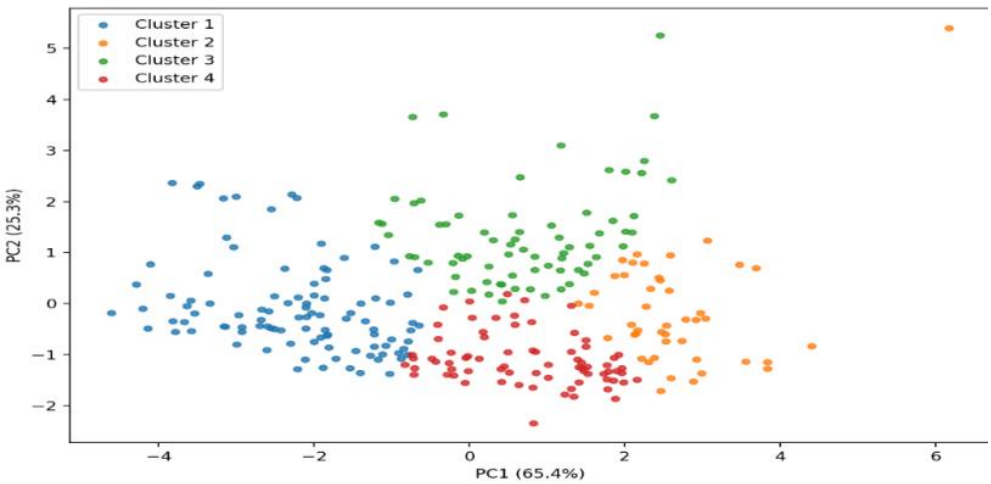


Fig. A2: PCA visualization of clusters (k = 4)
Source: Author's calculation