MDPI

Article

LSTM-Based Prediction of Mediterranean Vegetation Dynamics Using NDVI Time-Series Data

Christos Vasilakos [1],*, George E. Tsekouras [2] and Dimitris Kavroudakis [1]

1   Department of Geography, University of the Aegean, 81100 Mytilene, Greece; dimitrisk@aegean.gr
2   Department of Cultural Technology and Communications, University of the Aegean, 81100 Mytilene, Greece; gtsek@ct.aegean.gr
*   Correspondence: chvas@aegean.gr; Tel.: +30-22510-36451

Abstract: Vegetation index time-series analysis of multitemporal satellite data is widely used to study vegetation dynamics in the present climate change era. This paper proposes a systematic methodology to predict the Normalized Difference Vegetation Index (NDVI) using time-series data extracted from the Moderate Resolution Imaging Spectroradiometer (MODIS). The key idea is to obtain accurate NDVI predictions by combining the merits of two effective computational intelligence techniques; namely, fuzzy clustering and long short-term memory (LSTM) neural networks under the framework of dynamic time warping (DTW) similarity measure. The study area is the Lesvos Island, located in the Aegean Sea, Greece, which is an insular environment in the Mediterranean coastal region. The algorithmic steps and the main contributions of the current work are described as follows. (1) A data reduction mechanism was applied to obtain a set of representative time series. (2) Since DTW is a similarity measure and not a distance, a multidimensional scaling approach was applied to transform the representative time series into points in a low-dimensional space, thus enabling the use of the Euclidean distance. (3) An efficient optimal fuzzy clustering scheme was implemented to obtain the optimal number of clusters that better described the underline distribution of the low-dimensional points. (4) The center of each cluster was mapped into time series, which were the mean of all representative time series that corresponded to the points belonging to that cluster. (5) Finally, the time series obtained in the last step were further processed in terms of LSTM neural networks. In particular, development and evaluation of the LSTM models was carried out considering a one-year period, i.e., 12 monthly time steps. The results indicate that the method identified unique time-series patterns of NDVI among different CORINE land-use/land-cover (LULC) types. The LSTM networks predicted the NDVI with root mean squared error (RMSE) ranging from 0.017 to 0.079. For the validation year of 2020, the difference between forecasted and actual NDVI was less than 0.1 in most of the study area. This study indicates that the synergy of the optimal fuzzy clustering based on DTW similarity of NDVI time-series data and the use of LSTM networks with clustered data can provide useful results for monitoring vegetation dynamics in fragmented Mediterranean ecosystems.

Keywords: remote sensing; NDVI; machine learning; LSTM; spatiotemporal forecasting

1. Introduction

Monitoring vegetation changes in Mediterranean-type ecosystems is crucial in several studies of climate change, hydrology, and ecology [1]. Remote sensing plays a key role in ecological studies regarding environmental changes in an ecosystem context while the usage of vegetation indices has been widely explored for such studies [2]. Forecasting vegetation indices over time is critical for decision-making to reduce losses from environmental hazards such as land degradation and drought, especially in areas at risk of desertification [3,4]. Moreover, drought affects agricultural production, so forecasting vegetation indices provide information on vegetation health and allow the farmer to prepare for water scarcity [5].

Various studies have analyzed the spatiotemporal variations of the Normalized Difference Vegetation Index (NDVI) for the investigation of vegetation changes [6–9]. For example, Bai et al. [10] investigated the vegetation change as depicted from the NDVI, and the roles of climate change, CO_2 fertilization and human activities. They applied their research in a quite complex ecosystem including mountain, oasis and desert. According to their results, the spatial heterogeneity in driving forces of vegetation change emphasizes the need to distinguish between spatial and temporal environmental management in areas with complex ecosystems. The NDVI has been also analyzed in a desert/grassland transition zone [11]. In that research, from 1982–2015, the NDVI showed an increasing trend while most of the vegetation in the transition zone showed a tendency towards recovery.

An extensive literature has been developed on crop prediction based on vegetation indices [12–15]. Usually the vegetation indices derived from remote sensing data are used in conjunction with other independent variables, including climate data, air-temperature, total precipitation, isolation, as well as soil characteristics [16]. Besides yield estimation, forecasted vegetation indices have been used for other applications as well, including natural hazards [17–19], tree mortality [20,21], and lake surface fluctuations [22]. Most of the previous studies have developed empirical models using statistical methods. Recently, more advanced approaches have been explored based on machine learning. In a recent study, Wang et al. [23] used machine learning approaches to develop prediction models for winter wheat farming yield. They compared two linear regression methods and four machine learning methods, including support vector machine (SVM), random forest (RF), adaptive boosting (AdaBoost), and deep neural network (DNN), for estimation of winter wheat yield within the growing season for their study area. Based on their results, the machine learning methods seem to perform better than linear regression models. In [24], Karateke et al. employed an optimized adaptive neural-fuzzy inference system (ANFIS) and hybrid wavelet−ANFIS (WANFIS) model to estimate NDVI variation. The comparison between the forecasted and the actual NDVI values from MODIS obtained a mean absolute percentage error as high as 1.5%.

Early studies as well as the current work focus on deep learning approaches [3]. Non-stationary NDVI time-series data are used for forecasting while one of the well documented approaches is the use of long short-term memory (LSTM) neural networks. Several works have found that LSTM outperforms the traditional regression-based methods, hence it is beyond our scope to prove the superiority of LSTM over traditional regression methods. For example, Cui et al. [25] claim that auto-regressive integrated moving average (ARIMA) assumptions about linearity and stationarity cannot be applied in NDVI time series while the abnormal changes due to various disturbances cannot be accurately predicted. Furthermore, Wang et al. [26] argue that LSTM is suitable for NDVI time-series forecasting due to its ability to store internally information gained from a long period of the past. In other approaches, [27,28] authors combined LSTM and wavelet transform (WT). The WT was used for the time-series decomposition into different components while the LSTM was used for forecasting and the synergy of both methods provided trustworthy results. The LSTM was also applied for NDVI and fraction of photosynthetically active radiation (fPAR) time-series modeling. The problem was formulated as a dynamic sequence-to-sequence modeling task incorporating meteorological variables as dynamic predictors [29]. As predictor variables, some authors used a number of static variables representing soil characteristics and land cover types with climate time series [30]. Their aim was to identify memory effects and gain a better understanding of the scales on which they operate in various environments. According to their findings, memory effects are important on a global scale, while the global NDVI was predicted with an RMSE of 0.056. Finally, other studies used modified LSTM approaches such as bidirectional LSTM [27] and convolutional long short-term memory (ConvLSTM) [31]. A more extensive systematic literature review on deep learning-based approaches for vegetation forecasting can be found in [3].

One of the fundamental problems of time-series data mining is the representation of the data [32]. Hence, time-series data forecasting without any data clustering requires that

one model to be built for each pixel. This requires significant computational effort and validation of a substantial number of models. The low spatial resolution of MODIS NDVI products (i.e., 1000 m by 1000 m) does not represent the homogeneous pixels required to take into account the large variability in phenology signals found in a fragmented Mediterranean landscape. To deal with the above representation problem, different dimensionality reduction mechanisms were proposed for time-series transformation. One of the proposed methods is the similarity measure between time series [32]. Based on this measure, the time-series data can be clustered to a higher-level abstraction [33]. One of the most commonly used similarity measures for time-series data is the dynamic time warping (DTW) [34–37] applied also in remote sensing time-series data [38–41].

The goals of this paper were threefold: (1) to cluster the NDVI time-series data based on the DTW similarity algorithm into an optimal number of clusters; (2) to correlate the clusters with the land-use/land-cover types of the study area; and (3) to develop and evaluate an LSTM model for each identified cluster and apply it for one year; i.e., 12 monthly time steps. The analysis was applied to an insular environment of the Mediterranean coastal region and to the best of our knowledge, the combination of DTW and LSTM for NDVI forecasting has not been examined for this type of mixed and fragmented ecosystem.

2. Materials and Methods

2.1. Study Area and Data

Lesvos island is in the northeast part of the Aegean Sea (Greece), with a total area of 1636 km^2 and seashore of about 382 km. It includes a variety of geological formations, a number of climatic conditions and substantial number of vegetation types (Figure 1). Climate conditions of the island could be categorized as Mediterranean, including warm and dry summer periods along with a mild and moderately rainy winter season. Annual average precipitation is about 710 mm while annual average air temperature is about 17 °C with high daily fluctuations. Terrain is somewhat hilly and rough (highest peak of 960 m a.s.l). Dominant slopes have values of greater than 20%, covering about two-thirds of the island.

Since 1950, agricultural areas have been abandoned due to low productivity, resulting in shrub regeneration and wildfires [42]. Soil types on the island are widely cultivated, mostly producing rain-fed crops such as cereals, vines and olives. Other vegetation types include phrygana or garrigue-type shrubs in grasslands, evergreen-sclerophylous or maquis-type shrubs, pine forests, deciduous oaks, olive groves and other agricultural lands.

We obtained the NDVI (MOD13A3 product—Version 6 of MODIS Terra sensor) by submitting the extent of the study areas through the Application for Extracting and Exploring Analysis Ready Samples (AppEEARS) platform [43]. This is a monthly maximum value composite (MVC) NDVI product with a spatial resolution of 1000 m. We retrieved the complete database, i.e., from 2000 till end of 2020. More specifically, 239 images for the period 2000−2019 were used for training of the models while 12 images from 2020 were used for the spatial validation of the resultant models. The final database included 2704 time series of 239 time steps as training data and 12 time steps as validation data. During quality control, a certain number of null values within each time series were replaced with interpolated values.
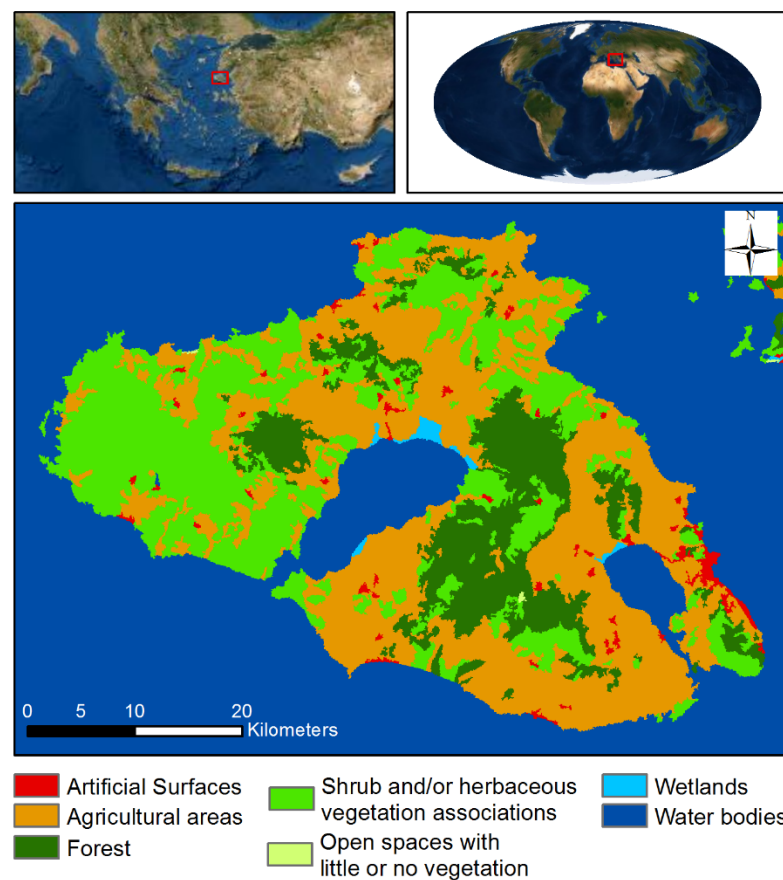
Figure 1. The land cover of Lesvos Island in the northeast Aegean sea according to the corine land cover (CLC) database 2018, version 2020_20u1 (https://land.copernicus.eu/pan-european/corine-land-cover/clc2018 accessed on 10 June 2021).

2.2. Methodology

2.2.1. Optimal Fuzzy Clustering of the DTW Distances

The NDVI is a nonlinear, nonstationary and seasonal time series, thus the analysis should be performed for each pixel independently. Based on a nonlinear autoregressive model (NAR), the future value $y(t)$ is forecasted using d past values according to Equation (1).

$$y(t) = f(y_{t-1}, y_{t-2}, \ldots, y_{t-d}) \tag{1}$$

where f is a nonlinear known function that will be estimated through the proposed method. To avoid building a separate model for each pixel, all time-series data were clustered into an optimal number of clusters based on a fuzzy clustering method. The clustering method was based on the dynamic time warping (DTW) similarity algorithm [44]. This technique can be used to determine the best alignment between two time series by dealing with temporal deformations and speeds in time-varying data (Figure 2). In addition, DTW accounts for the misalignment of peaks and dips between two sequences. A detailed description of the calculation of DTW can be found in [41].

The clustering procedure consisted of three steps and used tools already existing in the literature (Figure 3). The first step concerned data reduction, where a set of representative time series was selected. In the second step, a multidimensional scaling procedure mapped the representative times series to points in the $\Re^p$ Euclidean space. Finally, the third step applied an optimal fuzzy clustering to the points obtained in the previous step. These steps are analyzed in the next paragraphs.
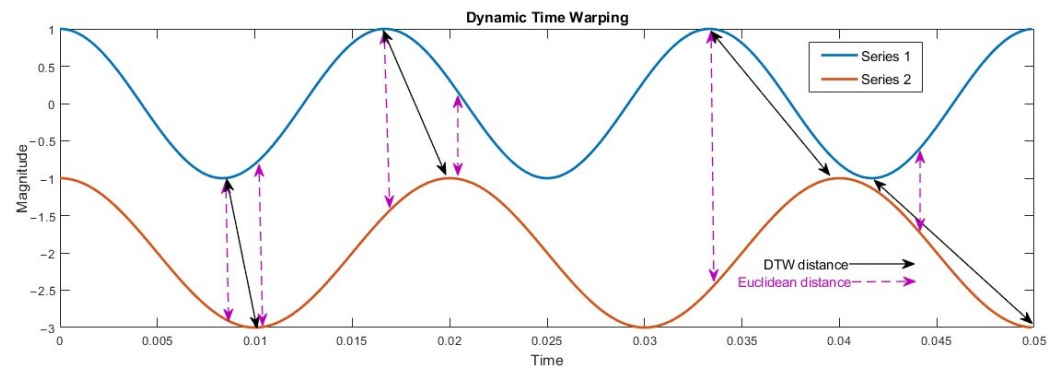
Figure 2. Dynamic time warping (DTW) (solid arrowed lines) versus Euclidean distance (dotted arrowed lines) for two arbitrary time series. Reproduced from [41] (CC BY 4.0).
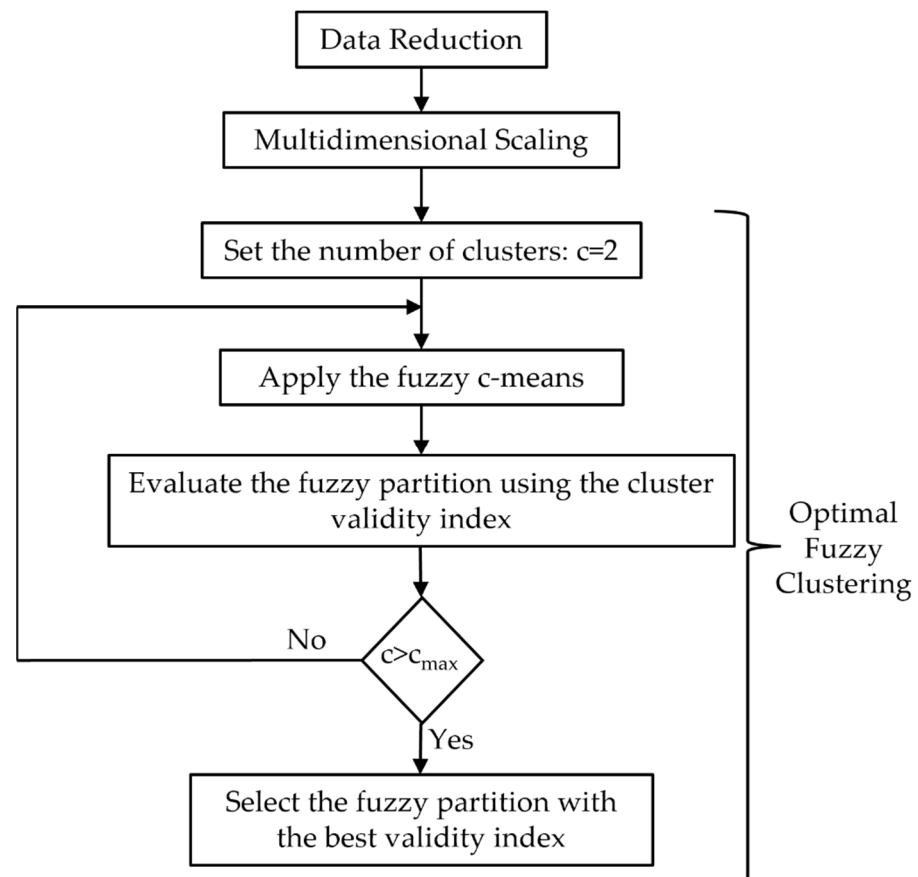
Figure 3. The basic steps of the optimal clustering algorithm: (i) the data reduction process generates several representative time series, where each time series from the original set is assigned to a representative times series; (ii) the multidimensional scaling transforms the representative time series into points within a low-dimensional feature space; (iii) the fuzzy c-means algorithm is applied several times with a different number of clusters; (iv) for each number of clusters the corresponding validity index is calculated; (v) the fuzzy partition with the minimum validity index is finally selected.

Let us assume that we are given N times series $T = \{T_1, T_2, \ldots, T_N\}$ and their pairwise distances $DTW_{ij}(1 \leq i \leq N; 1 \leq j \leq N)$. The similarity degree between the i-th and j-th time series is defined as:

$$Sim_{ij} = exp(-\alpha DTW_{ij}) \tag{2}$$

where $\alpha \in (0,1)$. The i-th and j-th time series are considered similar if $\exp(-\alpha DTW_{ij}) \geq \beta$, with $\beta \in (0,1)$. The target of the reduction process is to obtain a set of groups, where each group includes a number of similar time series. Moreover, each group is described by a representative times series. Based on Equation (2), the potential of the i-th time series is calculated in terms of the subsequent formula [45,46]:

$$P_i = \sum_{j=1}^{N} Sim_{ij} = \sum_{j=1}^{N} \exp(-\alpha DTW_{ij})(1 \leq i \leq N) \tag{3}$$

A high value of P_i directly implies that many time series are close to T_i. Therefore, a time series with a high potential value is a good nominee to be a representative time series. Thus, the key idea of the data reduction process is to determine a specific time series and use it to represent the corresponding group of time series. The number of the representative time series (and therefore the number of groups) is denoted as n with $n << N$. Consequently, instead of taking account of the N time series, we finally consider the n representative time series, obtaining a smaller (i.e., reduced) amount of data. The algorithm to generate the representative time series and the corresponding groups is given below:

The values for the parameters used in this paper are: $\alpha = -log(0.5)/DTW_{mean}$, where DTW_{mean} is the mean DTW distance, and $\beta = 0.75$. Algorithm 1 selects n representative times series $\{g_1, g_2, \ldots, g_n\}$ from the set T, with $n << N$. To each $g_k (1 \leq k \leq n)$ there are assigned several time series from the set T, according to the Step 5, forming the k-th time series group.

Algorithm 1: Data Reduction Process [45,46]

1: Select values for the parameters $\alpha, \beta \in (0,1)$ and set $n = 0$.
2: Using Equation (2) calculate the values of the potential of all the N time series.
3: Set $n = n + 1$.
4: Calculate $\underset{T_i \in T}{Pmax}\{P_i\}_{max}$. Select the time series that corresponds to the P_{max} as the n-th

representative time series: $g_n = \left\{ T_{i_0} \in T : P_{i_0} = \underset{T_i \in T}{Pmax}\{P_i\}_{max} \right\}$.

5: Remove from the set T all times series for which $\exp(-\alpha DTW_{nj}) \geq \beta$, and assign them to the n-th group, the representative element of which is the g_n time series.
6: If T is empty stop; else turn the algorithm to Step 2.

In the next step, multidimensional scaling (MS) [47,48] is used to transform the time series $\{g_1, g_2, \ldots, g_n\}$ into the points $\{x_1, x_2, \ldots, x_n\}$ lying in a lower p-dimensional Euclidean space $\Re^p$. Thus, $x_k = [x_{k1} x_{k2} \ldots x_{kp}]$. The DTW_{ks} $(1 \leq k \leq n; 1 \leq s \leq n)$ between the time series g_k and g_s is transformed into the Euclidean distance $\|x_k - x_s\|$ between two points x_k and x_s. An effective way to do this is to minimize the differences $|\|x_k - x_s\| - DTW_{ks}|$, using the following objective function [41]:

$$J_{MS} = \frac{1}{\sum_{s<k} DTW_{ks}} \sum_{s<k} \frac{(\|x_k - x_s\| - DTW_{ks})^2}{DTW_{ks}} \tag{4}$$

The above function is minimized using the standard gradient-descent optimization approach. The interested reader is referred to [41,47,48] for further details. To satisfy visualization, a typical value for the parameter p, also used in this paper, is $p = 2$.

In the final step, as shown in Figure 3, we use the well-known fuzzy c-means [49,50] to perform optimal fuzzy clustering. The fuzzy c-means obtains a partition of the set points $\{x_1, x_2, \ldots, x_n\}$ into a number of c clusters $C_1, C_2, \ldots, C_c$ with cluster centers $\{v_1, v_2, \ldots, v_c\}$, with $v_i \in \Re^p (1 \leq i \leq c)$, where the membership degree of the data vector x_k to the i-th cluster is $u_i(x_k)$. The optimal fuzzy clustering concerns the determination of the optimal

number of clusters which derives a fuzzy partition of the set $\{x_1, x_2, \ldots, x_n\}$, such that the resulting clusters are well separated and compact [49–51]. Specifically, the FcM runs for $c = 2, 3, \ldots, c_{max}$, where c_{max} is the maximum number of tested fuzzy clusters. Each time, the value of a validity index is calculated. The final partition corresponds to the minimum of the obtained values of the validity index.

The validity index used here was developed in [51], and reads as:

$$VI = \frac{\sum_{i=1}^{c} \left(\frac{\sum_{k=1}^{n} (u_i(x_k))^m \|x_k - v_i\|^2}{\sum_{k=1}^{n} u_i(x_k)} \right)}{\sum_{i=1}^{c+1} \sum_{\substack{j=1 \\ |j \neq i}}^{c+1} \left(\left(\left[\sum_{\substack{\ell=1 \\ |\ell \neq j}}^{c+1} \left(\frac{\|z_j - z_i\|}{\|z_j - z_\ell\|} \right)^2 \right]^{-1} \right)^2 \|z_j - z_i\|^2 \right)} \tag{5}$$

with $[z_1 z_2 \ldots z_c z_{c+1}]^T = [v_1 v_2 \ldots v_c x]^T$, and $x = \frac{\sum_{k=1}^{n} x_k}{n}$ is the mean vector of the set $\{x_1, x_2, \ldots, x_n\}$. For a detailed description of the index, the interested reader is referred to the referenced paper [51].

The implementation of the above validity index obtains an optimal fuzzy partition, where the optimal number of clusters is denoted as c_{opt}. Thus, the points $\{x_1, x_2, \ldots, x_n\}$ are partitioned into the set of clusters $\left\{ C_1, C_2, \ldots, C_{c_{opt}} \right\}$.

Since there is a one-to-one correspondence between the points $\{x_1, x_2, \ldots, x_n\}$ and the time series $\{g_1, g_2, \ldots, g_n\}$, the time series also are partitioned into the same number of clusters $\left\{ R_1, R_2, \ldots, R_{c_{opt}} \right\}$. The representatives (i.e., cluster centers) of the above clusters are denoted as $\left\{ r_1, r_2, \ldots, r_{c_{opt}} \right\}$. Finally, each $r_i (1 \leq i \leq c_{opt})$ is a time series calculated as the mean of all elements of the set $\{g_1, g_2, \ldots, g_n\}$ that belong to the cluster R_i.

2.2.2. LSTM Training and Validation

The optimal cluster centers described above were used for the time-series processing. According to the literature, the architecture of recurrent neural networks (RNNs) makes them suitable for time-series data processing including classification and prediction tasks [52]. The training process of RNNs is usually based on gradient-based training algorithms, i.e., the back propagation algorithm, hence they suffer from local minima and the vanishing gradient problem [53,54]. LSTM networks are a special type of RNN that can learn long-term dependencies and avoid the vanishing gradient problem of RNN [55]. The typical structure of an LSTM network includes the sequence input layer, a predefined number of LSTM layers, a number of fully connected layers and the regression output layer (Figure 4). With this structure, at the time step t the cell is fed with the input X_t and the hidden state H_{t-1}. Then, the output state C_t and the hidden state H_t at time step t are given by:

$$C_t = f_t \odot C_{t-1} + i_t \odot g_t \tag{6}$$

and

$$H_t = o_t \odot \sigma c(C_t) \tag{7}$$

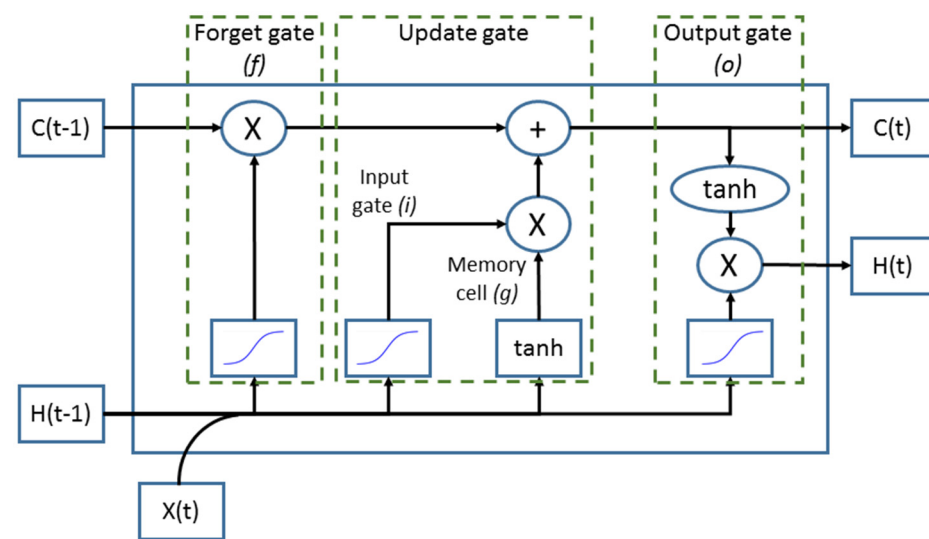with the input gate i_t, the forget gate f_t, the memory cell g_t and the output gate o_t respectively:

$$i_t = \sigma g(W_i X_t + R_i H_{t-1} + b_i) \tag{8}$$

$$f_t = \sigma g\left(W_f X_t + R_f H_{t-1} + b_f\right) \tag{9}$$

$$g_t = \sigma c(W_g X_t + R_g H_{t-1} + b_g) \tag{10}$$

$$o_t = \sigma g(W_o X_t + R_o H_{t-1} + b_o) \tag{11}$$

where W, R, b are the learnable parameters, σg the logistic function and σc the hyperbolic tangent function.

Figure 4. The typical structure of a long short-term memory (LSTM) cell including its fundamental elements, i.e., the forget gate, the input gate, the output gate and the cell state.

One of the drawbacks of maximum value composite vegetation indices from MODIS that may affect the analysis is cloud contamination. Based on previous studies [56], the cluster centers were smoothed by applying the Savitzky–Golay filter [57,58], fitting a third-degree polynomial based on the 10 neighboring data points. Then, input time-series data of the 239 time steps were split into two consecutive subsets (training and test), containing 94% and 6% of the original dataset, respectively. Both datasets were used during the training procedure. Various numbers of networks were tested including a variable number of LSTM cells based on the "Adam" optimizer [59] and the best network according to the root mean square error was saved for further validation. During the validation phase, the saved networks were used for the prediction of the next 12 forecasted steps (year 2020) for each pixel according to their cluster assignment. The produced forecasted NDVI rasters were compared with the NDVI rasters retrieved from USGS for the year 2020. The overall workflow is summarized in Figure 5.
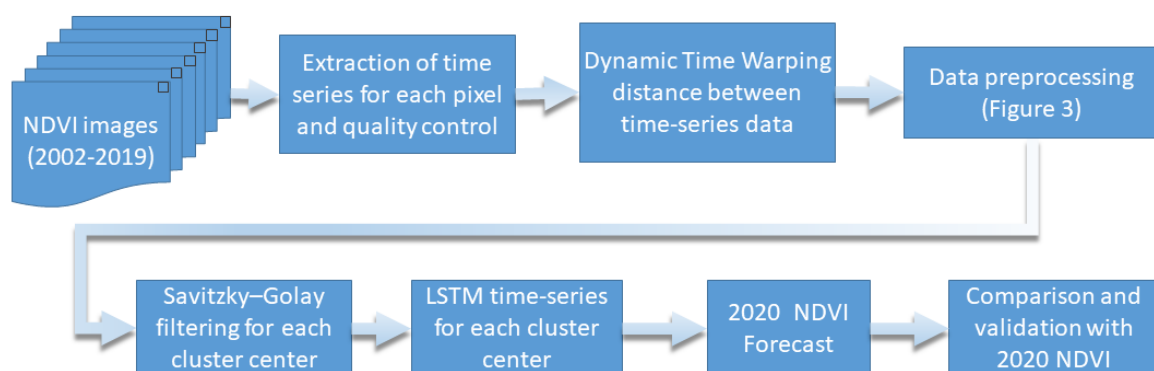

Figure 5. Methodology of the normalized difference vegetation index (NDVI) 12 –month forecast based on long short-term memory (LSTM) neural networks.

3. Results and Discussion

3.1. Optimal Data Clustering and CORINE Cross Correlation

Applying Algorithm 1 using the abovementioned parameter values, we obtained $n = 245$ representative time series $\{g_1, g_2, \ldots, g_{245}\}$. Based on their DTW distances, the application of the multidimensional scaling results in the $n = 245$ points lying on the two-dimensional Euclidean space, which are depicted in Figure 6 (see blue dots). To implement

the optimal fuzzy clustering, we selected $c = \left\lfloor \sqrt{245} \right\rfloor_{max}$, where $\lfloor \Box \rfloor$ stands for the floor function (i.e., integer part). The result of the optimal fuzzy clustering process is reported in Figure 7. In that figure it is clear that the optimal number of clusters is equal to $c_{opt} = 9$. The resulting cluster centers are illustrated in Figure 6 as red rectangles. Thus, the final partition consists of nine clusters $\{R_1, \ldots, R_9\}$ with cluster centers $\{r_1, \ldots, r_9\}$.
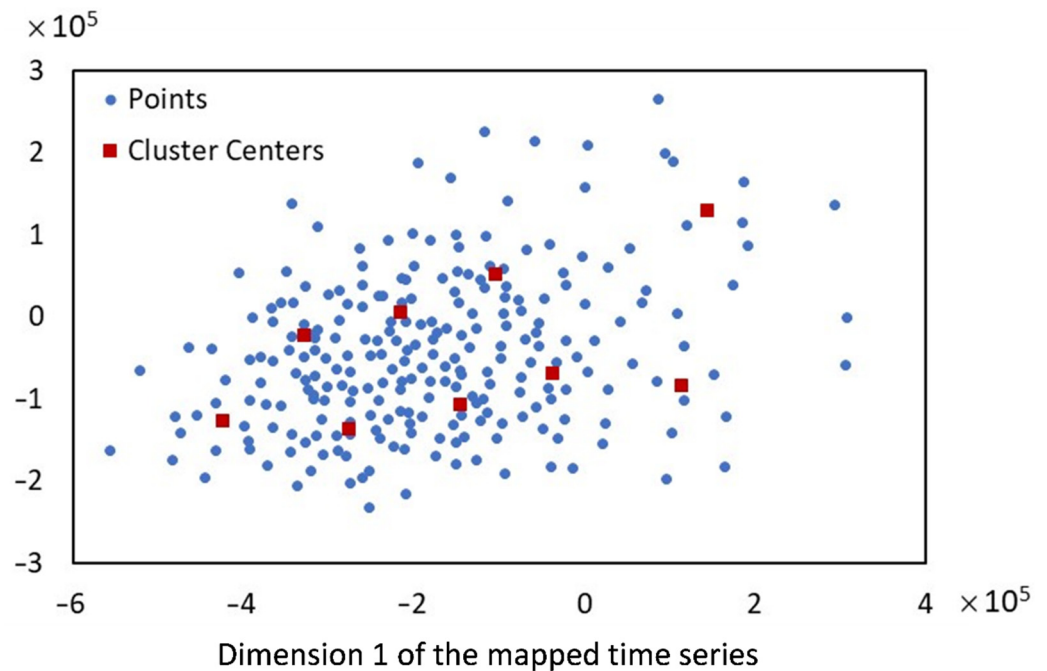
Figure 6. Results of the multidimensional scaling and the optimal fuzzy clustering. Blue dots represent the two-dimensional points to which the 245 time series are mapped through the multidimensional scaling process. Red rectangles represent the position of the nine optimal cluster center positions obtained by the optimal fuzzy clustering. Axes are dimensionless.
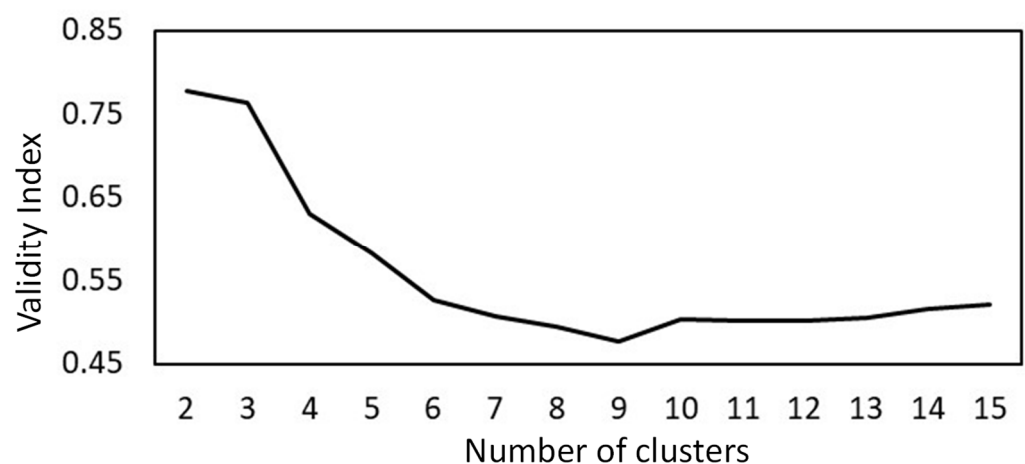

Figure 7. The validity index as a function of the number of clusters. The optimal number of clusters corresponds to the minimum value of the validity index, which is equal to 9.

Figure 8 shows the spatial distribution of the clusters. The comparison with the actual CORINE land-use/land-cover (LULC) data, presented in Figure 1, reveals that there is some evidence about the correlation between clusters and land cover types. To explore the coexistence of cluster classes and LULC, we compared the frequencies of coexistence in the data. Overall, we have 22 LULC types distributed in 9 clusters. Table 1 presents the

number and the frequency of different LULC types belong to each cluster. Clusters 4, 8, 3, 7, 1, 2 and 9 seem to have more than 10 distinct LULC types. This suggests that most of the clusters represent a quite diverse landscape.
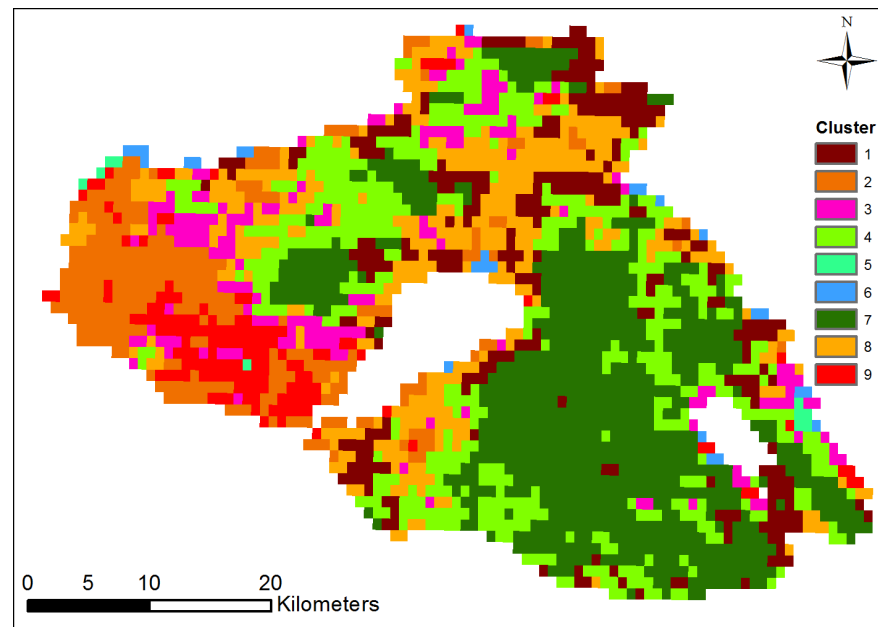

Figure 8. Spatial distribution of optimal nine clusters. Pixels of each cluster show similar NDVI time series based on the dynamic time warping (DTW) similarity algorithm.

Table 1. Length of distinct LULC for each class.

Cluster	4	8	3	7	1	2	9	6	5
No. of LULC types	16	15	14	13	12	11	11	7	3
% of LULC types	69.6	65.5	60.9	56.5	52.2	47.8	47.8	30.4	13

However, even better results can be achieved when applying a cross tabulation between clusters and CORINE LULC types (Figure 9). It is obvious that LULC "321" (natural grassland) is by far the LULC type which is overrepresented in all clusters. Secondly, LULC types "112" (discontinuous urban fabric), "211" (non-irrigated arable land), "223" (olive groves) and "242" (complex cultivation patterns) are very well represented in the majority of clusters. As can be seen in Figure 9, agricultural areas (LULC: 211, 223, 242 and 243) seem to appear in clusters 1, 2, 3, 4 and 7, 8, 9. This grouping is somehow clearer than others in the same figure. Another finding is that clusters 5 and 6 seems to have a somehow limited number of observations compared to all other clusters in the study area.

Furthermore, we evaluated all possible cross-tabulations for each cluster separately. We wanted to explore what was the most dominant LULC type in each cluster. The LULC type "223" is dominant in 4 clusters, followed by LULC type "321", which is dominant in 3 clusters. It is more than clear that the initial dataset of the study area includes a greater percentage of olive groves ("223") and a mix of natural grassland ("321"). This is why the above two LULC types are present as well as dominant in a relatively higher number of clusters.
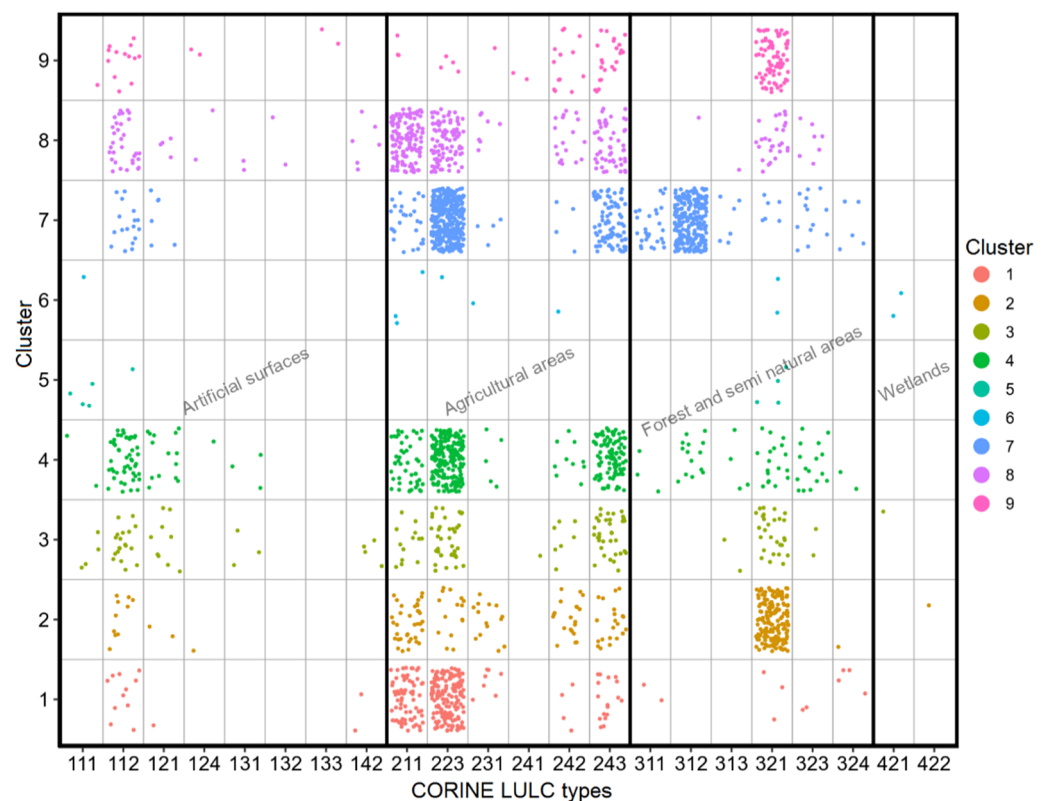
Figure 9. Cross tabulation of clusters and CORINE land-use/land-cover types.

3.2. LSTM Training and Validation

Figure 10 shows the results of the NDVI time-series training for the nine clusters. The testing datasets were the last 14 time steps/months of the training series. The RMSE for the testing datasets were fairly low for all clusters, meaning the networks were successfully trained with an RMS error ranging from 0.017 to 0.079. From the visual interpretation it is clear that there is an agreement between the observed and predicted series. More specifically, for all networks the difference between the predicted and the forecasted NDVI is less than 0.1 for the first 12 time steps. The difference is greater for clusters 4, 5 and 6 during the last two time steps.

Within the validation phase, the trained network of each cluster was applied for each pixel based on its assigned cluster in order to estimate the next 12 time steps, i.e., the year 2020. The comparison of forecasted and actual values of the NDVI for the next 12 months may provide useful insight on the underlying spatial distribution of the NDVI values as well as the quality of the LSTM analysis. Figure 11 depicts the monthly differences between estimated and observed NDVI values based on all pixels regardless of the LULC types.

The average estimated NDVI values for the study area were greater in the first 9 months of the year while this was reversed for the last 3 months of the year. Furthermore, we evaluated the relationship between the two values (observed versus predicted) on a monthly basis to gain a better understanding of the deviation of predicted values (Appendix A). A scatterplot and a histogram were produced for each month depicting the relationship between predicted (Y axis) and observed (X axis) NDVI values. According to these figures, February (Figure A2) seems to be the month with the greater deviation between the two values while September (Figure A9) seems to be the month with the smaller deviation. These results tie in well with previous studies where authors predicted the vegetation dynamics by using LSTM and MODIS (MOD13Q1) NDVI data but with a higher temporal resolution than in our approach [1].

Finally, twelve maps were produced for the predicted NDVI for each month of 2020 (Figure 12). The spatial comparison between the actual and the predicted maps revealed that there is a similarity between the two spatial distributions. Furthermore, the time variation in relation to space shows an agreement between observed and predicted values. It should be noted that the absolute differences were less than 0.1 for most of the part of the study area for the entire study year (Figure 13). The highest differences were observed at the western part where natural grasslands exist and especially during the winter−spring period. One possible cause might be the different precipitation and temperature between the training period and the validation year 2020 given the interannual variation of the NDVI, precipitation and temperature [60–63]. Additionally, some great differences are observed in some of the coastal pixels, probably due to water−land mixture.
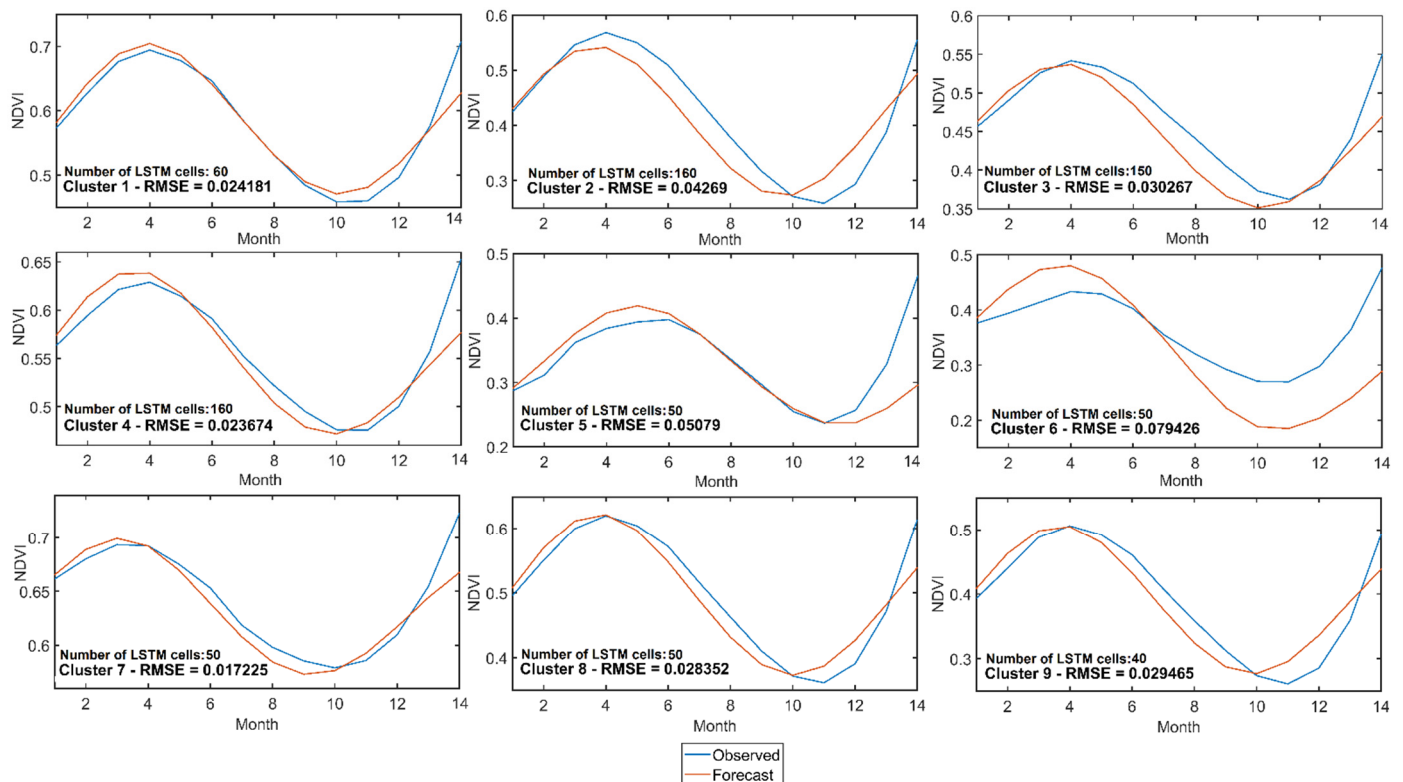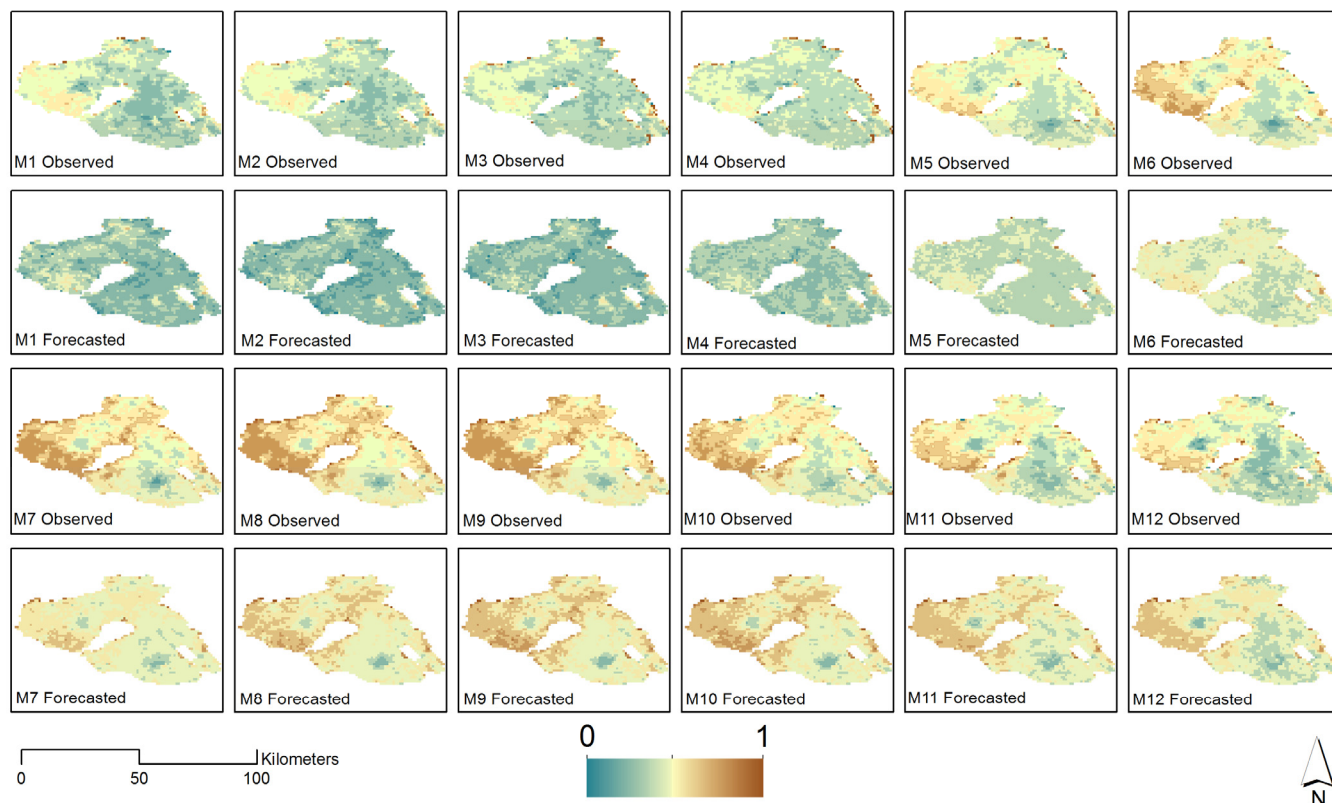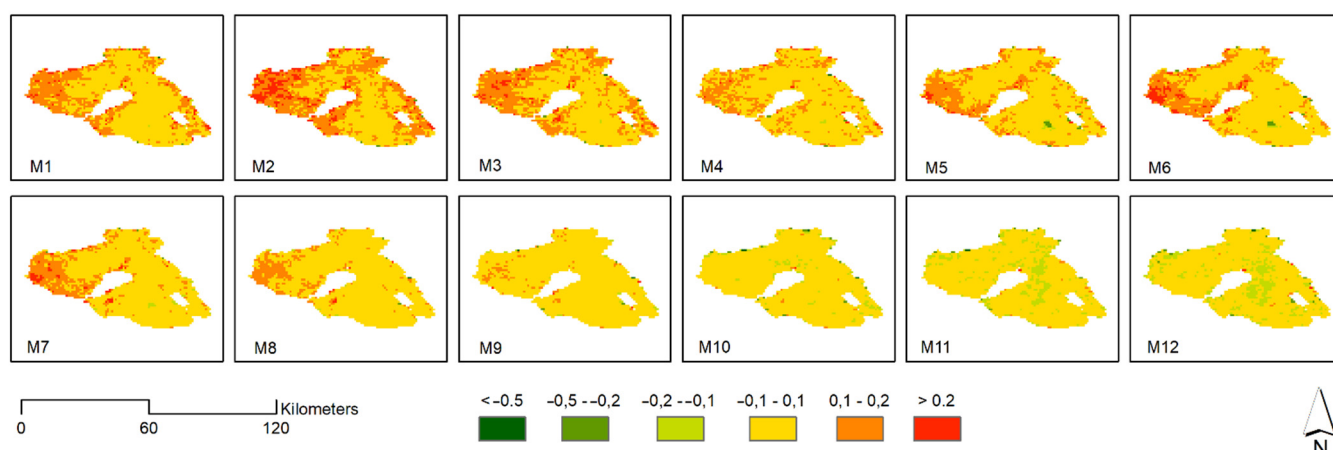

Figure 10. Observed versus forecasted NDVI of the testing dataset.
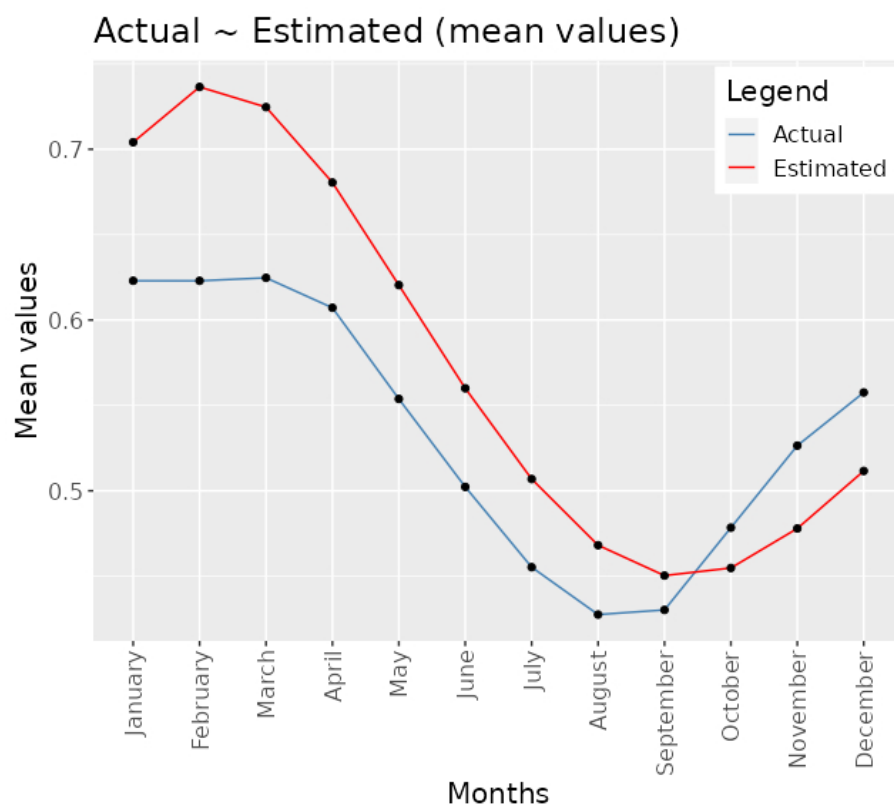
Figure 11. Monthly differences between estimated and predicted NDVI values. For each month, the mean value for the study area was calculated.

Figure 12. Spatial comparison of observed versus forecasted NDVI of the validation dataset. Symbology of maps is stretched to 0–1.

Figure 13. Spatial distribution of the difference between observed and forecasted NDVI of the validation dataset.

4. Conclusions

Understanding vegetation dynamics and taking necessary protective measures is crucial for Mediterranean ecosystems, given the number of major disturbances such as wildfires and droughts, as a result of man-made activities in the areas affected, limiting the ecosystem's regeneration. In the present study, we forecasted the Normalized Difference Vegetation Index (NDVI) based on NDVI time-series data derived from the Moderate Resolution Imaging Spectroradiometer (MODIS). The proposed approach combines the DTW technique, optimal fuzzy clustering and long short-term memory (LSTM) neural networks. The results indicate that the above synergy can provide useful insights for monitoring vegetation dynamics in fragmented Mediterranean ecosystems.

Machine learning methods and semi-empirical models have been proven quite successful in identifying climatic influences both in terms of spatial and temporal grounds. Evaluation of their results as well as recalibration of their properties in terms of sensitivity, are two of the most notable aspects for further research. Recalibration of machine learning model properties, based on spatial characteristics and spatial representation of classes, is also a topic for further research as it has a substantial impact on results' quality. On these grounds, "representation" of all classes as well as measures for handling unbalanced or skewed samples would definitely contribute towards more robust results of machine learning approaches in general. To this end, in the recent literature there are some approaches for auto algorithm selection and hyperparameter optimization based on data variability and composition [64–67]. Nevertheless, recent computational approaches have been proven to be quite promising in estimating environmental variables and incorporating spatiotemporal values in prediction and clustering.

Future research steps could take place regarding the subsequent issues. First, the method could be applied to a larger amount of the data retained for testing the predictions. Second, since fuzzy systems have been proven to be very effective tools in handling uncertainties in data, improved approximation capabilities could be obtained by considering sophisticated polynomial neural-fuzzy networks. Moreover, investigating prediction of vegetation disturbances might prove important for mapping forest degradation due to logging, wildfires, livestock grazing and other human activities [68].

Author Contributions: Conceptualization, C.V.; methodology, C.V. and G.E.T.; validation, C.V. and D.K.; formal analysis, C.V., G.E.T. and D.K.; writing—review and editing, C.V., G.E.T. and D.K. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A
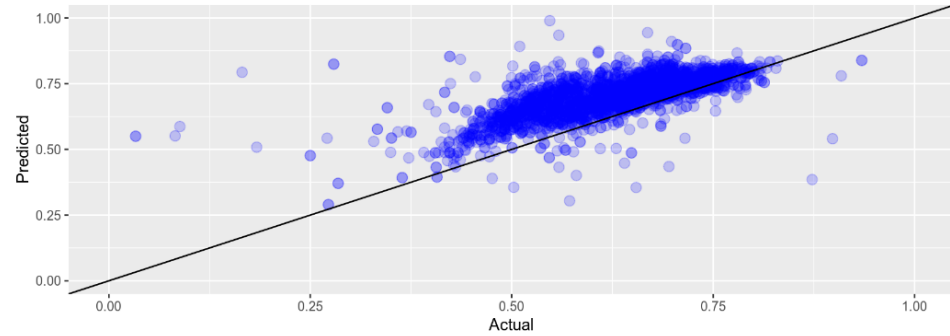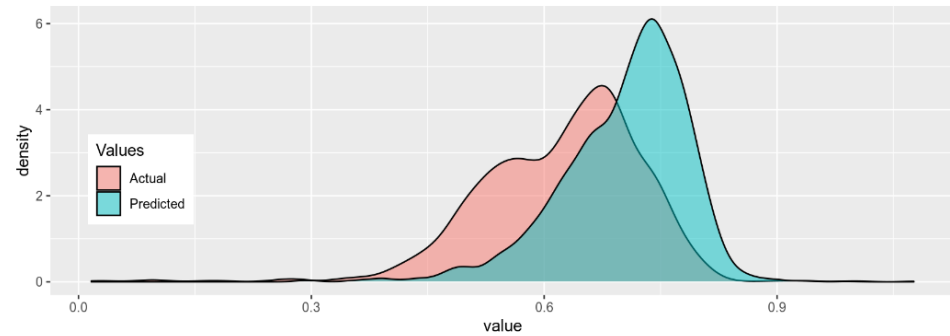
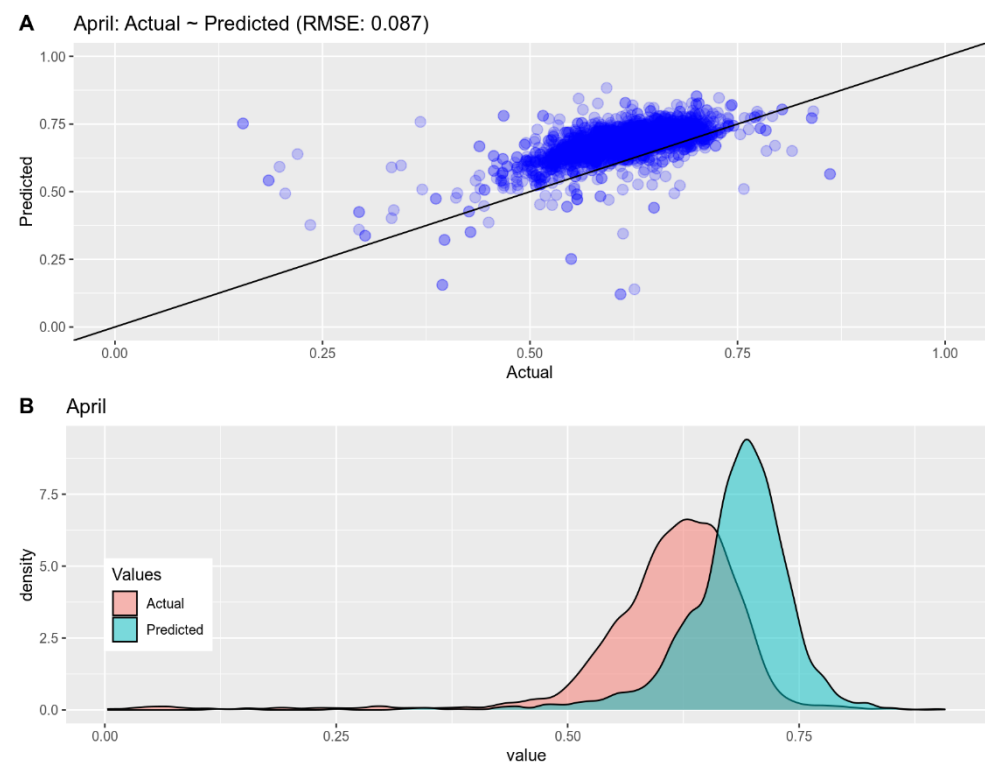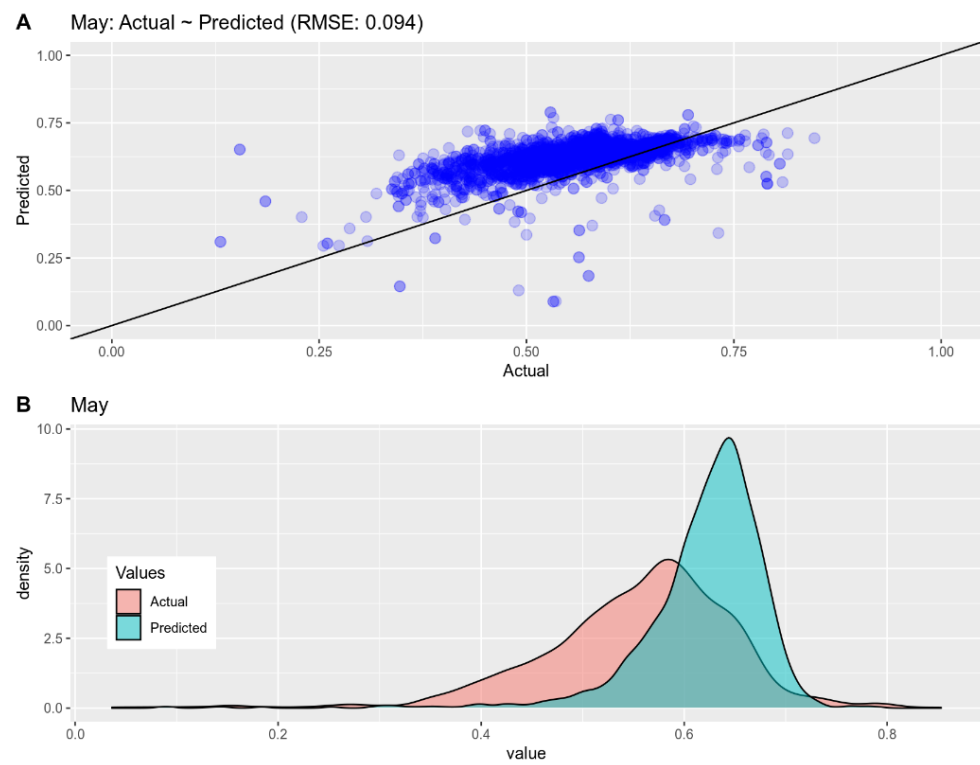**Figure A1.** Comparison between actual~predicted for January 2020.
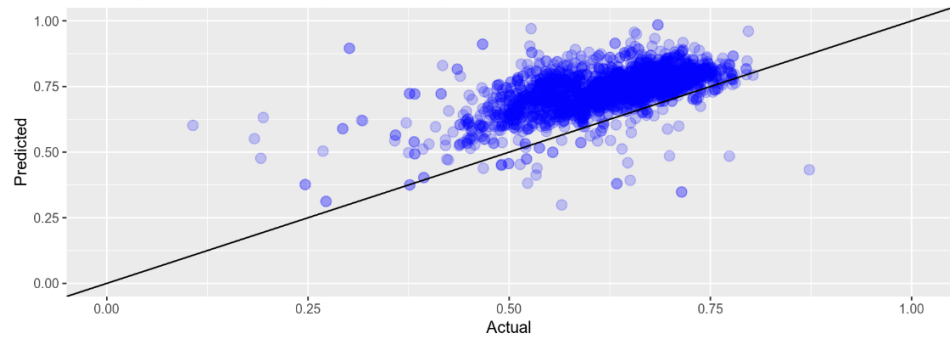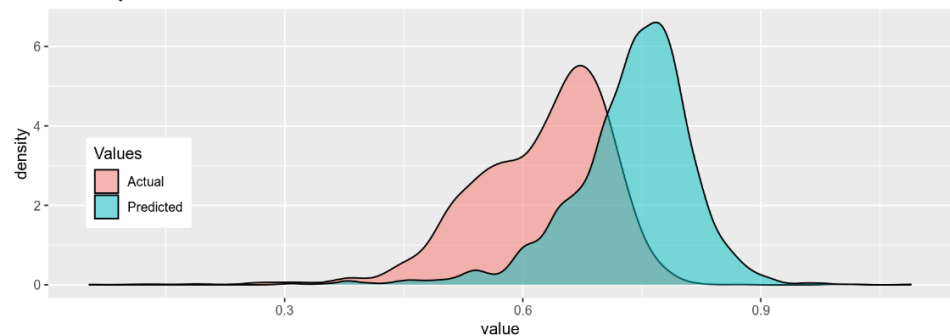
Figure A2. Comparison between actual~predicted for February 2020.
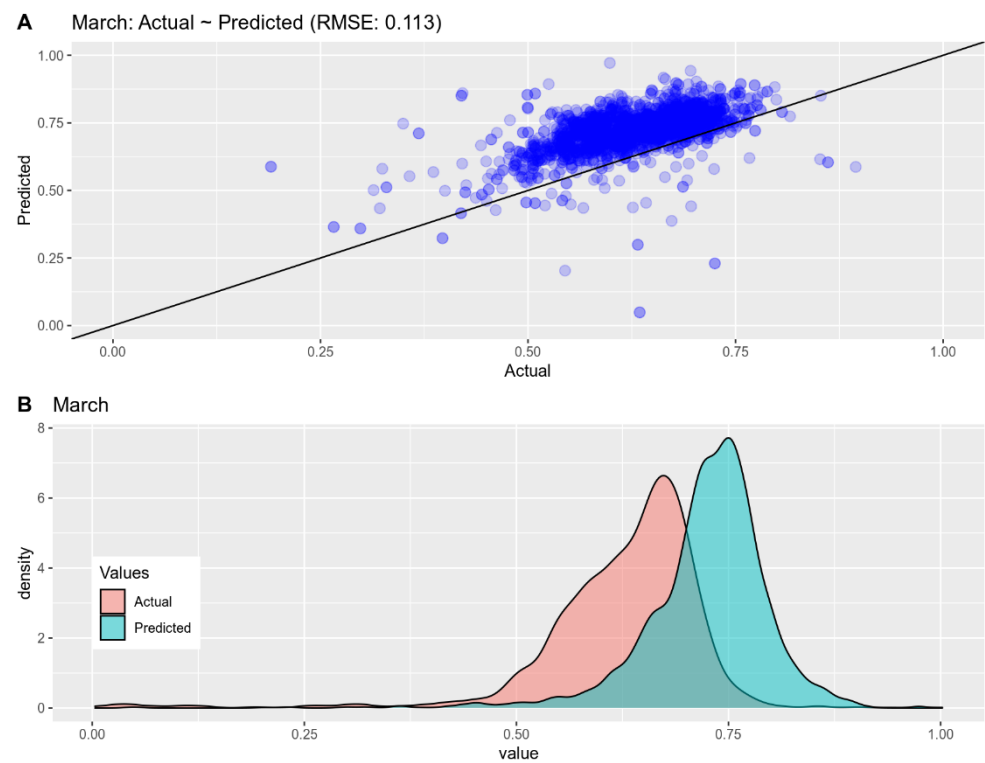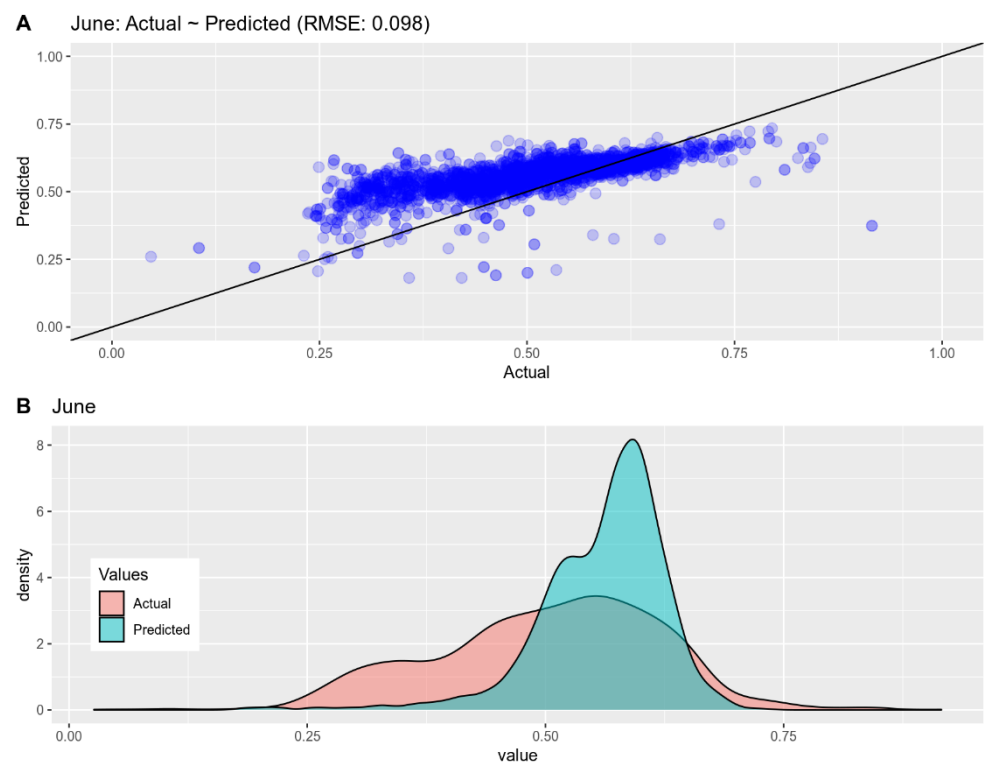
Figure A3. Comparison between actual~predicted for March 2020.

Figure A4. Comparison between actual~predicted for April 2020.

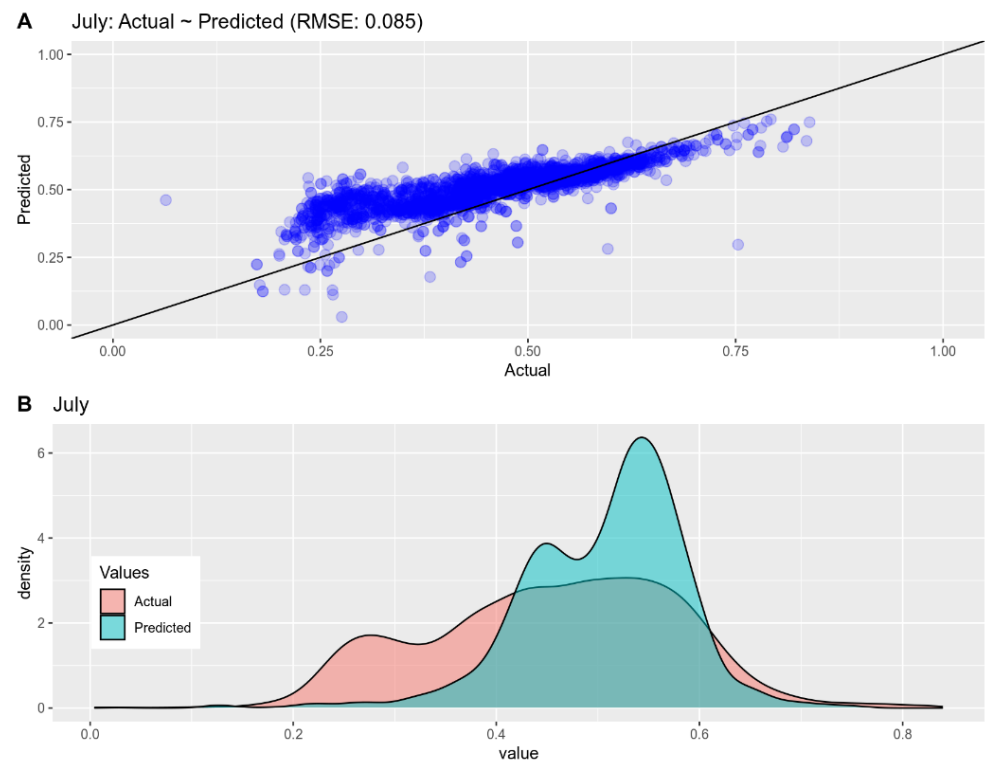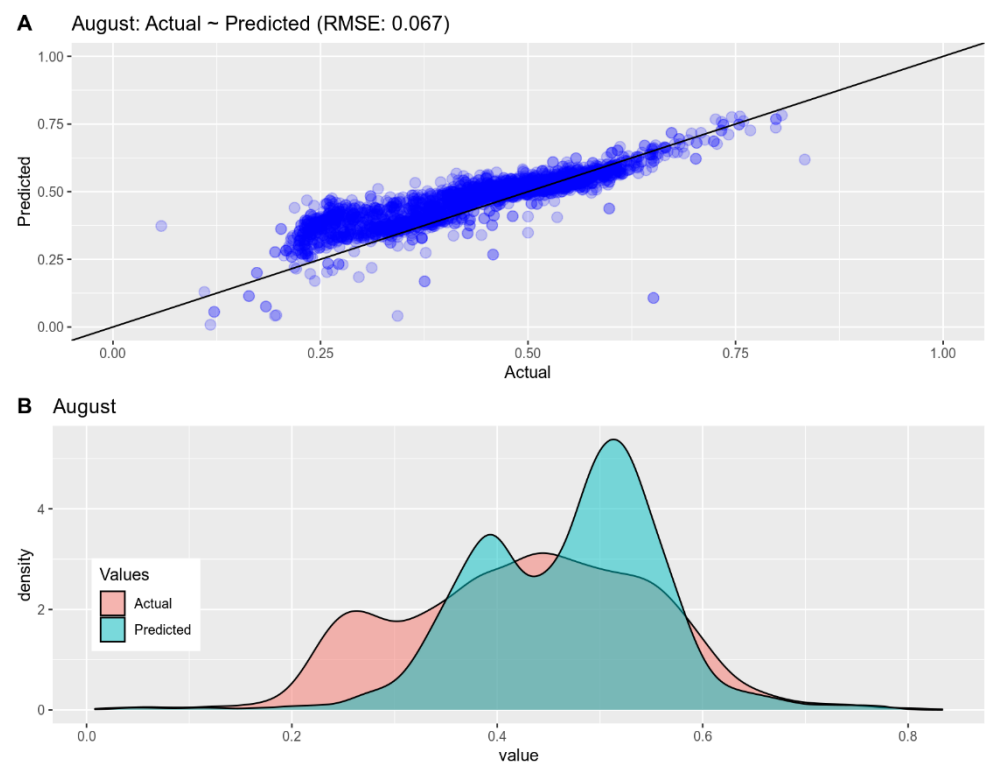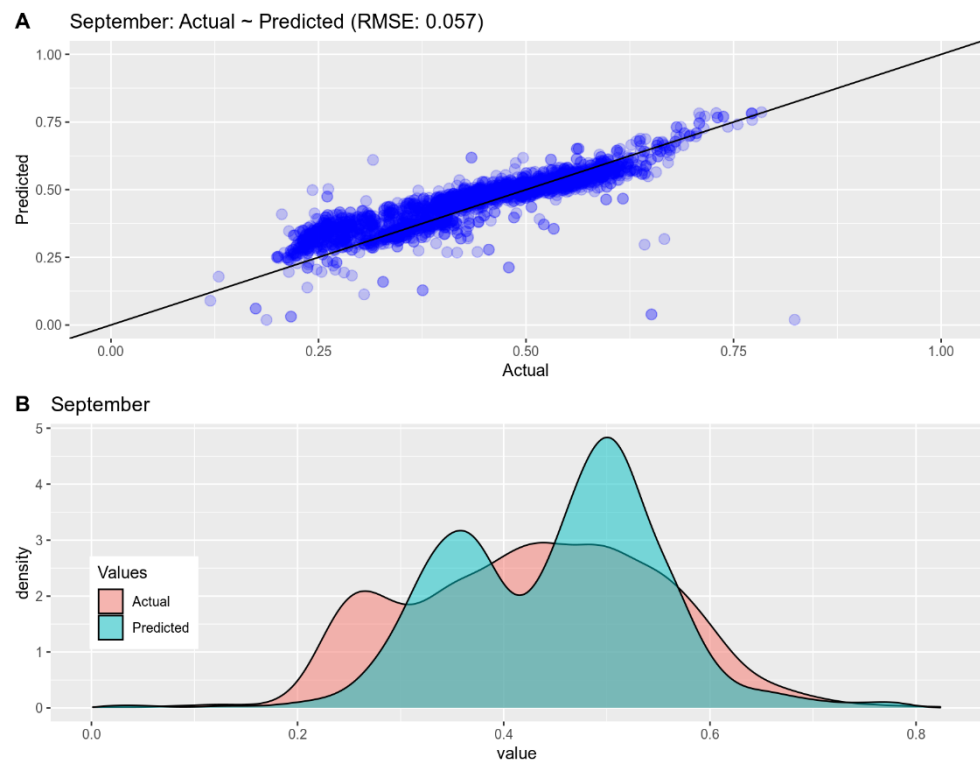**Figure A5.** Comparison between actual~predicted for May 2020.

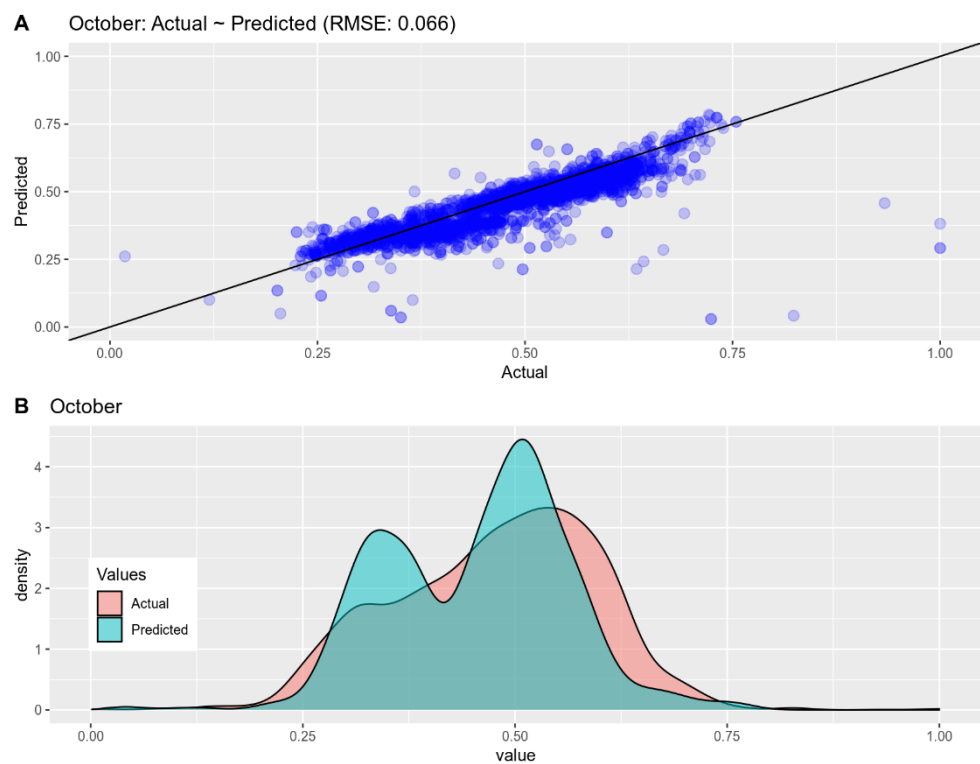**Figure A6.** Comparison between actual~predicted for June 2020.

Figure A7. Comparison between actua~predicted for July 2020.

Figure A8. Comparison between actual~predicted for August 2020.

Figure A9. Comparison between actual~predicted for September 2020.

Figure A10. Comparison between actual~predicted for October 2020.

A   November: Actual ~ Predicted (RMSE: 0.074)
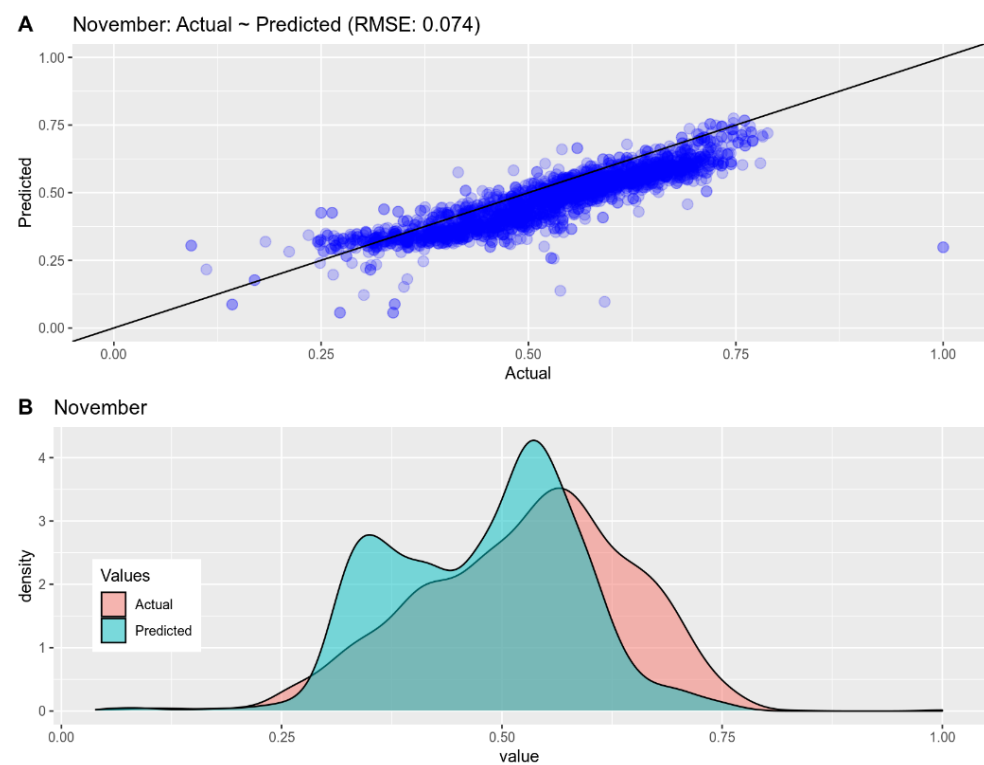
B   November

Figure A11. Comparison between actual~predicted for November 2020.

A   December: Actual ~ Predicted (RMSE: 0.077)
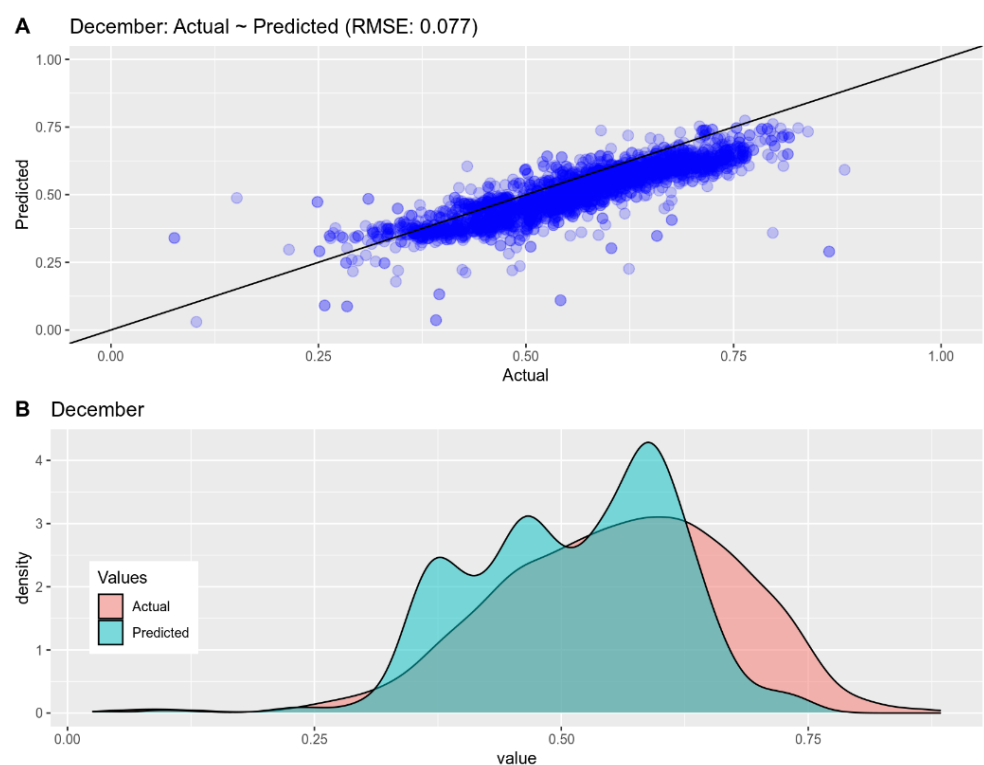
B   December

Figure A12. Comparison between actual~predicted for December 2020.

References

1.   Reddy, D.S.; Prasad, P.R.C. Prediction of Vegetation Dynamics Using NDVI Time Series Data and LSTM. *Model. Earth Syst. Environ.* **2018**, *4*, 409–419. [CrossRef]

2.   Salas, E.A.L.; Henebry, G.M. A New Approach for the Analysis of Hyperspectral Data: Theory and Sensitivity Analysis of the Moment Distance Method. *Remote Sens.* **2013**, *6*, 20. [CrossRef]

3. Ferchichi, A.; Abbes, A.B.; Barra, V.; Farah, I.R. Forecasting Vegetation Indices from Spatio-Temporal Remotely Sensed Data Using Deep Learning-Based Approaches: A Systematic Literature Review. *Ecol. Inform.* **2022**, *68*, 101552. [CrossRef]

4. Mutti, P.R.; Lúcio, P.S.; Dubreuil, V.; Bezerra, B.G. NDVI Time Series Stochastic Models for the Forecast of Vegetation Dynamics over Desertification Hotspots. *Int. J. Remote Sens.* **2020**, *41*, 2759–2788. [CrossRef]

5. Nay, J.; Burchfield, E.; Gilligan, J. A Machine-Learning Approach to Forecasting Remotely Sensed Vegetation Health. *Int. J. Remote Sens.* **2018**, *39*, 1800–1816. [CrossRef]

6. Pettorelli, N.; Vik, J.O.; Mysterud, A.; Gaillard, J.M.; Tucker, C.J.; Stenseth, N.C. Using the Satellite-Derived NDVI to Assess Ecological Responses to Environmental Change. *Trends Ecol. Evol.* **2005**, *20*, 503–510. [CrossRef]

7. Beck, P.S.A.; Atzberger, C.; Høgda, K.A.; Johansen, B.; Skidmore, A.K. Improved Monitoring of Vegetation Dynamics at Very High Latitudes: A New Method Using MODIS NDVI. *Remote Sens. Environ.* **2006**, *100*, 321–334. [CrossRef]

8. 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. [CrossRef]

9. Forkel, M.; Carvalhais, N.; Verbesselt, J.; Mahecha, M.D.; Neigh, C.S.R.; Reichstein, M. Trend Change Detection in NDVI Time Series: Effects of Inter-Annual Variability and Methodology. *Remote Sens.* **2013**, *5*, 2113. [CrossRef]

10. Bai, X.; Fu, J.; Li, Y.; Li, Z. Attributing Vegetation Change in an Arid and Cold Watershed with Complex Ecosystems in Northwest China. *Ecol. Indic.* **2022**, *138*, 108835. [CrossRef]

11. Wu, S.; Gao, X.; Lei, J.; Zhou, N.; Wang, Y. Spatial and Temporal Changes in the Normalized Difference Vegetation Index and Their Driving Factors in the Desert/Grassland Biome Transition Zone of the Sahel Region of Africa. *Remote Sens.* **2020**, *12*, 4119. [CrossRef]

12. Fernandez-Beltran, R.; Baidar, T.; Kang, J.; Pla, F. Rice-Yield Prediction with Multi-Temporal Sentinel-2 Data and 3D CNN: A Case Study in Nepal. *Remote Sens.* **2021**, *13*, 1391. [CrossRef]

13. Dang, C.; Liu, Y.; Yue, H.; Qian, J.X.; Zhu, R. Autumn Crop Yield Prediction Using Data-Driven Approaches: Support Vector Machines, Random Forest, and Deep Neural Network Methods. *Can. J. Remote Sens.* **2021**, *47*, 162–181. [CrossRef]

14. Sun, J.; Di, L.; Sun, Z.; Shen, Y.; Lai, Z. County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model. *Sensors* **2019**, *19*, 4363. [CrossRef]

15. Cao, J.; Zhang, Z.; Luo, Y.; Zhang, L.; Zhang, J.; Li, Z.; Tao, F. Wheat Yield Predictions at a County and Field Scale with Deep Learning, Machine Learning, and Google Earth Engine. *Eur. J. Agron.* **2021**, *123*, 126204. [CrossRef]

16. Hara, P.; Piekutowska, M.; Niedbała, G. Selection of Independent Variables for Crop Yield Prediction Using Artificial Neural Network Models with Remote Sensing Data. *Land* **2021**, *10*, 609. [CrossRef]

17. Zambrano, F.; Vrieling, A.; Nelson, A.; Meroni, M.; Tadesse, T. Prediction of Drought-Induced Reduction of Agricultural Productivity in Chile from MODIS, Rainfall Estimates, and Climate Oscillation Indices. *Remote Sens. Environ.* **2018**, *219*, 15–30. [CrossRef]

18. Nam, K.; Wang, F. The Performance of Using an Autoencoder for Prediction and Susceptibility Assessment of Landslides: A Case Study on Landslides Triggered by the 2018 Hokkaido Eastern Iburi Earthquake in Japan. *Geoenviron. Disasters* **2019**, *6*, 19. [CrossRef]

19. Malik, A.; Rao, M.R.; Puppala, N.; Koouri, P.; Thota, V.A.K.; Liu, Q.; Chiao, S.; Gao, J. Data-Driven Wildfire Risk Prediction in Northern California. *Atmosphere* **2021**, *12*, 109. [CrossRef]

20. Verbesselt, J.; Robinson, A.; Stone, C.; Culvenor, D. Forecasting Tree Mortality Using Change Metrics Derived from MODIS Satellite Data. *For. Ecol. Manag.* **2009**, *258*, 1166–1173. [CrossRef]

21. Rogers, B.M.; Solvik, K.; Hogg, E.H.; Ju, J.; Masek, J.G.; Michaelian, M.; Berner, L.T.; Goetz, S.J. Detecting Early Warning Signals of Tree Mortality in Boreal North America Using Multiscale Satellite Data. *Glob. Chang. Biol.* **2018**, *24*, 2284–2304. [CrossRef] [PubMed]

22. Soltani, K.; Amiri, A.; Zeynoddin, M.; Ebtehaj, I.; Gharabaghi, B.; Bonakdari, H. Forecasting Monthly Fluctuations of Lake Surface Areas Using Remote Sensing Techniques and Novel Machine Learning Methods. *Theor. Appl. Climatol.* **2021**, *143*, 713–735. [CrossRef]

23. Wang, Y.; Zhang, Z.; Feng, L.; Du, Q.; Runge, T. Combining Multi-Source Data and Machine Learning Approaches to Predict Winter Wheat Yield in the Conterminous United States. *Remote Sens.* **2020**, *12*, 1232. [CrossRef]

24. Karateke, S.; Zontul, M.; Bozkurt, N.E.; Aslan, Z. Wavelet-ANFIS Hybrid Model for MODIS NDVI Prediction. *J. Appl. Remote Sens.* **2021**, *15*, 024519. [CrossRef]

25. Cui, C.; Zhang, W.; Hong, Z.M.; Meng, L.K. Forecasting NDVI in Multiple Complex Areas Using Neural Network Techniques Combined Feature Engineering. *Int. J. Digit. Earth* **2020**, *13*, 1733–1749. [CrossRef]

26. Wang, W.; Hu, P.; Yang, Z.; Wang, J.; Zhao, J.; Zeng, Q.; Liu, H.; Yang, Q. Prediction of NDVI Dynamics under Different Ecological Water Supplementation Scenarios Based on a Long Short-Term Memory Network in the Zhalong Wetland, China. *J. Hydrol.* **2022**, *608*, 127626. [CrossRef]

27. Rhif, M.; Abbes, A.B.; Martinez, B.; Farah, I.R. Deep Learning Models Performance For NDVI Time Series Prediction: A Case Study On North West Tunisia. In Proceedings of the 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS), Tunis, Tunisia, 9–11 March 2020; pp. 9–12.

28. Rhif, M.; Ben Abbes, A.; Martinez, B.; Farah, I.R. A Deep Learning Approach for Forecasting Non-Stationary Big Remote Sensing Time Series. *Arab. J. Geosci.* **2020**, *13*, 1174. [CrossRef]

29. Reichstein, M.; Besnard, S.; Carvalhais, N.; Gans, F.; Jung, M.; Kraft, B.; Mahecha, M. Modelling Landsurface Time-Series with Recurrent Neural Nets. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 7640–7643.

30. Kraft, B.; Jung, M.; Körner, M.; Requena Mesa, C.; Cortés, J.; Reichstein, M. Identifying Dynamic Memory Effects on Vegetation State Using Recurrent Neural Networks. *Front. Big Data* **2019**, *2*, 31. [CrossRef]

31. Ahmad, R.; Yang, B.; Ettlin, G.; Berger, A.; Rodríguez-Bocca, P. A Machine-Learning Based ConvLSTM Architecture for NDVI Forecasting. *Int. Trans. Oper. Res.* **2020**. [CrossRef]

32. Fu, T.C. A Review on Time Series Data Mining. *Eng. Appl. Artif. Intell.* **2011**, *24*, 164–181. [CrossRef]

33. Alqahtani, A.; Ali, M.; Xie, X.; Jones, M.W. Deep Time-Series Clustering: A Review. *Electronics* **2021**, *10*, 3001. [CrossRef]

34. Paparrizos, J.; Gravano, L. Fast and Accurate Time-Series Clustering. *ACM Trans. Database Syst.* **2017**, *42*, 1–49. [CrossRef]

35. Izakian, H.; Pedrycz, W.; Jamal, I. Fuzzy Clustering of Time Series Data Using Dynamic Time Warping Distance. *Eng. Appl. Artif. Intell.* **2015**, *39*, 235–244. [CrossRef]

36. Javed, A.; Lee, B.S.; Rizzo, D.M. A Benchmark Study on Time Series Clustering. *Mach. Learn. Appl.* **2020**, *1*, 100001. [CrossRef]

37. Johnpaul, C.; Prasad, M.V.N.K.; Nickolas, S.; Gangadharan, G.R. Trendlets: A Novel Probabilistic Representational Structures for Clustering the Time Series Data. *Expert Syst. Appl.* **2020**, *145*, 113119. [CrossRef]

38. Baumann, M.; Ozdogan, M.; Richardson, A.D.; Radeloff, V.C. Phenology from Landsat When Data Is Scarce: Using MODIS and Dynamic Time-Warping to Combine Multi-Year Landsat Imagery to Derive Annual Phenology Curves. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *54*, 72–83. [CrossRef]

39. Guan, X.; Huang, C.; Liu, G.; Meng, X.; Liu, Q. Mapping Rice Cropping Systems in Vietnam Using an NDVI-Based Time-Series Similarity Measurement Based on DTW Distance. *Remote Sens.* **2016**, *8*, 19. [CrossRef]

40. Guan, X.; Liu, G.; Huang, C.; Meng, X.; Liu, Q.; Wu, C.; Ablat, X.; Chen, Z.; Wang, Q. An Open-Boundary Locally Weighted Dynamic Time Warping Method for Cropland Mapping. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 75. [CrossRef]

41. Vasilakos, C.; Tsekouras, G.E.; Palaiologou, P.; Kalabokidis, K. Neural-Network Time-Series Analysis of MODIS EVI for Post-Fire Vegetation Regrowth. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 420. [CrossRef]

42. Henderson, M.; Kalabokidis, K.; Marmaras, E.; Konstantinidis, P.; Marangudakis, M. Fire and Society: A Comparative Analysis of Wildfire in Greece and the United States. *Hum. Ecol. Rev.* **2005**, *12*, 169–182.

43. Didan, K. MOD13A3 MODIS/Terra Vegetation Indices Monthly L3 Global 1 km SIN Grid V006. Available online: https://doi.org/10.5067/MODIS/MOD13A3.006 (accessed on 15 February 2021).

44. Sakoe, H.; Chiba, S. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Trans. Acoust.* **1978**, *26*, 43–49. [CrossRef]

45. Chiu, S.L. Fuzzy Model Identification Based on Cluster Estimation. *J. Intell. Fuzzy Syst.* **1994**, *2*, 267–278. [CrossRef]

46. Tsekouras, G.E.; Anagnostopoulos, C.; Gavalas, D.; Dafhi, E. Classification of Web Documents Using Fuzzy Logic Categorical Data Clustering. In *IFIP International Federation for Information Processing*; Springer: Berlin/Heidelberg, Germany, 2007; Volume 247, pp. 93–100.

47. Cox, M.A.A.; Cox, T.F. Multidimensional Scaling. In *Handbook of Data Visualization*; Springer: Berlin/Heidelberg, Germany, 2008.

48. Sammon, J.W. A Nonlinear Mapping for Data Structure Analysis. *IEEE Trans. Comput.* **1969**, *C-18*, 401–409. [CrossRef]

49. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Springer: Boston, MA, USA, 1981; ISBN 978-1-4757-0452-5.

50. Karayiannis, N.B.; Bezdek, J.C. An Integrated Approach to Fuzzy Learning Vector Quantization and Fuzzy C-Means Clustering. *IEEE Trans. Fuzzy Syst.* **1997**, *5*, 622–628. [CrossRef]

51. Tsekouras, G.E.; Sarimveis, H. A New Approach for Measuring the Validity of the Fuzzy C-Means Algorithm. *Adv. Eng. Softw.* **2004**, *35*, 567–575. [CrossRef]

52. Hüsken, M.; Stagge, P. Recurrent Neural Networks for Time Series Classification. *Neurocomputing* **2003**, *50*, 223–235. [CrossRef]

53. Cai, X.; Zhang, N.; Venayagamoorthy, G.K.; Wunsch, D.C. Time Series Prediction with Recurrent Neural Networks Trained by a Hybrid PSO-EA Algorithm. *Neurocomputing* **2007**, *70*, 2342–2353. [CrossRef]

54. Qiu, R.; Wang, Y.; Rhoads, B.; Wang, D.; Qiu, W.; Tao, Y.; Wu, J. River Water Temperature Forecasting Using a Deep Learning Method. *J. Hydrol.* **2021**, *595*, 126016. [CrossRef]

55. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]

56. 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. [CrossRef]

57. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]

58. Steinier, J.; Termonia, Y.; Deltour, J. Smoothing and Differentiation of Data by Simplified Least Square Procedure. *Anal. Chem.* **1972**, *44*, 1906. [CrossRef] [PubMed]

59. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015.

60. Regmi, R.; Ma, Y.; Ma, W.; Baniya, B.; Bashir, B. Interannual Variation of NDVI, Precipitation and Temperature during the Growing Season. *Appl. Ecol. Environ. Sci.* **2020**, *8*, 218–228.

61. Wang, J.; Rich, P.M.; Price, K.P. Temporal Responses of NDVI to Precipitation and Temperature in the Central Great Plains, USA. *Int. J. Remote Sens.* **2003**, *24*, 2345–2364. [CrossRef]

62. Wu, T.; Feng, F.; Lin, Q.; Bai, H. Advanced Method to Capture the Time-Lag Effects between Annual NDVI and Precipitation Variation Using RNN in the Arid and Semi-Arid Grasslands. *Water* **2019**, *11*, 1789. [CrossRef]

63. Kang, L.; Di, L.; Deng, M.; Shao, Y.; Yu, G.; Shrestha, R. Use of Geographically Weighted Regression Model for Exploring Spatial Patterns and Local Factors behind NDVI-Precipitation Correlation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 4530–4538. [CrossRef]

64. Kerschke, P.; Hoos, H.H.; Neumann, F.; Trautmann, H. Automated Algorithm Selection: Survey and Perspectives. *Evol. Comput.* **2018**, *27*, 3–45. [CrossRef]

65. Kong, J.; Kowalczyk, W.; Nguyen, D.A.; Back, T.; Menzel, S. Hyperparameter Optimisation for Improving Classification under Class Imbalance. In Proceedings of the 2019 IEEE Symposium Series on Computational Intelligence, SSCI 2019, Xiamen, China, 6–9 December 2019.

66. Veloso, B.; Gama, J.; Malheiro, B.; Vinagre, J. Self Hyper-Parameter Tuning for Stream Recommendation Algorithms. In Proceedings of the Communications in Computer and Information Science, Dublin, Ireland, 10–14 September 2018; Volume 967, pp. 91–102.

67. Yuan, Y.; Wang, W.; Pang, W. A Genetic Algorithm with Tree-Structured Mutation for Hyperparameter Optimisation of Graph Neural Networks. In Proceedings of the 2021 IEEE Congress on Evolutionary Computation (CEC 2021), Kraków, Poland, 28 June–1 July 2021.

68. Gao, Y.; Quevedo, A.; Szantoi, Z.; Skutsch, M. Monitoring Forest Disturbance Using Time-Series MODIS NDVI in Michoacán, Mexico. *Geocarto Int.* **2021**, *36*, 1768–1784. [CrossRef]