Article

Identification of Ecosystem Functional Types from Coarse Resolution Imagery Using a Self-Organizing Map Approach: A Case Study for Spain

Ana Pérez-Hoyos *, Beatriz Martínez, Francisco Javier García-Haro, Álvaro Moreno and María Amparo Gilabert

Departament de Física de la Terra i Termodinàmica, Universitat de València, Dr. Moliner, 50, 46100 Burjassot (València), Spain; E-Mails: Beatriz.Martinez@uv.es (B.M.); J.Garcia.Haro@uv.es (F.J.G.-H.); Alvaro.Moreno@uv (A.M.); M.Amparo.Gilabert@uv.es (M.A.G.)

* Author to whom correspondence should be addressed; E-Mail: Ana.Perez-Hoyos@uv.es; Tel.: +34-96-354-32-56; Fax: +34-963-543-385.

Abstract: Ecosystem state can be characterized by a set of attributes that are related to the ecosystem functionality, which is a relevant issue in understanding the quality and quantity of ecosystem services and goods, adaptive capacity and resilience to perturbations. This study proposes a major identification of Ecosystem Functional Types (EFTs) in Spain to characterize the patterns of ecosystem functional diversity and status, from several functional attributes as the Normalized Difference Vegetation Index (NDVI), Land Surface Temperature (LST) and Albedo. For this purpose, several metrics, related to the spatial variability in seasonal and annual patterns (e.g., relative range), have been derived from remote sensing time series of 1 km MODIS over the period 2000–2009. Moreover, precipitation maps from data provided by the AEMet (Agencia Estatal de Meteorología) and the corresponding aridity and humidity indices were also included in the analysis. To create the EFTs, the potential of the joint use of Kohonen's Self-Organizing Map (SOM) and the *k-means* clustering algorithm was tested. The EFTs were analyzed using different remote sensing (*i.e.*, Gross Primary Production) and climatic variables. The relationship of the EFTs with existing land cover datasets and climatic data were analyzed through a correspondence analysis (CA). The trained SOM have shown feasible in providing a comprehensive view on the functional attributes patterns and a remarkable potential for the quantification of

ecosystem function. The results highlight the potential of this technique to delineate ecosystem functional types as well as to monitor the spatial pattern of the ecosystem status as a reference for changes due to human or climate impacts.

1. Introduction

Ecosystem transformation and status in the earth system are raising increasing concern [1] because the ecosystems are vital for human and animal well-being. Ecosystem provides food and other products and performs a wide range of functions that support existence [2]. Hence, there is increasing awareness of the dependence of the services and goods of humankind on the Earth's ecosystem (*i.e.*, natural and modified anthropologically) [3,4]. Nevertheless, terrestrial ecosystems are permanently changing at a variety of spatial and temporal scales [5], especially in recent decades, where they have experienced strong changes due to human activity [6].

Accurate information on the distribution, properties and status of ecosystems is essential as input for meteorological and biogeochemical models [7,8], for global change research [9], biodiversity conservation [10], to develop sustainable strategies for a wide range of human activities [11] and for ecosystem management [12].

Stratifying the ecosystem into relatively homogeneous regions or patterns allows reducing the complexity of the landscape into something that is more manageable and understandable [13]. Ecological land classification, a hierarchically structured, multifactor approach for mapping ecological units at multiple scales, has been shown to be helpful in quantifying variation in fundamental ecological processes [14]. In this respect, diverse approaches and nomenclatures have been applied and defined at continental [15], regional [16] and local scales [17]. Depending on the final application, these stratifications are namely different. For example, ecoregions have been defined as geographic zones that represent geographical groups or associations of similarly functioning ecosystems [18]. Other approaches of this type include the maps of isogrowth [19], the delimitation of biozones [20], pheno-classes [13] and pheno-regions [21].

Traditionally, the description and characterization of ecosystems rely only on the structural attributes of the vegetation such as physiognomy, landform, dominant species or floristic composition [22]. However, the structural attributes of ecosystems may not be sensitive enough to assess the impact of current environmental changes if the response of the vegetation structure has a long time lag [23]. Hence, the ecosystem status characterization can also be complemented considering also the functional attributes (or state variables) [24]. Functional groups correspond to sets of organisms at any level of biological organization (*i.e.*, microorganisms, plants) sharing a certain set of common structural or process features [25]. These attributes (e.g., primary production or heat fluxes) capture critical aspects of carbon, energy and water fluxes between the biota and the environment [26]. Functional attributes offer advantages over the traditional use of structural variables, since they respond faster to environmental disturbances than structural ones [24]. Additionally, they can easily be monitored at

different spatial scales, over large areas, which is particularly useful to track the effects of global environmental changes, evaluate the effectiveness of management practices and analyze vegetation dynamics [27]. Moreover, functional attributes are associated with ecosystem services [28], which represent the goods and services utilized by humanity.

In this sense, Ecosystem Functional Types (EFTs) offer an optimal framework for functional classification of ecosystems because they represent the spatial heterogeneity in ecosystem functioning, complementing and improving the description based only on structural features [29]. EFTs are analogous to the more extended approach of Plant Functional Types [30] but defined at a broader level of organization [31]. EFTs incorporate multiple functioning properties in the classification related to different aspects of the exchange of the energy and matter of ecosystems [23], this allows characterizing ecosystem heterogeneity and responses to spatio-temporal environmental change.

Satellite remote sensing is the only practical means by which land surface biogeophysical variables can be obtained at the temporal and spatial scales required for global and regional use [30]. Several time series of remotely sensed images are available at coarse and medium spatial resolution. In particular, Normalized Difference Vegetation Index (NDVI)-based time series are fundamental to the remote sensing of vegetation phenology and to extract numerical observations related to vegetation dynamics [27]. However, NDVI has shown to saturate over forested regions, while being sensitive to canopy background variation in arid and semi-arid areas [32]. Therefore, additional information is required to complement NDVI information and compensate for this weakness [33]. Nevertheless, a few attempts have been made to integrate additional information into ecosystem monitoring, mainly including indicators of energy partitioning into sensible and latent heat fluxes and light reflection [34].

Accurate and efficient EFTs mapping via remote sensing depends largely on the selection of an appropriate classification approach [8]. Different classifiers may produce different results even when the same functional variables are used [35]. Many classifiers have been proposed to delineate EFTs such as ISODATA [23], CLARA [34] and partitioning based on thresholds [16,36]. A promising alternative to these methods is the self-organizing map (SOM) [37]. The SOM is an efficient way to derive maps from multi-dimensional and complex data. This method allows showing the data in a more comprehensive fashion and in fewer dimensions [38]. The detection of useful knowledge (in the form of patterns, structures and relationship) within the raw data is enhanced. Furthermore, the SOM method has overcome some of the limitations of the clustering methods since (i) they are more robust and are not significantly affected by missing data, (ii) do not require any prior assumption about the underlying distribution of the data, (iii) are capable of generalizing in noisy environments and (iv) make it possible to learn complex patterns in a limited time [39] with a decreased computational cost, also being an appropriate tool when the sample size is large [40]. Although this method has been shown to be more successful than the above mentioned algorithms to monitor and classify land cover [41] and has been previously used in ecology [42,43] and remote sensing [44], the potential of this methodology has not yet been applied to delineate EFTs.

The aim of this study is to identify EFTs in Spain for the period 2000–2009 as areas exhibiting similar responses to environmental conditions and similar ecosystem processes. The remarkable landscape diversity in Spain makes this area a suitable region to identify the EFTs. This work follows preliminary investigation on the same topic [16], which revealed the interest and potential of the EFTs classification encouraging the authors to further in-depth research. The current extended study uses images of higher

spatial resolution (1 km) compared with previous studies (e.g., [23]), and a robust methodology based on a two-level procedure, based on the combined use of the SOM and *k-means* clustering method to derive a map of EFTs definition. This method also allows expanding the number of variables involved in the definition of the EFTs to the extended use of the NDVI, and evaluates its performance. Additionally we analyze, for the different EFTs seasonal patterns and the correspondence between functional structural descriptors (*i.e.*, land cover) and ecoregions with the derived ecosystems.

2. Study Area

Despite its relatively small extent, 504,645 km^2, Spain has a remarkable landscape diversity and a wide range of ecosystems due to its relief, climate and geological features (Figure 1). This is the result of its geographic position, in the southwestern Europe, and an elevation that varies from sea level to 3479 m (Mount Mulhacén in Sierra Nevada, in the south) [36]. The central part of Spain is dominated by a vast plateau surrounded by a number of mountain ranges (Sistema Central, Sierra Morena, the Cordillera Cantábrica and the Sistema Ibérico). Other mountain ranges include the Pyrenees in the north, Sistema Penibético in the southeast and Sierra Nevada in the south.

Figure 1. Study area with the localization of the regions, the Basin Rivers and the mountains.
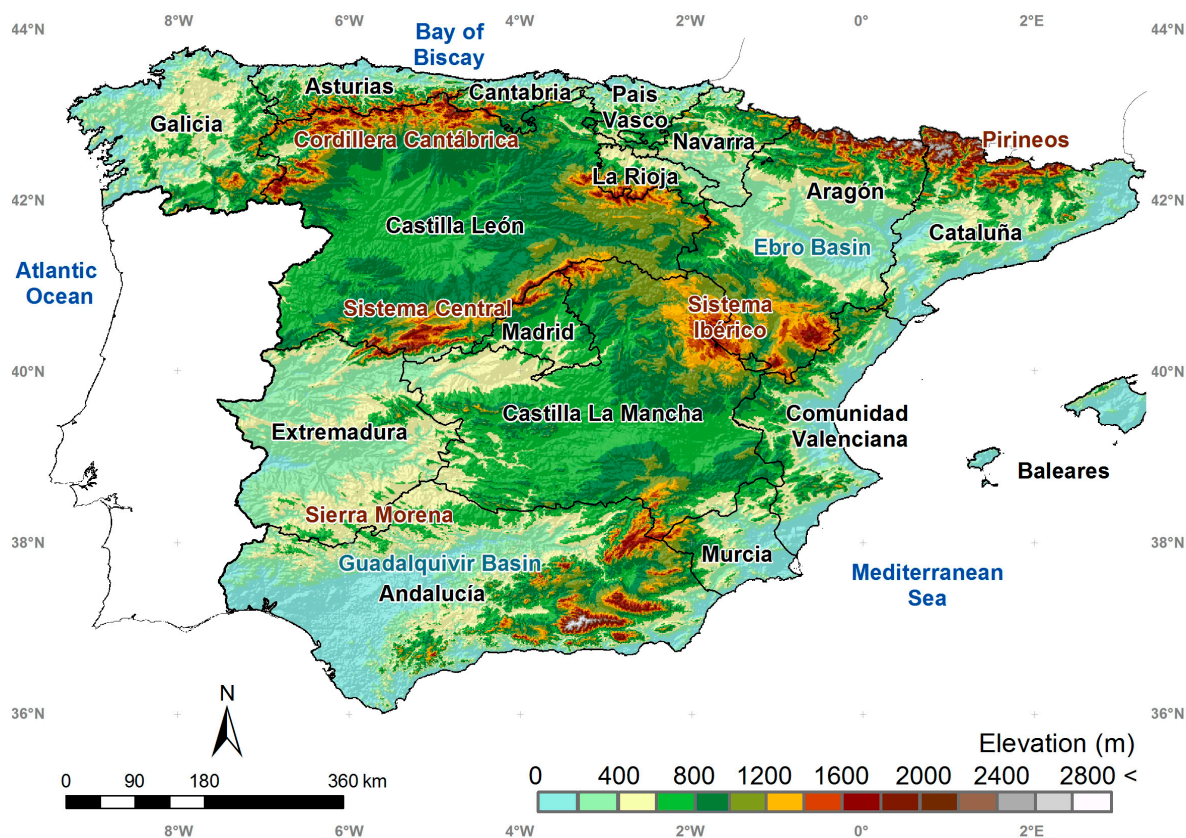
Synoptic air masses create a NW to SE gradient of water availability (total annual precipitation ranges from 2000 mm to 120 mm), ranging from Atlantic humid climate zones on the north coast to Mediterranean climate on the east coast, the SE corner becoming the most arid zone of Europe [45].

Spain has a mosaic of land cover that includes significant areas of traditional and newly developed agriculture (49% of land), dominated by non-irrigated land. These areas are embedded in a matrix of

natural and semi-natural vegetation (47% of land), mainly occupied by shrublands, and broadleaved or coniferous forests. Vegetation distribution is strongly controlled by aridity, and drought conditions have a marked influence on vegetation cover and activity. In particular, the droughts that affect the semiarid areas represent an important factor in environmental degradation, because they limit the development of vegetation cover and make the soil more sensitive to erosion by wash off from intense precipitation [5].

3. Data Set

3.1. Normalized Difference Vegetation Index (NDVI)

NDVI provides an index of ecosystem function since it has proven to be a proxy for the status of the aboveground biomass at the landscape level due to the high correlation with green-leaf density, net primary production and the fraction of photosynthetically active radiation absorbed by green biomass and CO_2 fluxes [46].

The 16-day Terra MODIS (MOD13A2) NDVI product at 1 km spatial resolution has been used. MOD13A2 ingest level two daily surface reflectance product (MOD09 series), which provides red and near-infrared surface reflectance corrected by the effect of atmospheric gases, thin cirrus clouds and aerosol [47,48]. The MOD13A2 vegetation product contains a data quality assessment (QA data) product holding information on overall usefulness and cloud conditions on a per-pixel basis [49].

3.2. Land Surface Temperature (LST)

LST is an indicative variable of the net surface energy balance driven by long-wave radiation surface emission [50]. LST is one of the key parameters in the physics of land-surface processes, combining surface-atmosphere interactions and the energy fluxes between the atmosphere and the land surface [51] and influencing processes such as evapotranspiration and vegetation stress [52]. The Terra MODIS 8-day LST with 1 km resolution (MOD11A2, collection v005) was used. The MODIS land surface temperature is derived from two thermal infrared band channels, *i.e.*, 31 (10.78–11.28 μm) and 32 (11.77–12.27 μm) using the split-window algorithm [53] which corrects atmospheric effects and emissivity using a look-up table based on global land surface emissivity in the thermal infrared band. An accuracy assessment across a wide set of test sites indicates an accuracy better than 1 km with a root mean square (RMS) less than 0.5 K in most cases [54].

3.3. Albedo

Albedo is a key geophysical parameter in ecosystem characterization since it indicates the amount of solar energy absorbed by the surface and therefore the surface energy budget [55]. A large number of studies recognize the important role of Albedo influencing gross productivity [56], surface temperature [57], evapotranspiration [58] and physical, physiological and biogeochemical processes [59].

The Albedo information was obtained from the 16-day MODIS MCD43B3 combined Aqua and Terra products [60,61]. It provides a 1 km data describing both the directional hemispherical reflectance (black-sky Albedo) and the bihemispherical reflectance (white-sky Albedo). The overall surface Albedo was derived by averaging the above two estimates.

Black-sky and white-sky Albedos are provided for seven spectral bands (MODIS channels 1–7) and the three broadbands (0.3–0.7, 0.7–3.0, and 0.3–5.0 μm). These products make use of a kernel-driven linear model of the Bidirectional Reflectance Distribution Function (BRDF), which relies on the weighted sum of an isotropic parameter and two functions (or kernels) of viewing and illumination geometry. The accuracy assessment against continuous field measurements results in a root mean square error (RMSE) of 0.013 and a bias around −0.02 [62].

3.4. Precipitation Data

Climatological data such as precipitation and temperature control differences in the Earth's vegetation cover, affect growth rate, plant reproduction, and frost damage. Monthly precipitation maps were derived at 2 km spatial resolution. The accurate estimation of the spatial distribution of precipitation requires a very dense network of measuring gaugement. The climatic data used in the study was obtained from the AEMet (Agencia Estatal de Meteorología) and corresponds to between 4000 and 7000 (depending on the period) recording stations distributed around Spain. Geostatistical approaches [63,64], such as ordinary kriging, were chosen for interpolation since they provide a measure of prediction error and allow observations to be complemented by secondary attributes that are more densely sampled. Monthly and annual averages for the period analyzed were computed from the precipitation monthly maps.

4. EFTs Classification Scheme

EFTs were identified following a multi-step approach (Figure 2). First, the time series of LST, Albedo and NDVI were pre-processed and filtered through a Savitzky-Golay [65] filter to smooth out noise caused mainly by cloud contamination, atmospheric variability, and bi-directional effects (Section 4.1). Second, different metrics related to vegetation phenology and ecosystem functioning (Section 4.2) were derived from these series and their contribution to discriminate EFTs was explored. Third, a SOM procedure was applied to extract and visually display the topological structure of high-dimensional input data (Section 4.3). Finally, the *k-means* method was carried out based on the SOM prototypes, with the aim of partitioning the input data into EFTs (Section 4.4). A more detailed description of the methodology follows.

4.1. Time-Series Filtering

Although the MODIS data are 8 or 16 day Maximum Value Composite (MVC) products [66], these still include noise that can affect the time series signal [67]. For this reason, spectral data was smoothed using a Savitzky-Golay filter to mitigate the effects of sensor noise and more efficiently reduce contamination [68]. This method is based on simple polynomial least-square calculations and has showed effectiveness in eliminating the cloudy values and exceptional data in MODIS time-series data [69]. However, instead of fitting a least-square curve to the total length of spectrum all at once, the method fits the spectral data piece-by-piece with the size equal to a user-defined value (*i.e.*, filter size) using a special form of matrix calculations [65]. This requires two key parameters: the filter size and the degree of polynomial orders. The degree of the smoothing polynomial ranges from 2 to 4. A smaller value will produce a smoother result but introduce more bias; conversely, a higher value will reduce filter bias,

but may "over fit" the data and give a noisier result [68]. In this study, it was found that a combination of a filter size of 5 and a degree of polynomial orders of 3 provided the best result in most cases.

Figure 2. Flowchart of the methodology used to derive the ecosystem functional types (EFTs).
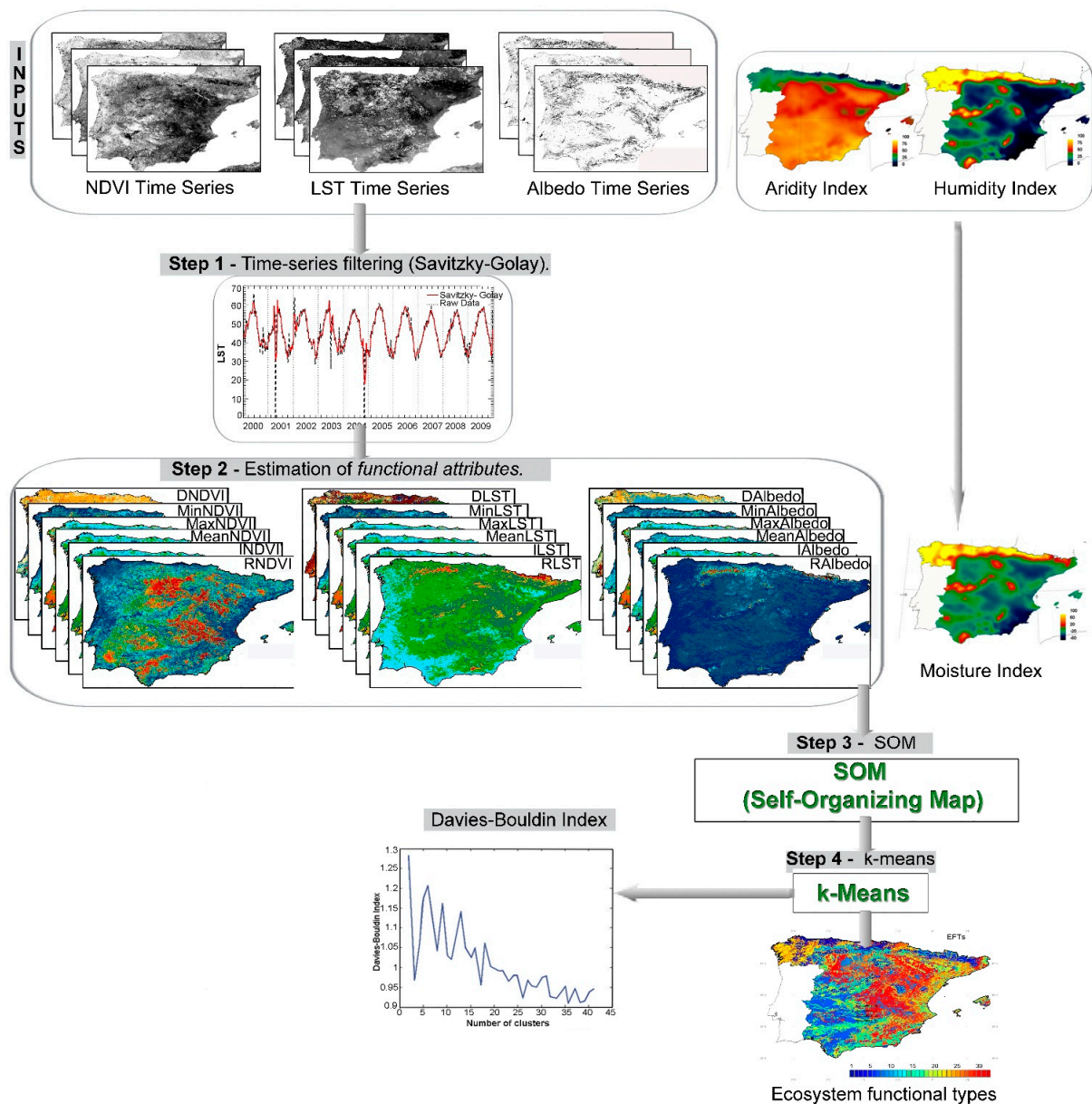
4.2. Estimation of Functional Attributes

In order to summarize the information provided by the NDVI, LST and Albedo time series, several metrics of the annual average curve for the period 2000–2009 have been derived for every pixel of the image composite. The metrics used in this paper include: maximum value (Max), minimum value (Min), mean value (Mean), annual integral (I) as the area under the seasonal curve, date of maximum (D) as the date when the maximum value occurs, relative range (R) as the difference between maximum value and minimum value divided by the annual integral. A total of eighteen functional attributes have been derived, six for each of the NDVI, LST and Albedo series. These cumulative remote sensing indices or metrics [70] have been reported to capture significant features of ecosystem functioning. In particular,

the annual integral of NDVI (INDVI) shows the ecosystem's long-term capacity to sustain biomass whereas the date of maximum NDVI (DNDVI) shows its resilience for recovering from disturbances, mainly rainfall fluctuations [45]. To keep the continuous nature of the annual period and the relative distance between months (*i.e.*, December is as close to February as May is to July), we transformed months into polar coordinates to compute the date of maximum. Months were therefore characterized by their sine and cosine values (30° for January and 360° for December).

Additionally, the Thornthwaite moisture index was computed as the difference between monthly precipitation (P) and potential evapotranspiration (PET) [71]. This climate index is a useful indicator of the supply of water in an area relative to the demand for water (P) under prevailing climatic conditions (PET). Therefore, P-PET is used as an estimate of the atmospheric balance between water supply and water demand, which reflects the portion of total precipitation used to nourish vegetation over a certain area. To derive PET we apply the Thornthwaite method [72] to monthly precipitation maps (2000–2009 period) from the AEMet dataset. The budget between PET and rainfall yields the yearly humidity and aridity indices. The linear combination of them produces the moisture index.

4.3. Kohonen Self-Organizing Map (SOM)

The self-organizing map (SOM) is an unsupervised, non-parametric and competitive neural network model used to reduce data dimensionality in non-linear data processing and which allows identifying groups of observations with similar patterns [37,73].

The SOM (Figure 3) is composed of two layers: the input and output layer. The input layer represents the input feature vector and has as many neurons (n) as it has variables (*i.e.*, functional attributes). The output layer or competitive layer is an array of neurons or nodes (M) spatially structured in a regular lattice ($m \times m$), usually in a two-dimensional grid. Each output layer neuron is connected to the n components of the input layer x (*i.e.*, $x_1, \ldots, x_n$) by a synaptic weight w_{ij}. An n-dimensional vector of synaptic weights (prototype) (w_j) is associated with each neuron j in the grid, where n refers to the dimensionality of an input data pattern [74].

Figure 3. Schematic layout of a self-organizing map adapted from [74].
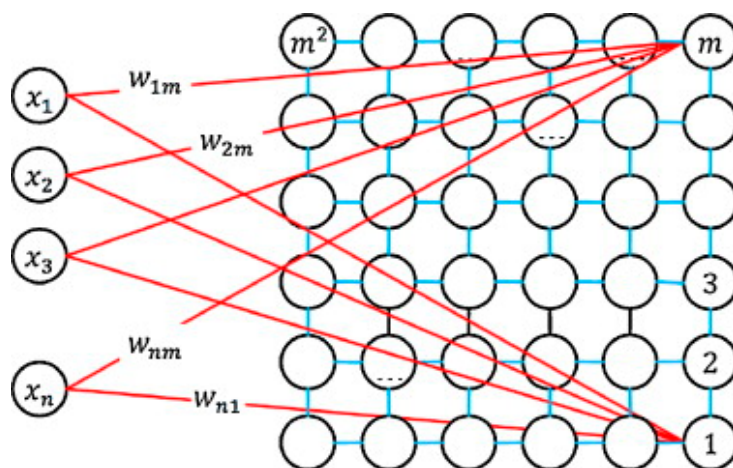
The SOM is trained iteratively, which could be summarized in four basic steps (for more in-depth analysis of the algorithm, refer to [41]):

(1) *Initialization:* the *M* weight vectors are initialized to a random value between 0 and 1.

(2) *Competition:* each input vector in the training set, *x*, is compared with each weight vector, w_j, to determine the neuron *j* which is the closest in terms of distance. This winning output node or neuron *c(j)*,which is called Best-Matching Unit (BMU), is typically computed using the minimum-distance Euclidean through the following rule:

$$c(j) = \arg \min_{1 \leq j \leq M} \left\{ \left\| x - w_j \right\| \right\} i = 1, ..., N \tag{1}$$

(3) *Cooperation:* in this stage it is necessary to define a neighborhood function that allows to identify the output nodes close to the BMU, *c(j)*, to be updated in the next step.

(4) *Updating:* the weight vector of neurons close to the BMU, as well as the weight vector of the BMU itself, are updated according to:

$$w_{ij}(t+1) = w_{ij} + \alpha(t) h_{ci}(r(t)) [x(t) - w_{ij}(t)] \tag{2}$$

where *t* denotes time, $\alpha(t)$ is learning rate with an initial value between 0 and 1, $h_{ci}(r(t))$ denotes the neighborhood kernel around the winner unit *c*, with neighborhood radius *r(t)*.

This update procedure stretches the winning neuron and its topological neighbors towards the sample vector. Neighboring neurons are pulled in the same direction, and thus weight vectors of neighboring neurons resemble each other for adopting competitive learning rules and neighborhood (the topological neighborhood relationships). Neurons with the same properties in the competitive layer would be close to each other, and the others would be far away when the clustering process is finished [41].

(5) *Repetition*: Repeat steps 2 and 3 until the network convergence, where the learning rate and neighborhood decrease monotonically with time.

Before applying the SOM algorithm, variables were standardized in order to ensure that all variables had the same *a priori* relevance for the model. Several parameters in the SOM need to be set up manually, e.g., number of neurons, and thus were evaluated in order to apply the better ones. The quality of the results is measured through quantization (*Qe*) and the topographic error (*Te*) [38]. *Qe* is the average Euclidean distance between each data vector and its BMU [75]. This error measure evaluates the fitting of the neural map to the data. *Te* is a measure of the continuity of the mapping [76], which measures the proportion of all data vectors for which first and second BMU are not adjacent vectors.

4.4. K-means Classification

K-means algorithm was applied to the prototypes obtained during the SOM learning to determine the clusters inherent in the structure of the data at the SOM's output layer. *K-means* [77] consists in a self-organized identification process of initial and arbitrary group means or centers, which are specified for each one of the *k* clusters. Each pixel is then assigned to the nearest cluster center in an iterative way using the minimum distance criterion. In each iteration the center of each cluster is updated by using the average vector of the pixels assigned to the cluster, unless a standard deviation or distance threshold is specified or there is no significant change in pixel assignments from one iteration to the next. Hence, the objective is to minimize a measure of dispersion within the clusters and to maximize the distance between clusters.

Selecting the optimal number of clusters is a critical issue because there are no predefined classes. To avoid subjectivity in delimiting the number of EFTs (clusters), different sets of validity indices have been proposed (e.g., Silhouette, Dunn's, Calinski Harabasz indices) [78]. In this paper, the Davies-Bouldin index [79], which is a function of the ratio of the sum of within-cluster scatter to between-cluster separation, has been selected because is less sensitive to the position of a small group of data set members and the expected behavior in case of more than two clusters is good [80].

5. Results and Discussion
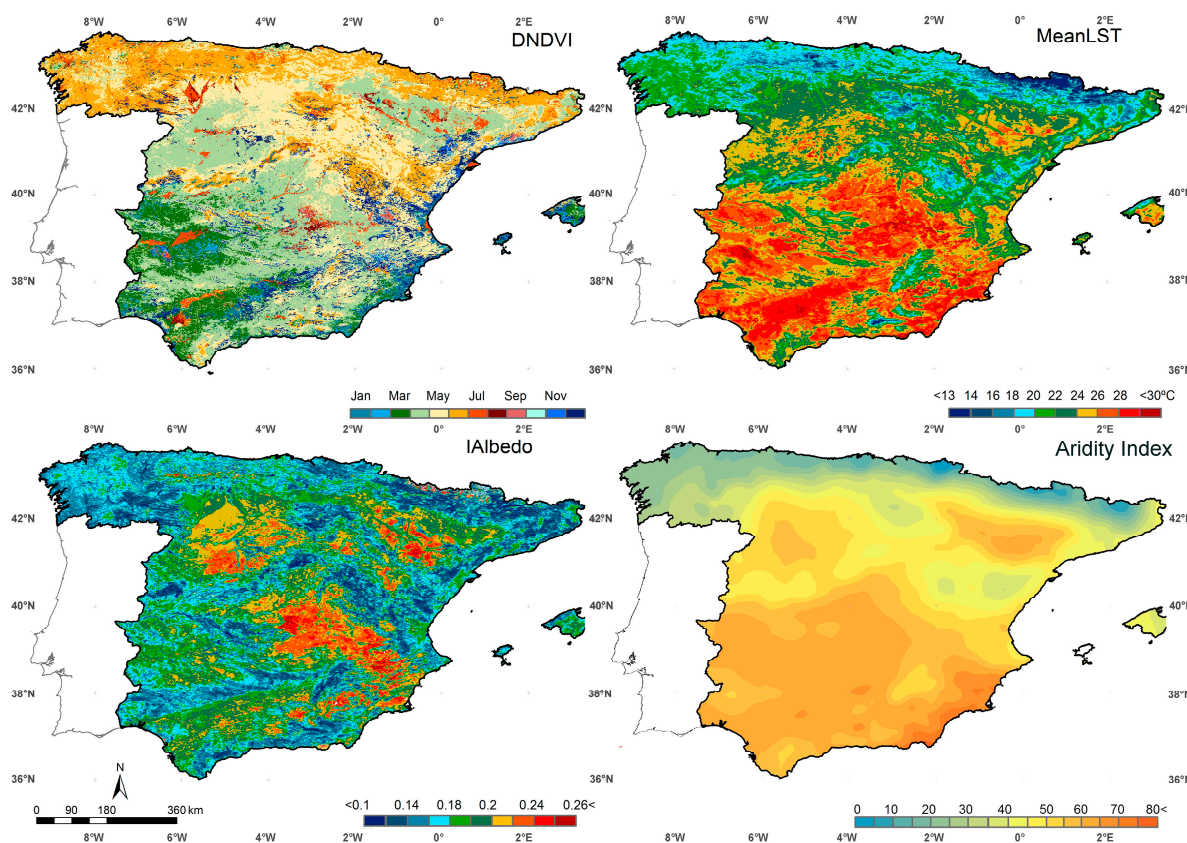
5.1. Examples of Functional Attributes

The analysis of the functional variables shows clear and contrasted patterns, allowing itself to glimpse the main functional traits underlying the ecosystems in Spain. Figure 4 shows some examples of key functional attributes (e.g., DNDVI, MeanLST, IAlbedo) used in this study to derive the EFTs. The date of the maximum NDVI (DNDVI) is associated to the phenological cycle of each vegetation type. Pixels showing a spring-peaking vegetation (April) are located in Castilla y León and Castilla-La Mancha, and mainly covered by non-irrigated crops such as wheat (*Triticum spp*). Areas with DNDVI in March correspond to *dehesas* grassland (Extremadura) where the herbal growth (*Oxalis sp.*) that recovers the orchard canopies during winter is eliminated, by tilling, in summer. Summer-peaking vegetation (July–August) is generally located in irrigated areas such as the Ebro basin, Vegas Del Guadiana and Tiétar and the Alagón-Árrago basin. Summer-peak also occurs in rice paddies located in the Ebro River Delta, around the Albufera Lagoon and in the Guadalquivir River marshes. Forested areas, in the north of Spain, also showed high values during June–July (orange colors) whereas winter-peaking mainly corresponds to citrus crops along the Mediterranean coastline (blue colors).

Orography, together with other geographic variables such as latitude or longitude, plays a significant role in the distribution of MeanLST. This functional attribute reveals a progressively hotter gradient when moving from north to south. Lowest MeanLST values (around 18 °C), are mainly located in the northern forested areas of Spain. Several patches of cooler areas are also located in the south, such as Los Alcornocales (*Quercus suber*) Natural Park in the Sierra del Aljibe (Cádiz). This is due to the effect of the mountains that intercept moisture from SE-prevailing winds coming directly from the Mediterranean that diminish the severity of drought in summer. Extreme values of MeanLST, about 30 °C, correspond to areas of sparse vegetation such as Tabernas Desert in the southeast characterized by high aridity and low moisture index, as reported by previous studies [45]. These areas have a period of drought during the summer with a low vegetative productivity due to the connection of LST with the length of the vegetative period and evapotranspiration [81].

Results in the IAlbedo confirm that the land surface albedo of any geographic location varies depending on the predominant land cover type, together with other variables such as the density and structure of land cover [82]. The highest IAlbedo values are located in non-irrigated croplands (Castilla y León, Castilla-La Mancha and Andalucía) because the albedo of these types of crops is mainly controlled by the reflection of the bare soil. The albedo generally decreases as one moves from grassland (Extremadura) or cropland (Castilla) to more dense vegetation such as broadleaved or needleleaved forests (north of Spain).

The spatial distribution of the averaged Aridity Index, which represents a measure of water availability in the ecosystem, is related to the relief and a longitudinal gradient, which determines the spatial distribution of precipitation and temperature. Hence, aridity is more intense in the south of Spain, where low precipitation and high temperatures are recorded [83].

Figure 4. Example of several functional variables (Date of maximum Normalized Difference Vegetation Index (DNDVI), Normalized Difference Vegetation Index (NDVI), Mean Land Surface Temperature (MeanLST), IAlbedo, Aridity Index).

5.2. Training SOM

The SOM component planes of the functional attributes are displayed in Figure 5. This method, allows visualizing multiple variables simultaneously and their relationship, being and efficient way to analyze the effect of each variable on the patterning input dataset. The 19 subfigures are linked by position: in each subfigure, the hexagon in a certain position corresponds to the same map unit, and the color of each hexagon informs on the value of the component in that neuron—warm colors for high values and cold colors for low values.

Figure 5. Visualization of the two-plane components of the Self-Organizing Map (SOM) trained from time series of NDVI, Albedo, LST and climatic data of Spain.
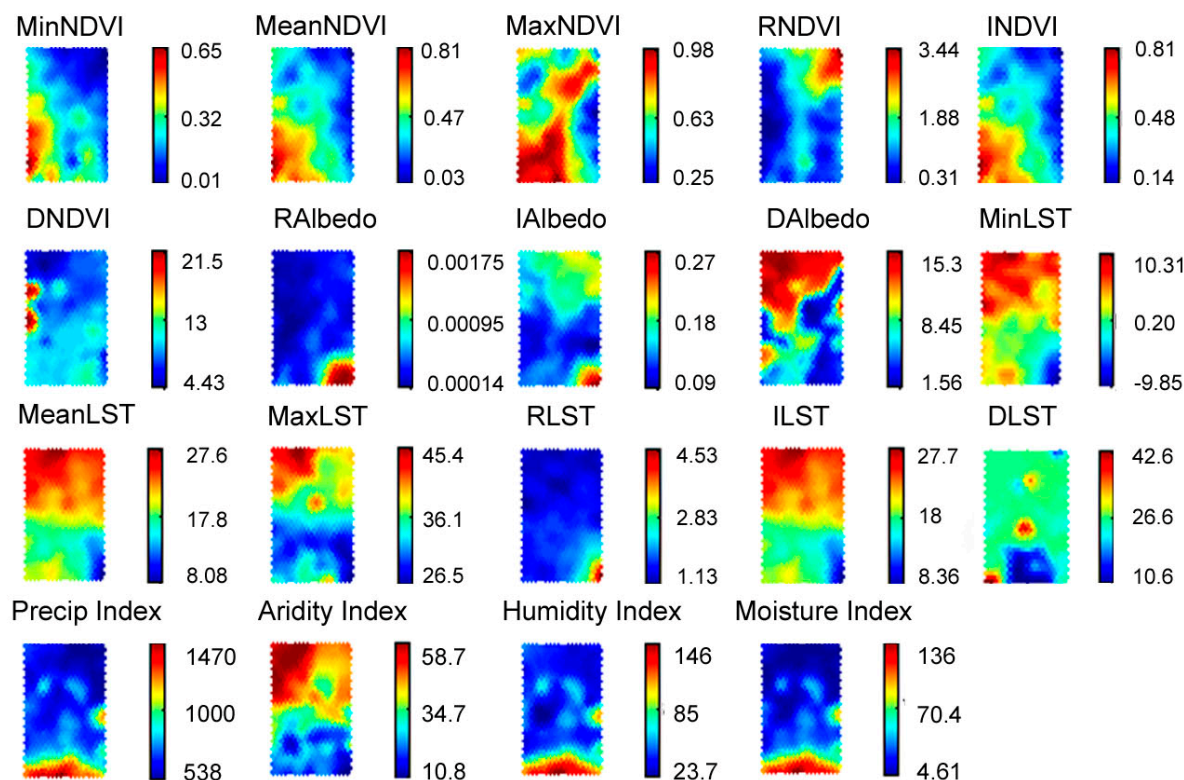

Comparing the color gradient of the component planes some correlation patterns among the variables could be discerned, indicating a number of general features summarized below. First, multiple variables exhibit a similar color indicating a strong positive correlation, such as precipitation, humidity and moisture indices, demonstrating that a decrease in the precipitation encompasses a decrease of moisture index [84]. Moreover, the integrals and mean values of variables (e.g., MeanLST and ILST) also show the same pattern, confirming that the information provided is similar and corroborating that SOM patterns make physical sense. Hence, these variables have the same relative influence in the characterization of the ecosystem function. For this reason, and to avoid redundancy, only one of the related variables in each category may be needed to adequately describe the ecosystem.

Second, several attributes have a strong negative correlation (same pattern but opposite colors). For example, between aridity index and precipitation as previously pointed out in the Mediterranean region [85], especially during the warm season.

Third, several attributes present very homogeneous patterns, so these variables contribute little to discriminate among EFTs, e.g., RLST and RAlbedo.

Figure 6 shows the final variables chosen for deriving EFTs based on the SOM maps presented in Figure 5 and following two criteria: (i) avoid variables which are redundant (*i.e.*, MeanLST and ILST) and (ii) include all the representative attributes involved in ecosystem functioning (e.g., attributes concerning the biomass production (NDVI) and heat fluxes (LST)). One of the most appealing characteristic of SOM is its faithful representation of the original data, e.g., two data points are represented close to each other in the resulting map when they have similar features.

Figure 6. Visualization of the two-plane components of the SOM trained from time series of NDVI, Albedo, LST and climatic data in Spain used as input to derive the EFTS.
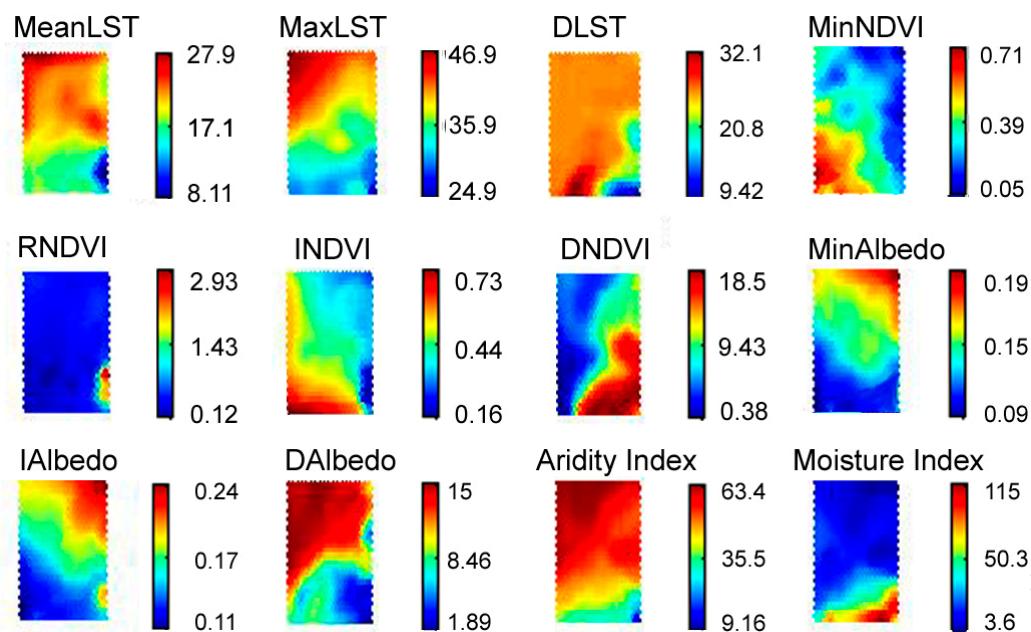
Several SOM neural network architectures were evaluated through the *Qe* and *Te* errors in order to optimize performance. The main discrepancies in these error parameters appear to regard the number of neurons. Since no strict rules to determine the optimal number of neurons exist, different map sizes were evaluated (Table 1) resulting in a mild drop of the *Qe* with increasing number of neurons. *Qe* is expected to decrease because a large number of dimensions better represent the sample; more neurons produce a better-input pattern because each data vector will be closer to its best matching unit. Nevertheless, a large number of neurons may cause over-fitting, while a small number can cause insufficient learning. *Te* represents the proportion of all data vectors for which first and second BMUs are not adjacent. The lower the *Te* is, the better the SOM preserves the topology. However, *Te* exhibits a tendency to increase with increasing SOM dimensionality due to a large number of non-neighbors in a large SOM matrix. Here the optimum size of 600 neurons was selected based on a compromise between *Qe* and *Te* [86].

Table 1. Map quality measures at different map sizes of the trained SOM.

| Map Size	100	200	300	400	500	600	700	800	900
Qe	2.08	1.82	1.67	1.57	1.44	1.39	1.35	1.31	1.30
Te	0.045	0.057	0.049	0.05	0.048	0.034	0.033	0.032	0.031

For the rest of the parameters, the best fit was achieved by those parameters proposed by [41]. Therefore, an hexagonal topology, a linear initialization with an initial rate of $\alpha(0) = 0.7$ and a batch training were selected. A Gaussian neighborhood function with linear descending rate was also used to adjust the weights of the neurons.

5.3. Clustering of the SOM

A *k-means* algorithm to estimate the optimal number of sub-structures (k) (EFTs) classified the prototype vectors from the SOM. Davies-Bouldin index was iteratively ran 100 times to minimize the possibility to find a suboptimal solution when the initial prototypes are not properly chosen and the sensitivity originate for the selection of the initial cluster [87]. According to the results (see Figure 7) the optimum number of EFTs was 36; nevertheless, other local solutions were detected (e.g., $k = 26$). In fact, the existence of several solutions is in agreement with the essence of the ecosystem heterogeneity where patterns can be recognized at different levels of complexity and spatial detail [34].

The final clustering partition is graphically displayed in the upper-right Figure 7. A detailed analysis of each cluster could be inferred from comparison with two-plane components (Figure 6). For instance, the upper-left cluster (red) was mainly characterize by medium INDVI (around 0.5) indicating medium biomass content, low RNDVI (0.4) that means a low intra annual variability, high MeanLST (around 25 °C) and high aridity index (around 62 °C).

Figure 7. (**Left-Upper**) Clusters obtained with the modified *k-means*, (**Right-Upper**) validity index values for Davies-Bouldin as a function of the number of clusters ranging from 2 to 36 and (**Down**) similarity among the dendrogram resulting from hierarchical agglomerative cluster analysis.
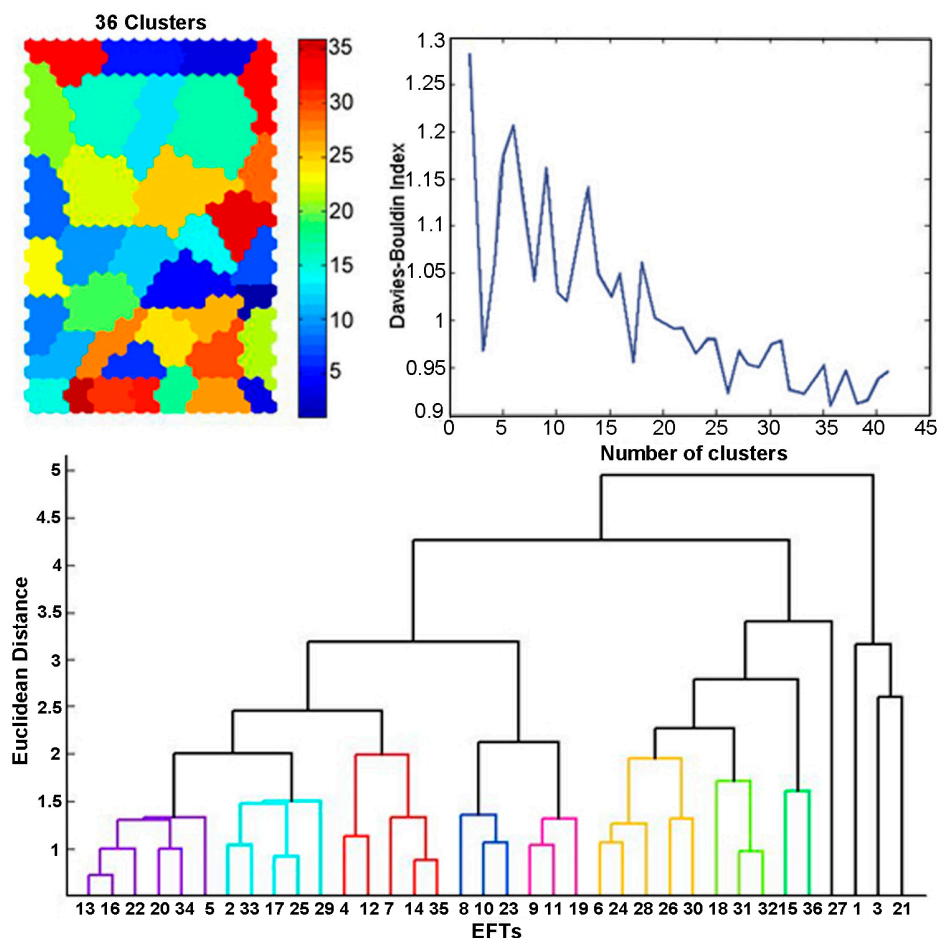
To facilitate summarizing and reporting results the 36 clusters were further grouped into 12 clusters based on dendrogram of a hierarchical cluster analysis. From the 36 clusters considered *a priori*, only representative ecosystems in Spain, which reached a minimum surface of 3%, were taken into account. A high cophenetic correlation of 0.85 confirmed the goodness of fit of the dendrogram. Several cut-point were evaluated, resulting that a distance of 2 provides a coherent number of EFTs. Similar to [34] the EFTs were numerated according the cluster similarities of the dendrogram; for example, EFT.1 and EFT.2 were closely related, whereas EFT.12 presents the highest differences.

5.4. Ecosystem Functional Types

Figure 8 shows the final 12 EFTs Map for Spain, excluding water and artificial surfaces (black color). Almost half of the territory is occupied by two ecosystems, EFT.1 (22.5% of surface) and EFT.2 (21.5% of surface). As mentioned above, these two ecosystems present the closest behavior (see Figure A—Appendix for supplementary material). EFT.1, mainly located in south-west (Extremadura and Andalucía regions), is characterized by high albedo means (0.16) due to the effect of the background, high LST (around 28 °C) and a moderate biomass (0.5 NDVI). According to Figure 7, EFT.1 encompasses high intra-ecosystem variability. This EFT could be divided into six subgroups in a more detailed analysis (see Figure B—Appendix), where each subgroup shows a marked spatial distribution, for example, the group 13 is mainly located in the sparse area of Almeria, whereas subgroup 34 is mainly located in the non-irrigated croplands of Andalucía.

EFT.2 is largely continuous and encompassing non-irrigated crops located in the center of Spain and is characterized by a high albedo (0.25) due to the strong influence of soil in these low dense canopy crops. Compared to EFT.1 this ecosystem shows a lower LST mean (around 25 °C) due to north south decreasing climatic gradient and a slightly lower NDVI mean NDVI (0.35). This ecosystem has a spring-peak of NDVI and a high intra-annual variability.

**Figure 8.** Distribution of ecosystem functional types (EFTs) in Spain.
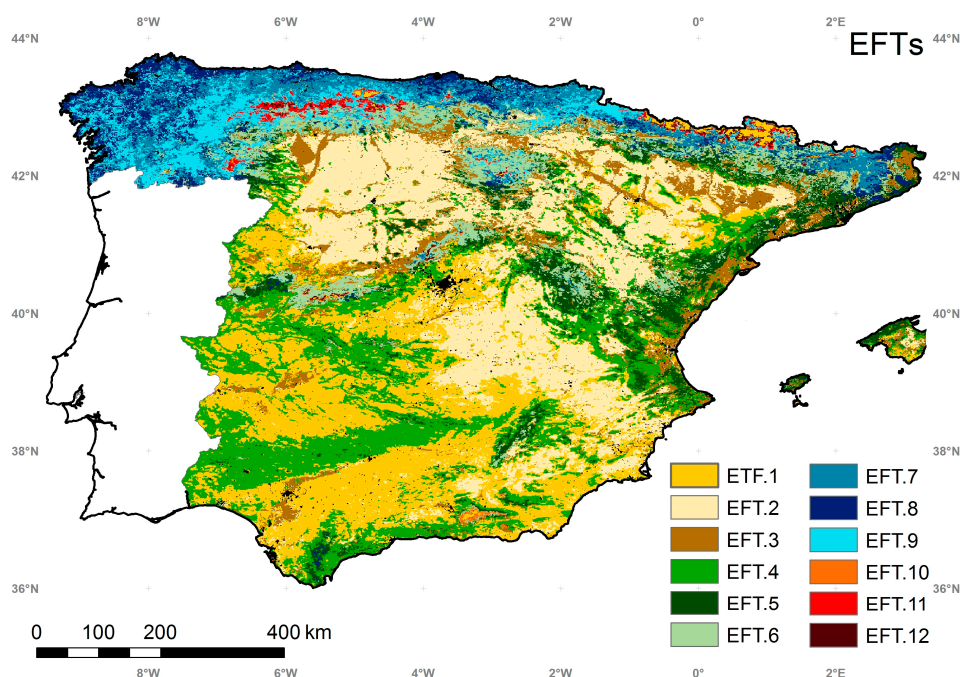
Higher intra-annual variability is also assigned to EFT.3, mainly composed by irrigated crops and rice paddies with a summer-peak and mean albedo values of 0.17. Orange trees and several areas of olive trees located in the eastern part of the Peninsula, with a winter-peak also belong to this EFT. This ecosystem illustrates how different land cover can have a similar ecosystem function. In this case, Albedo is the main ecosystem functional trait that conforms this ecosystem and differentiates it from other ecosystems. Similar to EFT.1, in a deeper analysis these subgroups (e.g., irrigated land-subgroup 7 and 4-and citrus-subgroup 12) exhibiting different NDVI patterns, can be analyzed separately.

EFT.7, EFT.8 and EFT.9 are located in the north, with a similar pattern in LST and Albedo means. The main source of variation in these ecosystems regards the NDVI pattern. Generally, these areas show low intra-annual variability and a DNDVI around June, however the surrogate of ANPP calculated as the NDVI mean indicates differences of productivity with values of 0.67, 0.72 and 0.63 respectively that are directly related with high amounts of annual precipitation. Conversely, EFT.10, EFT.11 and EFT.12 reveal low values of mean NDVI mean mainly located in mountainous areas such as the Pyrenees, and variable LST depending on the altitude.

To analyze the differences among EFTs, we examined the functional variation for each type according to independent variables such as the Gross Primary Production (GPP) [88] and the rate of Evapotranspiration (ETP). Seasonal differences among EFTs are produced (Figure 9). The rate of evapotranspiration increases in summer; however, the maximum varies from spring (e.g., EFT.1) to summer (e.g., EFT.9). The amount of GPP and ETP is also variable, with a direct relationship between the magnitudes. Thus EFTs characterized by a low PET presenting higher GPP values (e.g., EFT.8), and are mainly located in the north of the Peninsula. Conversely, EFT.1 and EFT.2, mainly located in the southeast correspond to the high PET and low GPP.

Figure 9. Mean seasonal profile of the Gross Primary Production (GPP) and the rate of Evapotranspiration (ETP) for the ecosystem functional types (EFTs).
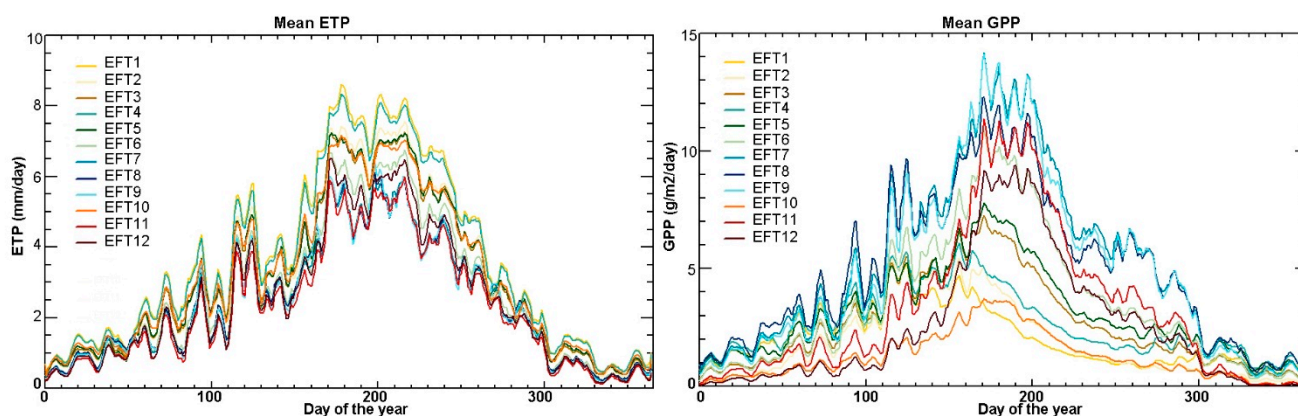
The contribution of each of the twelve functional attributes used in the identification of EFTs was analyzed through a multivariate separability measure such as Wilk's Lambda analysis [89]. This statistical is related to the likelihood ratio criterion and ranges between 0 and 1, where low values indicate a high contribution of the variable in the description of the ecosystem. According to the results (see Table A—Appendix), the Moisture Index (0.25), IAlbedo (0.28) and the Aridity Index (0.29) were the variables that presented a major contribution in the identification of ecosystems, whilst DAlbedo

(0.52), DLST (0.52) and the RNDVI (0.78) were the variables with a minor role. In order to determine the impact of all the variables considered *a priori* (Figure 5) the Wilk's lambda was computed again. The results presented similar contributions, with the lowest values relapsed in moisture, IAlbedo and the Aridity Index and where the highest values in Relative Ranges of LST, Albedo and NDVI.

5.5. Intercomparison of the EFTs Map with a Land Cover and Ecoregion Classification

To evaluate the final EFTs map and provide a further explanation, the map was compared with a hybrid land cover map derived from the combination of four datasets (CORINE, GLC2000, MODIS and GlobCover) [90] and with the ecoregions map [18]. The correspondence analysis (CA) [91] and the Minnick's coefficient of areal correspondence [92] between the EFTs and the compared maps were used for this purpose.

Figure 10 depicts the correspondence analysis of the EFTs with a land cover and an ecoregion classification by means of a bi-dimensional plot. On the left, the CA for the land cover solution accounts for 77.3% of the total inertia (62.2% and 15.1% for axes 1 and 2, respectively) that is a measure of variance of dispersion in the data. The degree of closeness of points in the CA is indicative of the relationships between the EFTs and land cover datasets, items similar to each other will tend to be near to each other in the graph.

Several clear associations are observed, such as EFT.7 with broadleaved forest and EFT. 3 with irrigated croplands, as have also been showed by [36]. Two ecosystem functional types (*i.e.*, EFT.1 and EFT.2) were specific of cultivated land cover, illustrating the high functional heterogeneity of this land cover [34] that can exhibit a multiple phenological response to environmental conditions, climate variability, plant communities and topography [13]. Moreover, the function describing the extent of overlap between two classes was analyzed by means of the Minnick's coefficient (*Cm*) as follows:

$$Cm = \frac{A \cap B}{A \cup B - (A \cap B)} \tag{3}$$

where *A* is the map of EFTs and *B* is the land cover or ecoregions map, respectively. $A \cap B$ and $A \cup B$ refer to the intersection and the union of *A* and *B* ,respectively.

Values higher than 0.05 are indicative of a moderate to high spatial association (see Tables B and C—Appendix). In the case of land covers, more than half of them are associated with an EFT, indicating that effectively there are a relationship between land cover and ecosystem function. Minnick's values support CA results and highlight several new associations. Strong values (>0.10) included the relation between Cultivated areas and EFT.2. EFT.3 showing a strong relationship with Irrigated land cover datasets. Minnick's values also show cases were a single land cover class could exhibit a multiple functional response such as needleleaved with EFT.4 (0.10) and EFT.5 (0.28). In contrast, there are ecosystems (*i.e.*, EFT.9) that are related to several land covers simultaneously (*i.e.*, Broadleaved (0.12), Mixed (0.10) and Cultivated areas (0.09)). We infer that the EFTs map provides unique information for augment available land cover datasets.

In the Ecoregions, the first two CA axes explained 84.9% of the variance with the EFTs. This higher value indicates that Ecoregions are more related to EFTs than land covers, as well as demonstrates the consistency of these two maps. Clear associations are shown in the CA as the Northeastern Mediterranean Forest and EFT.3. A single Ecoregion, such as the Cantabrian Mixed Forest, encompasses

EFT.7, EFT.8 and EFT.9 with 0.36, 0.26 and 0.37 Minnick's values respectively. Hence, the EFTs map complements this map, by providing an even finer level of resolution. In this case, eight ecoregions contain 12 EFTs in a broader level and 36 EFTs in case more detailed levels were considered. Regarding Minnick's values, higher degrees of overlap among ecoregions and ecosystems are found. Moreover, EFT.11 and EFT.12 that showed a lack of correlation with any land cover, presented a slightly relation with the Pyrenees Conifer ecoregion because these EFTs are mainly located in a specific spatial distribution.

Figure 10. Correspondence analysis between Ecosystem Functional Types (EFTs) and (**Left**) ecoregions (SISW: Southeastern Iberian Shrub and Woodlands, SISM: Southwest Iberian Mediterranean Sclerophyllous and Mixed Forest, ISSF: Iberian Sclerophyllous and Semi-deciduous Forest, CMF: Cantabrian Mixed Forest, PCM: Pyrenees Conifer and Mixed Forests, NIMF: Northwest Iberian Montane Forest, ICF: Iberian Conifer Forest, NSMF: Norhteastern Spain and Southern France Mediterranean Forest) and (**Right**) hybrid land cover (CUL: Cultivated and Managed areas, IRR: Irrigated Areas, MOS: Mosaic of cropland/natural, NEED: Needleleaved Forest, BROAD: Broadleaved Forest, MIX: Mixed Forest, SHR: Shrubland, HER: Herbaceous, SPA: Sparse Vegetation and BAR: Bare areas).
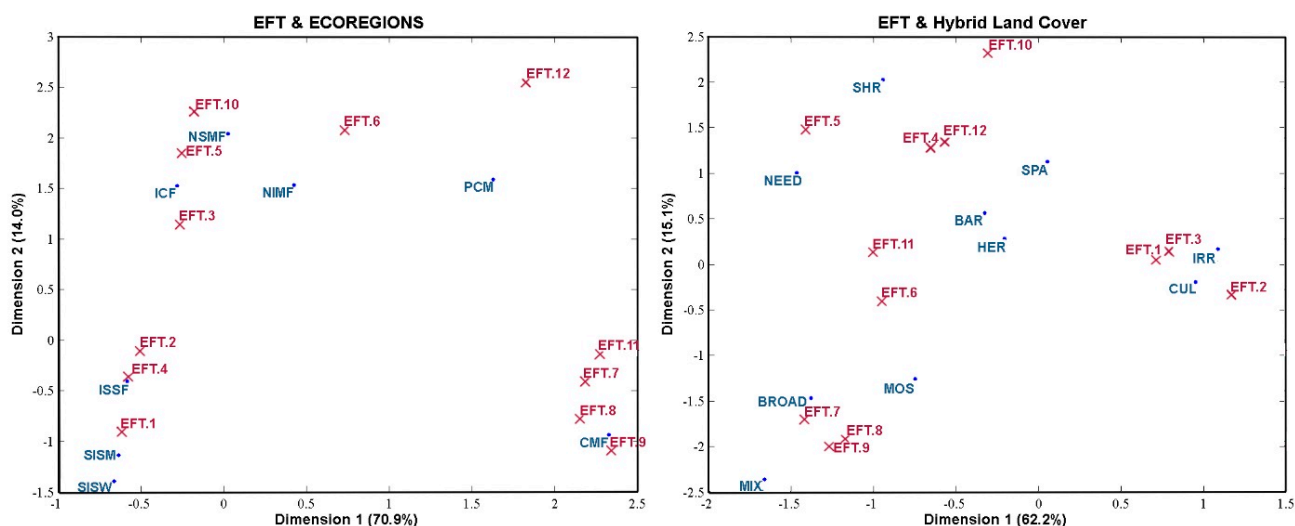
6. Discussion

Throughout the last decade, the role of ecosystem functioning in both environmental management and biodiversity conservation has significantly increased [92]. Ecosystem characterization results in a useful tool for several tasks, such as direct evaluation of ecosystem services, selection of protected areas, planning and monitoring ecological restorations among others.

Usually, the characterization of EFTs relies on estimators of the seasonal dynamics of carbon gains, such as the NDVI, EVI, fAPAR that have been proven to be useful parameters to characterize the ecosystem diversity [36]. Nevertheless, remote sensing provides additional functional variables to complement ecosystem functioning descriptors [93]. Hence, in this article we expanded the concept of ecosystems, adopting a more comprehensive method that extends the traditional concept of EFTs [23] and incorporates other variables related to energy partitioning into sensible and latent heat fluxes and light reflection that have only been tested in a local framework [34]. Moreover, we include climatic

variables, such as precipitation, that have previously been demonstrated to be a regional driver on the function of ecosystem [94], although they have never been included in the definition of EFTs. In fact, results from a Wilk's lambda analysis to determine which variable played the greatest role in the formation of the EFTs, revealed that climatic variables such as the Moisture Index (*i.e.*, Wilk's lambda 0.25) presented the main contribution in identifying the ecosystem functioning. Our findings support and reflect the outcomes of various studies addressing water-limited ecosystems, in which rainfall determines plant water stress and affects ecosystem function, renders it vulnerable to degradation and reduces its productivity [4]. The difference between maximum and minimum NDVI (RNDVI) revealed the lowest contribution in the definition of EFTs (*i.e.*, Wilk's lambda 0.79). This is in agreement with [24] that also found a low weight of this variable due to the bioclimatic control on the NDVI signal in the Mediterranean Iberia that underlies the variation in ecosystem responses observed.

The identification of EFTs and the analysis of the functional attributes by a SOM and *k-means* in this study highlights the power of this technique for interpreting highly functional variables and for identifying EFTs. Although this method has become a focus of particular interest across various ecological studies and has been applied successfully, as far as we are aware, this is the first time that SOMs have been applied to remotely sensed data to investigate ecosystem function. Therefore, this approach demonstrates an important new methodology for ecosystem characterization and analysis.

The SOM was used as a visualization tool to identify patterns in the dataset, allowed us to elucidate the interplay and relationships among difficult variables such as the functional types, where numerous biological and environmental factors are involved in a complex manner. For instance, high LST encompasses a high Aridity Index (Figure 5). This method allowed us to deal with a large amount of variables demonstrating its effectiveness in the handling of a large amount of variables. Comparing with classical techniques, that can only deal with accurate visualization of whole data sets when the number of features required is lower than three (e.g., RNDVI, INDVI, DNVI), we evaluated a high amount of variables that is a promising method of evaluating the effect of new attributes on ecosystem in the future.

On the other hand, the SOM along with the *k-means* technique made it feasible to classify behavioral patterns and organize behaviors in different classes according to functional type attributes. One of the main challenges using a SOM along with the *k-means* technique is the definition of the number of clusters that will best represent the ecosystem. To minimize the problem associated with the random selection of the centroid in the *k-means* algorithm, we used 100 runs. Data was patterned into 36 clusters that was gathered according to hierarchical agglomerative to facilitate summarizing results. However, this multi-level approach tolerates returning to a higher detail level if necessary. Each of the twelve clusters represented well, each individual characteristic in terms of the main functional attributes considered (*i.e.*, LST, NDVI and Albedo). In compliance with Appendix, we could postulate the twelve clusters as individual functional entities. Moreover, results from other independent attributes related to clime and biomass (*i.e.*, GPP and PET) also demonstrated the discrepancies among ecosystems (see results in Section 5.3) which confirms that they function differently.

EFTs direct validation was a non-trivial task due to lack of ground truth or *in situ* measurements. Moreover, the ground measures could not be not equivalent due to differences in how they quantify the functional information [95]. Hence, the assessment of the EFTs maps was carried out, as previous studies [13], by visual exploration of the map with other derived EFTs in the area and by comparing the EFTs map with land cover and ecoregions datasets.

The resulting EFTs map was consistent with preceding studies presented in [29] and [36], revealing a similar spatial distribution that decreases gradually southwards and eastwards in terms of biomass production. Although, our characterization of the heterogeneity in ecosystem functioning goes a step further increasing the spatial resolution from 8 km (e.g., GIMMS, AVHRR) to 1 km (e.g., MODIS). This is a benefit, mainly in more fragmented and complex landscapes or intensively used areas, as Spain, where the coarse spatial resolution is limited for describing the spatial variability of vegetation [29]. Hence, a 1 km spatial resolution improves the identification of that phenomenon. Moreover, these studies only rely on the classification according to vegetation such as NDVI, whereas our EFTs also considers the contribution of energy balance resulting in a more comprehensive characterization of ecosystems. For instance, Galicia was classified by [36] as a mainly continuous ecosystem with high INDVI and low RNDVI (Da1). This result is coherent with a previous test obtained from our methodology using only NDVI variables. However, including additional information such as LST allowed complementing the description and variability of the ecosystem mainly in forested areas where NDVI saturates a high level. Thus, LST in this area acts as main force to characterize the ecosystems and represents their variability in this patchy area.

To evaluate our EFTs we analyzed the relation between EFTs and land cover datasets and ecoregions. We therefore analyzed the relation between EFTs and land cover, with a correspondence analysis of 77.3% (two first axis), that as previous findings in a global framework [29], indicates that EFTs have potential in the monitoring of human influence on ecosystem functioning and in supporting degradation studios. In the case of Ecoregions, the correspondence analysis and the Minnicks' values between EFTs and the Ecoregions indicated that they are highly related, however the EFTs map contained additional information about energy fluxes and phenology that is not inherent in the in the biogeographic strata. For instance, the Iberian sclerophyllous region was highly associated with three ecosystems (EFT.1, EFT.2 and EFT.3), indicating that the EFTs present a more detailed classification that we suggest that is more appropriate to an ecosystem management.

Our methodology has been proven adequate to characterize EFTs and gives us an appropriate framework to integrate other variables. In this study, we have focused on a 10-year average functioning of the ecosystems. Nevertheless, this method is a promising tool to characterize ecosystems on an annual-basis and to evaluate changes in ecosystem services and goods.

7. Summary

Ecosystem functional types (EFTs) defined as areas exhibiting similar response to environmental conditions and similar ecosystem processes are considered as a useful proxy to characterize the ecosystem status. This paper elaborated and attempt to introduce and assess the performance of a refined methodology such as the Self-Organizing Map (SOM) combined to *k-means* to identify the EFTs based on several metrics derived from coarse resolution data for the period 2000–2009. The case study presented is considered a suitable area to test the new approach for two reasons; firstly, this is a framework for the regional scale with high heterogeneity as demonstrated in previous studies that were carried out for the same area. Secondly, the earlier studies conducted used low-resolution data such as AVHRR [36] or with high spatial resolution data (*i.e.*, Landsat) but in a local framework [34]. Hence,

MODIS data at 1 km spatial resolution have been used in this paper allowing increasing the detail in the identification of EFTs.

One of the novel contributions of the approach presented refers to the inclusion of new satellite-based metrics or attributes, rarely introduced in previous studies. These metrics allow us to expand the traditional definition of EFTs (based only on NDVI as an indicator of biomass) and including indicators of energy, water balance and amount of solar energy, making it more comprehensive for identifying multifunctional patterns associated to both biomass production and feedbacks on climate.

Another relevant contribution is the use of a robust methodology based on the SOM combined to *k-means* to delineate EFTs. This approach, together with the above-mentioned inclusion of new ecosystem functional variables, improves both the repeatability and interpretability of the ecosystem classification. Moreover, this method has reduced subjectivity in the process of class identification since requires no prior knowledge of the relationships between variables. The flexible approach described in this paper is particularly considered suitable for dealing with a high amount of input. Form the results, it is clear that the approach could provide a powerful tool to visualize and analyze high-dimensional data. From an ecological point of view, SOM appeared as a value tool in: (i) analyze the relationship among variables in an easy way, (ii) to select the functional metrics that better represent the ecological variability, by leaving out redundant information and (iii) to group areas according to the similarity of their functionality.

Acknowledgments

Author Contributions

This paper was made in collaboration with all the authors. They contributed equally to the preparation, analysis, review and editing of this manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Millenium Ecosystem Assessment. *Ecosystems and Human Well Being*; Island Press: Washington, DC, USA, 2005.
2. Swift, M.J.; Izac, A.M.N.; Van Noordwijk, M. Biodiversity and ecosystem services in agricultural landscapes—Are we asking the right questions? *Agric. Ecosyst. Environ.* **2004**, *104*,113–134.
3. Ayensu, D.V.; Claasen, M.; Collins, A.; Dearing, L.; Fresco, M.; Gadgil, H.; Gitay, G.; Glaser, C.L.; Lohn, J.; Krebs, R.; *et al.* International ecosystem assessment. *Science* **1999**, *286*, 685–686.

4.   Hill, J.; Stellmes, M.; Udelhoven, T.; Roder, A.; Sommer, S. Mediterranean desertification and land degradation: Mapping related land use change syndromes based on satellite observations. *Glob. Planet. Change* **2008**, *64*, 146–157.

5.   Martínez, B.; Gilabert, M.A. Vegetation dynamics from NDVI time series analysis using the wavelet transform. *Remote Sens. Environ.* **2009**, *113*, 1823–1842.

6.   Baldi, G.; Nosetto, M.D.; Aragón, R.; Aversa, F.; Paruelo, J.M.; Jobbágy, E.G. Long-term satellite NDVI data sets: Evaluating their ability to detect ecosystem functional changes in South America. *Sensors* **2008**, *8*, 5397–5425.

7.   Lambin, E.F.; Geist, H.J.; Lepers, E. Dynamics of land-use and land-cover change in tropical regions. *Annu. Rev. Environ. Resour.* **2003**, *28*, 205–241.

8.   Kaptué-Tchuente, A.; Roujean, J.L.; Faroux, S. ECOCLIMAP-II: An ecosystem classification and land surface parameters database of Western Africa at 1 km resolution for the African Monsoon Multidisciplinary Analysis (AMMA) project. *Remote Sens. Environ.* **2010**, *114*, 961–976.

9.   Fischlin, A.; Midgley, G.F.; Price, J.; Leemans, R.; Brij, G.; Turley, C.; Rounsevell, M.; Dube, P.; Tarazona, J.; Velichko, A. Ecosystems, their properties, goods, and services, In *Climate Change 2007: Impacts, Adaptation and Vulnerability: Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*; Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden, P.J., Hanson, C.E., Eds.; Cambridge University Press: Cambridge, UK, 2007; pp. 211–272.

10.   Henle, K.; Alard, D.; Clitherow, J.; Coob, P.; Firbank, L.; Kull, T.; McCracken, D.; Moritz, R.F.A.; Niemelä, J.; Rebane, M.; *et al.* Identifying and managing the conflicts between agriculture and biodiversity conservation in Europe—A review. *Agric. Ecosyst. Environ.* **2008**, *124*, 60–71.

11.   Cihlar, J. Land cover mapping large areas from satellites: Status and research priorities. *Int. J. Remote Sens.* **2000**, *21*, 1093–1114.

12.   Lhermitte, S.; Verbesselt, J.; Verstraeten, W.W.; Coppin, P.A. Comparison of time series similarity measures for classification and change detection of ecosystem dynamics. *Remote Sens. Environ.* **2011**, *115*, 3129–3152.

13.   Gu, Y.; Brown, J.F.; Miura, T.; Van Leeuwen, W.J.D.; Reed, B.C. Phenological classification of the United States: A geographic framework for extending multi-sensor time-series data. *Remote Sens.* **2010**, *2*, 526–543.

14.   Host, G.E.; Polzer, P.L. A quantitative approach to developing regional ecosystem classifications. *Ecol. Appl.* **1996**, *6*, 608–618.

15.   Waring, R.H.; Coops, N.C.; Fan, W.; Nightingale, J.M. MODIS enhanced vegetation index predicts tree species richness across forested ecoregions in the contiguous USA. *Remote Sens. Environ.* **2006**, *103*, 218–226.

16.   Pérez-Hoyos, A.; Martínez, B.; Gilabert, M.A.; García-Haro, F.J. A multi-temporal analysis of vegetation dynamics in the Iberian peninsula using MODIS-NDVI data. *EARSeL eProc.* **2010**, *9*, 22–30.

17.   Huang, C.; Kim, S.; Song, K.; Townshend, J.R.G.; Davis, P.; Altstatt, A.; Rodas, O.; Yanosky, A.; Clay, R.; Tucker, C.J.; *et al.* Assessment of Paraguay's forest cover change using Landsat observations. *Glob. Planet. Change* **2009**, *67*, 1–12.

18. Olson, D.M.; DInersteing, E.; Wikramanayake, E.D.; Burgess, N.D.; Powell, G.V.N.; Underwood, E.C.; D'Amico, J.A.; Itoua, I.; Strand, H.E.; Morrison, J.C.; *et al.* Terrestrial ecoregions of the world: A new map of life on Earth. *Bioscience* **2001**, *51*, 933–938.

19. Azzali, S.; Menenti, M. Mapping isogrowth zones on continental scale using temporal Fourier analysis of AVHRR-NDVI data. *Int. J. Appl. Earth Obs. Geoinf.* **1999**, *1*, 9–20.

20. Soriano, A.; Paruelo, J.M. Biozones: Vegetation units defined by functional characters identifiable with the aid of satellite sensor images. *Glob. Ecol. Biogeogr. Lett.* **1992**, *2*, 82–89.

21. White, M.A.; Hoffman, F.; Hargrove, W.W.; Nemani, R.R. A global framework for monitoring phenological responses to climate change. *Geophys. Res. Lett.* **2005**, *32*, 1–4.

22. Garbulsky, M.F.; Paruelo, J.M. Remote sensing of protected areas to derive baseline vegetation functioning characteristics. *J. Veg. Sci.* **2004**, *15*, 711–720.

23. Paruelo, J.M.; Jobbágy, E.G.; Sala, O.E. Current distribution of ecosystem functional types in temperate South America. *Ecosystems* **2001**, *4*, 683–698.

24. Alcaraz, D.; Cabello, J.; Paruelo, J. Baseline characterization of major Iberian vegetation types based on the NDVI dynamics. *Plant Ecol.* **2009**, *202*, 3–29.

25. Steneck, R.S. Functional groups. In *Encyclopedia of Biodiversity*; Levin, S.A., Ed.; Academic Press: San Diego, CA, USA, 2001; pp. 21–139.

26. Virginia, R.A.; Wall, D.H. Principles of ecosystem function. In *Encyclopedia of Biodiversity*; Levin, S.A., Ed.; Academic Press: San Diego, CA, USA, 2001; pp. 345–352.

27. Pettorelli, N.; Chauvenet, A.L.M.; Duffy, J.P.; Cornforth, W.A.; Meillere, A.; Baillie, J.E.M. Tracking the effect of climate change on ecosystem functioning using protected areas: Africa as a case study. *Ecol. Indic.* **2012**, *20*, 269–276.

28. Volante, J.N.; Alcaraz-Segura, D.; Mosciaro, M.J.; Viglizzo, E.F.; Paruelo, J.M. Ecosystem functional changes associated with land clearing in NW Argentina. *Agric. Ecosyst. Environ.* **2012**, *154*, 12–22.

29. Ivits, E.; Cherlet, M.; Mehl, M.; Sommer, S. Ecosystem functional units characterized by satellite observed phenology and productivity gradients: A case study for Europe. *Ecol. Indic.* **2013**, *27*, 17–28.

30. Sun, W.; Liang, S.; Xu, G.; Fang, G.; Dickinson, R. Mapping plant functional types from MODIS data using multisource evidential reasoning. *Remote Sens. Environ.* **2008**, *112*, 1010–1024.

31. Duckworth, J.C.; Kent, M.; Ramsay, P.M. Plant functional types: An alternative to taxonomic plant community description in biogeography? *Prog. Phys. Geogr.* **2000**, *24*, 515–542.

32. Huete, A.; Liu, H.W.; Batchily. W.; van Leeuwen, W. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sens. Environ.* **1997**, *59*, 440–451.

33. Julien, Y.; Sobrino, J.A. The Yearly Land Cover Dynamics (YLCD) method: An analysis of global vegetation from NDVI and LST parameters. *Remote Sens. Environ.* **2009**, *113*, 329–334.

34. Fernández, N.; Paruelo, J.M.; Delibes, M. Ecosystem functioning of protected and altered Mediterranean environments: A remote sensing classification in Doñana, Spain. *Remote Sens. Environ.* **2010**, *114*, 211–220.

35. Lu, D.; Lueng, Q.W. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870.

36. Alcaraz, D.; Paruelo, J.M.; Cabello, J. Identification of current ecosystem functional types in the Iberian Peninsula. *Glob. Ecol. Biogeogr.* **2006**, *15*, 200–212.

37. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biolog. Cybern.* **1992**, *43*, 59–69.

38. Kohonen, T. *Self-Organizing Maps*, 3rd ed.; Springer: Berlin, Germany, 2001.

39. Vesanto, J.; Himberg, J.; Alhoniemi, E.; Parhankangas, J. *SOM Toolbox for Matlab 5*; Helsinki University of Technology: Espoo, Finland, 2000.

40. Gonçalves, M.L.; Netto, M.L.A.; Costa, J.A.F.; Zullo, J.K. An unsupervised method of classifying remotely sensed images using Kohonen self-organizing maps and agglomerative hierarchical clustering methods. *Int. J. Remote Sens.* **2008**, *29*, 3171–3207.

41. Filippi, A.M.; Jensen, J.R. Fuzzy learning vector quantization for hyperspectral coastal vegetation classification. *Remote Sens. Environ.* **2006**, *100*, 512–530.

42. Chon, T.S. Self-organizing maps applied to ecological sciences. *Ecol. Inform.* **2011**, *6*, 50–61.

43. Schmidt, A.; Hanson, C.; Kathilankal, J.; Law, B.E. Classification and assessment of turbulent fluxes above ecosystems in North-America with self-organizing feature map networks. *Agric. For. Meteorol.* **2011**, *151*, 508–520.

44. Xia, Z.; Rui, S.; Bing, Z.; Qingxi, T. Land cover classification of the North Chinga Plain using MODIS_EVI time series. *ISPRS J. Photogramm.* **2008**, *63*, 476–484.

45. Del Barrio, G.; Puigdefabregas, J.; Sanjuan, M.E.; Stellmes, M.; Ruiz, A. Assessment and monitoring of land condition in the Iberian Peninsula, 1989–2000. *Remote Sens. Environ.* **2010**, *114*, 1817–1832.

46. Kerr, J.T.; Ostrovsky, M. From space to species: ecological applications for remote sensing. *Trends Ecol. Evol.* **2003**, *18*, 299–305.

47. Vermote, E.F.; Vermeulen, A. Atmospheric correction algorithm: Spectral reflectances (MOD09), 1999. NASA-Official Site. Available online: http://modis.gsfc.nasa.gov/data/atbd/atbd_mod08.pdf (accessed on 8 October 2014).

48. Huete, H.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213.

49. Fensholt, R.; Proud, S.R. Evaluation of earth observation based global long term vegetation trends—Comparing GIMMS and MODIS global NDVI time series. *Remote Sens. Environ.* **2012**, *119*, 131–147.

50. Benali, A.; Carvalho, A.C.; Nunes, J.P.; Carvalhais, N.; Santos, A. Estimating air surface temperature in Portugal using MODIS LST data. *Remote Sens. Environ.* **2012**, *124*, 108–121.

51. Mannsteint, H. Surface energy budget, surface temperature and thermal intertia. In *Remote Sensing Applications in Meteorology and Climatology*; Vaughan, R.A., Reidel, D., Eds.; Springer: Dordrecht, The Netherlands, 1987; pp. 391–410.

52. Coll, C.; Caselles, V.; Valor, E.; Niclós, R. Comparison between different sources of atmospheric profiles for land surface temperature retrieval from a single channel thermal infrared data. *Remote Sens. Environ.* **2012**, *117*, 199–210.

53. Wan, Z.; Zhang, Y.; Li, Z.; Wang, R.; Salomonson, V.V.; Yves, A.; Bosseno, R.; Hanocq, J.F. Preliminary estimate of calibration of the moderate resolution imaging spectroradiometer thermal infrared data using Lake Titicaca. *Remote Sens. Environ.* **2002**, *80*, 497–515.

54. Wang, W.; Liang, S.; Meyers, T. Validating MODIS land surface temperature products using long-term nighttime ground measurements. *Remote Sens. Environ.* **2008**, *112*, 623–635.

55. Román, M.O.; Schaaf, C.B.; Woodcok, C.E.; Strahler, A.H.; Yang, X.; Braswell, R.H.; Curtis, P.S.; Davis, K.J.; Dragoni, D.; Goulden, M.L.; *et al*. The MODIS (Collection V005) BRDF/albedo product: Assessment of spatial representativeness over forested landscapes. *Remote Sens. Environ.* **2009**, *113*, 2476–2498.

56. Sellers, P.J.; Hall, F.G.; Kelly, R.D.; Black, A.; Baldocchi, D.; Berry, J. BOREAS in 1997: Experiment overview, scientific results, and future directions. *J. Geophys. Res. Atmos.* **1997**, *102*, 28731–28769.

57. Dickinson, R.E. Land surface processes and climate surface albedos and energy-balance. *Adv. Geophys.* **1983**, *25*, 305–353.

58. Lawrence, D.M.; Slingo, J.M. An annual cycle of vegetation in GCM. Part II: Global impacts on climate and hydrology. *Clim. Dyn.* **2004**, *22*, 107–122.

59. Wang, S.; Grant, R.F.; Verseghy, D.L.; Black, T.A. Modelling carbon dynamics of boreal forest ecosystems using the Canadian Land Surface Scheme. *Clim. Change* **2002**, *55*, 451–477.

60. Jin, Y.; Schaaf, C.B.; Woodcock, F.; Gao, X.; Li, A.H.; Strahler, A.H.; Lucht, W.; Liang, S. Consistency of MODIS surface bidirectional reflectance distribution function and albedo retrievals: 2. Validation. *J. Geophys. Res.* **2003**, *108*, 1984–2012.

61. Liang, S.L.; Fang, H.L.; Chen, M.Z.; Shuey, C.J.; Walthall, C.; Daughtry C. Validating MODIS land surface reflectance and albedo products: Methods and preliminary results. *Remote Sens. Environ.* **2002**, *83*, 149–162.

62. Salomon, J.G.; Crystal, C.B.; Strahler, A.H.; Gao, F.; Jin, Y. Validation of the MODIS bidirectional reflectance distribution function and albedo retrievals using combined observations from the aqua and terra platforms. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1555–1565.

63. Matheron, G. Principles of geostatistics. *Econ. Geol.* **1963**, *58*, 1246–1266.

64. Cressie, N. The origins of kriging. *Math. Geol.* **1990**, *2*, 239–253.

65. Savitzky, A.; Golay, M.J.E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639.

66. Holben, B.N. Characteristics of maximum-value composite images for temporal AVHRR data. *Int. J. Remote Sens.* **1986**, *7*, 1435–1445.

67. Hird, J.; McDermid, G.J. Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sens. Environ.* **2009**, *113*, 248–258.

68. Chen, J.; Jönsson, P.; Tamura, M.; Gu, Z.; Matsushita, B.; Eklundh, L. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sens. Environ.* **2004**, *91*, 332–344.

69. Ren, J.; Chen, Z; Zhou, Q.; Tang, H. Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. *Int. J. Appl. Earth Obs. Geoinf.* **2008**, *10*, 403–413.

70. Wang, Q.; Zhao, P.; Ren, H.; Kakubari, Y. Spatiotemporal dynamics of forest net primary production in China over the past two decades. *Glob. Planet. Change* **2008**, *61*, 267–274.

71. García-Haro, F.J.; Belda, F; Gilabert, M.A.; Meliá, J.; Moreno, A.; Poquet, D.; Pérez-Hoyos, A.; Segarra, D. Monitoring drought conditions in the Iberian Peninsula using moderate and coarse resolution satellite data. In Proceedings of the 2nd MERIS/(A)ATSR User Workshop', Frascaty, Italy, 22–26 September 2008.

72. Thornthwaite, C.W. An approach toward a rational classification of climate. *Geogr. Rev.* **1948**, *38*, 55–94.

73. Kohonen, T. The self-organizing map. *Neurocomputing* **1998**, *21*, 1–6.

74. Ghaseminezhad, M.H.; Karami, A. A novel self-organizing map (SOM) neural network for discrete groups of data clustering. *Appl. Soft Comput.* **2011**, *11*, 3771–3778.

75. Xu, R.; Wunsch, D. Survey of clustering algorithms. *IEEE Trans. Neural Netw.* **2005**, *16*, 645–678.

76. Kiviluoto, K. Topology preservation in self-organizing maps. In Proceedings of the IEEE International Conference on Neural Networks, 1996, Washington, DC, USA, 3–6 June 1996; pp. 294–299.

77. MacQueen, J.B. Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistic and Probability, Berkeley, CA, USA, 21 June–18 July 1965, 27 December 1965–7 January 1966.

78. Maulik, U. Performance evaluation of some clustering algorithms and validity indices. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 1650–1654.

79. Davies, D.L.; Bouldin, D.W. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *1*, 224–227.

80. Petrovic, S. A comparison between the silhouette index and the Davies-Bouldin index in labeling ids clusters. In Proceedings of the 11th Nordic Workshop of Secure IT Systems, Linköping, Sweden, 19–20 October 2006.

81. Zavala, L.M.; González, F.A.; Jordán, A. Intensity and persistence of water repellence in relation to vegetation types and soil parameters in Mediterranean SW Spain. *Geoderma* **2009**, *152*, 361–374.

82. Benavides, R.; Montes, F.; Rubio, A.; Osoro, K. Geostatistical modelling of air temperature in a mountainous region of Northern Spain. *Agric. For. Meteorol.* **2007**, *146*, 173–188.

83. Liang, S. Narrowband to broadband conversions of land surface albedo I: Algorithms. *Remote Sens. Environ.* **2001**, *76*, 213–238.

84. Vicente-Serrano, S.M.; Cuadrat-Prats, J.M.; Romo, A. Aridity influence on vegetation patterns in the middle Ebro Valley (Spain): Evaluation by means of AVHRR images and climate interpolation techniques. *J. Arid Environ.* **2006**, *66*, 353–375.

85. Tabari, H.; Talaee, H.P. Moisture index for Iran: Spatial and temporal analysis. *Glob. Planet. Change* **2013**, *100*, 11–19.

86. Gao, X.; Giorgi, F. Increased aridity in the Mediterranean region under greenhouse gas forcing estimated from high resolution simulations with a regional climate model. *Glob. Planet. Change* **2008**, *62*, 195–209.

87. Kamimura, R. Supposed maximum information for comprehensible representations in SOM. *Neurocomputing* **2011**, *74*, 1116–1134.

88. Naldi, M.C.; Campello, R.J.; Hruschka, E.R.; Carvalho, A.C. Efficiency issues of evolutionary k-means. *Appl. Soft Comput.* **2011**, *11*, 1938–1952.

89. Moreno, A. Retrieval and Assessment of CO_2 Uptake by Mediterranean Ecosystems Using Remote Sensing and Meteorological Data. Ph.D. Thesis, University of Valencia, Valencia, Spain, 2014.

90. Wilks, S.S. On the independence of k sets of normally distributed statistical variables. *Econom. J. Econom. Soc.* **1935**, *3*, 309–326.

91. Pérez-Hoyos, A.; García-Haro, F.J.; San-Miguel-Ayanz, J. Conventional and fuzzy comparisons of large scale land cover products: Application to CORINE, GLC2000, MODIS and GlobCover in Europe. *ISPRS J. Photogramm. Remote Sens.* **2012**, *74*, 185–201.

92. Benzécri, J.P. L'analyse des correspondences. In *L'Analyse Des Données*; Dunod: Paris, France, 1973.

93. Minnick, R.F. A method for the measurement of areal correspondence. *Mich. Acad. Sci. Arts Lett.* **1964**, *49*, 333–343.

94. Cabello, J.; Fernández, N.; Alcaraz-Segura, D.; Oyonarte, C.; Piñero, G.; Altesor, A.; Delibes, M.; Paruelo, J.M. The ecosystem functioning dimension in conservation: Insights from remote sensing. *Biodivers. Conserv.* **2012**, *21*, 3287–3305.

95. Alcaraz-Segura, D.; Paruelo, J.M.; Epstein, H.E.; Cabello, J. Environmental and human controls of ecosystem functional diversity in temperate South America. *Remote Sens.* **2013**, *5*, 127–154.

Appendix

Figure A. Box-plots of the three ecosystem function variables (LST, Albedo and NDVI) for the twelve EFTs.
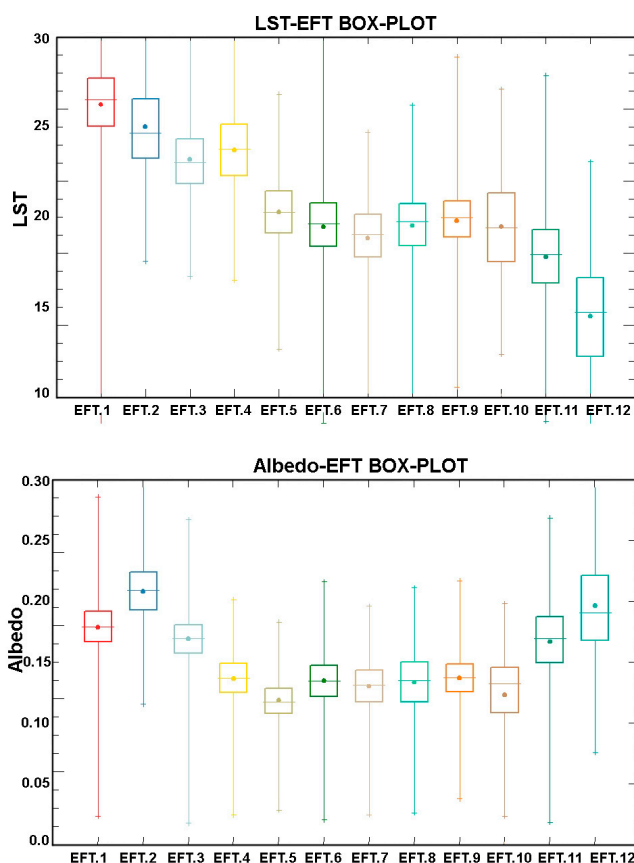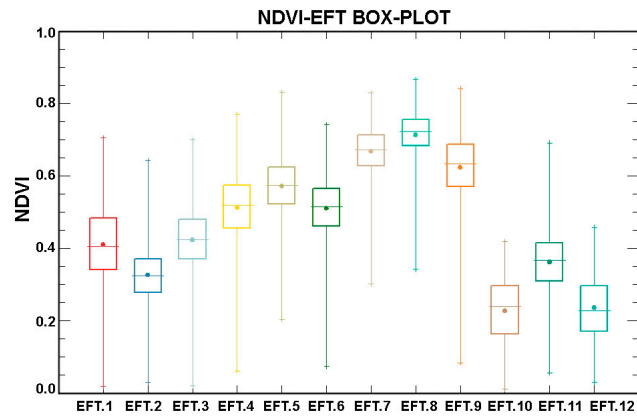
Figure A. *Cont.*

Figure B. Distribution of the 36 EFTs in Spain.
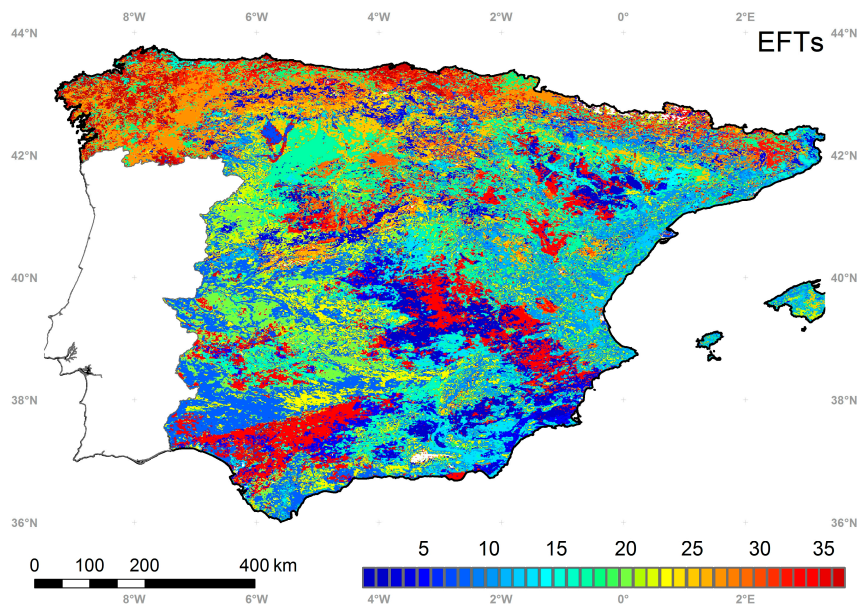
Table A. Wilk's lambda results to identify the partial contribution of each variable in the discrimination of EFTs.

Variable	Wilk's Lambda
Moisture	0.24913
IAlbedo	0.27629
Aridity Index	0.29232
ValminNDVI	0.3122
ValmaxLst	0.33468
INDVI	0.33787
DNDVI	0.34857
ValmeanLST	0.37458
ValminAlbedo	0.397
DAlbedo	0.51617
DLST	0.51617
RelNDVI	0.78532

Table B. Minnick's coefficient (Cm) to show the degree of overlap between the land cover dataset and the EFTs.

	EFT. 1	EFT. 2	EFT. 3	EFT. 4	EFT. 5	EFT. 6	EFT. 7	EFT. 8	EFT. 9	EFT. 10	EFT. 11	EFT. 12
Irrigated	0.03	0.04	0.23	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Cultivated	0.37	0.64	0.09	0.06	0.01	0.03	0.01	0.01	0.01	0.00	0.00	0.00
Mosaic Cultivated	0.04	0.02	0.02	0.04	0.03	0.08	0.07	0.06	0.09	0.00	0.01	0.00
Broadleaved	0.01	0.00	0.01	0.08	0.04	0.10	0.15	0.07	0.12	0.00	0.01	0.00
Needleleaved	0.01	0.00	0.01	0.10	0.28	0.09	0.07	0.06	0.03	0.00	0.01	0.00
Mixed Forest	0.00	0.00	0.00	0.02	0.05	0.04	0.09	0.11	0.10	0.00	0.01	0.00
Shrublands	0.03	0.00	0.02	0.21	0.11	0.05	0.02	0.01	0.03	0.00	0.01	0.00
Herbaceous	0.12	0.03	0.05	0.14	0.01	0.04	0.05	0.03	0.04	0.00	0.01	0.01
Sparse	0.07	0.03	0.04	0.08	0.03	0.04	0.00	0.00	0.01	0.01	0.02	0.02
Bare	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01
Wetlands	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Snow	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table C. Minnick's coefficient (Cm) to show the degree of overlap between the land cover dataset and the ecoregions.

	EFT. 1	EFT. 2	EFT. 3	EFT. 4	EFT. 5	EFT. 6	EFT. 7	EFT. 8	EFT. 9	EFT. 10	EFT. 11	EFT. 12
Cantabrian Mixed Forest	0.00	0.00	0.01	0.00	0.00	0.08	0.36	0.26	0.37	0.00	0.03	0.01
PyreneesConifer	0.01	0.00	0.00	0.00	0.01	0.09	0.09	0.03	0.03	0.00	0.05	0.08
IberianConifer	0.03	0.07	0.06	0.05	0.09	0.08	0.00	0.00	0.00	0.02	0.00	0.00
Iberian Sclerophyllous	0.40	0.39	0.08	0.26	0.06	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Northeastert Mediterranean Forest	0.01	0.01	0.14	0.06	0.18	0.03	0.04	0.03	0.00	0.00	0.00	0.00
Northwest Iberian Montane	0.02	0.08	0.06	0.03	0.05	0.21	0.03	0.03	0.06	0.00	0.01	0.00
Southeastern IberainShrubs	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Southwest Iberian Sclerophyllous	0.12	0.01	0.03	0.11	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00