# Exploiting the Temporal Dimension of Remotely Sensed Imagery with Deep Learning Models

Agustin Garcia Pereira
Insight Centre for Data
Analytics NUI Galway,
Republic of Ireland
agustin.garciapereira
@insight-centre.org

Lukasz Porwol
Insight Centre for Data
Analytics NUI Galway,
Republic of Ireland
lukasz.porwol
@insight-centre.org

Adegboyega Ojo
Maynooth University,
Republic of Ireland
adegboyega.ojo@mu.ie

Edward Curry
Insight Centre for Data
Analytics NUI Galway,
Republic of Ireland
edward.curry
@insight-centre.org

## Abstract

*The rapid rise of artificial intelligence and the increasing availability of open Earth Observation (EO) data present new opportunities to address important global problems such as the proliferation of agricultural systems which endanger ecological sustainability. Despite the plethora of satellite images describing a given location on earth every year, very few deep learning-based solutions have harnessed the temporal and sequential dynamics of land use to map agricultural practices. This paper compares different approaches to classify agricultural land use exploiting the temporal and spectral dimensions of EO data. The results show greater efficiency of the presented deep learning-based algorithms compared to state-of-the-art approaches when mapping agricultural classes.*

## 1. Introduction

Ecological agriculture practices such as intercropping, double cropping, crop rotations and the use of cover crops have shown to increase agriculture sustainability. The increasing tendency among farmers, decision-makers, and society in general to establish cropping systems that allow, not only the maximization of crop yield but also the provision of ecosystem benefits [1] is expected to rapidly increase the demand for spatial information about agricultural practices [2]. The 17 United Nation's Sustainable Development Goals (SDG) also tackle this issue. These goals present a list of indicators to help assess the progress made. For instance, indicators 2.4.1, 15.1.1, and 15.3.1 relate to land use and land cover data. To this end, the literature suggests that remote sensing has been an effective tool for monitoring the land surface properties resulting from human practices, and can greatly contribute to measuring these indicators in a cost-effective way [3]. Despite significant progress made in the area of remote sensing

and agriculture, an extensive literature review shows that only 9% of the total publications in the domain focus on cropping practices [4]. Moreover, due to the wide variety of agricultural practices and the difficulty and complexity in providing descriptions for large areas using satellite data, studies have been mostly limited to case studies [5]. Despite the fact that the earth is continuously monitored by satellites, drones and different types of sensors, most recent AI models or classifiers used in operational mapping generally use only single date spectral data for classification [6]. In this paper, we address the research gap identified in the literature by comparing different deep learning approaches to classify agricultural land use and practices and harnessing the temporal and spectral dimensions of earth observation data. More specifically, we address the following research questions: i) How much can agricultural classification performance be enhanced by considering the temporal dimension and not only the spectral dimension?; ii) Can a deep learning architecture outperform the state of the art algorithms used in the remote sensing domain?; iii) What structure will this architecture have, and which parameters should be used? iv) Can vegetation indices help improve the classification performance?

## 2. Background

The application of Deep Learning (DL) has shown outstanding results in many fields including remote sensing. Harnessing the ability of DL models to learn feature representations exclusively from raw data without the need of domain-specific knowledge, Deep Neural Networks (DNNs) have been used in tasks including image classification, object detection, semantic segmentation, classification from time series and anomaly detection [7] in remotely sensed imagery. Previous works in this domain used rule-based classification algorithms, like decision trees, and multitemporal vegetation indexes derived from spectral satellite data to classify vegetation cover [8], [9].

H CSS

However, these approaches are oblivious to the temporal dimension because they do not consider the sequential relationship of multitemporal observations. Other studies have extracted temporal features or phenological metrics from the time series and reported better accuracy results when compared to using raw time-series data [10]. Although many approaches to manually extracting temporal features are described in the literature, they face significant problems which are listed by Zhong et al. [11]: i) human experience and domain knowledge is required to manually design models and extract features. Moreover, features from a general model might not be suitable for a more specific problem; ii) manual feature engineering is time-consuming since human supervision is required when environmental and weather conditions change, and iii) fixed predefined models and mathematical constraints limit the ability to handle disparate patterns. To identify temporal patterns, humans do not make their decisions based on a single mathematical model or a group of them, and they struggle to list all the rules they apply during the process [12]. In this context, DL has shown to be able to discover complex data structures in high-dimensional data, making this technology applicable to many domains of science, business and government [13]. In remote sensing, DL and specifically Convolutional Neural Networks (CNNs), have been applied in different ways in the context of remote sensing. Two-dimensional CNNs have been widely adopted to extract 2-dimensional spatial features from the width and height dimensions of satellite imagery, pushing forward the state of the art in areas such as semantic segmentation [14], land use classification [15], and object detection [16]. W. Hu et al. [17] and Guidici et al. [18] utilized one-dimensional convolutions in the spectral dimension, whereas Y. Li et al. [19] used 3-dimensional convolutions on the spectral and spatial domains. Thus, convolutions in literature are mostly applied to the spectral and spatial dimensions, but rarely to the temporal dimension of remotely sensed time series.

## 3. Data Preparation
### 3.1. Context

The setting of the study, as the consequence of our ground truth data choices, is a surface of 4466 km2 or 1724 square miles in Sacramento County, in the west part of the United States of America and encompasses a one-year period ranging from January 2015 to December 2015. The area is fully covered by the data of Paths 44, 43, and Rows 34, 33 in the Worldwide Reference System-2 (WRS-2).

### 3.2. Data sources

In the search for data sources, we considered the metadata quality, distribution format, and costs concluding that California Land Use Survey was the dataset that best met our needs. This dataset is free to use and describes the land use in detail. Moreover, it presents data for 87 different agricultural classes in California, together with metadata about irrigation and agricultural practices, such as intercropping, double cropping, triple cropping, and mixed land use. Additionally, more than 95% of the land surveyed was visited during the process. Since different satellites have been launched at different dates, a match between the ground truth data and the availability of remotely sensed data represents a strong limitation at the time of selecting a satellite product. Considering that the land use survey described before is based on the agricultural fields for the year 2015, a combination of Landsat-7 and Landsat-8 was deemed as the best option. Among the available Landsat products, Landsat Level 2 is a research-quality, application-ready science product derived from Landsat Level 1 data [20]. A total of 178 Surface Reflectance image products were downloaded for the region of interest delimited in Fig. 1 for the year 2015. From this set, 88 images correspond to Landsat-8 and a total of 90 images correspond to Landsat-7. In terms of spectral data, six bands were selected for the application of this study. The blue band was only used during EVI computation and spectral indices evaluation. Despite the spectral ranges of the different bands are slightly different between Landsat-7 and Landsat-8, these differences have been studied in [21] suggesting that their impact on a model depends on the sensitivity of the model in question. Studies have shown the insignificant impact of these differences on classification models [11], [22].

### 3.3. Pre-processing

Two main approaches have been proposed in the literature to classify remotely sensed images: a pixel-based and an object-based approach [23]. In the first, the classification algorithms exploit the spectral differences between classes to classify a pixel [24]. On the other hand, an object-based approach also exploits the spatial and textural information of a group of pixels grouped in a meaningful way, relying heavily on a previous accurate segmentation process [23]. Despite an object-based approach might be a valid option for a single year land use classification, where the agricultural fields are well defined [25], the segmentation process depends on the assumption that agricultural fields will hold the same practices over the time period and that their boundaries will not change. This is not particularly true since the farmers can decide, for example, to grow corn two times

in the year, but when sowing for the second time, to leave a portion of that field idle. Thus, we decided that a pixel-based approach will be better in our case. We used the pipeline proposed in our previous contribution [26] to preprocess the data, creating a 2-day temporal grid and using linear interpolation to fill the gaps. The process artificially increased the observations from 46 to 176.
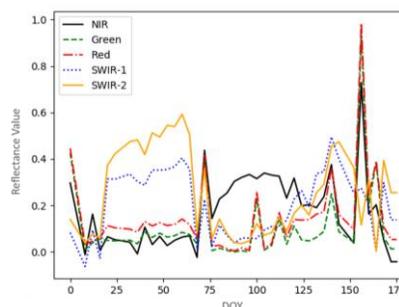
### 3.4. Datasets Created

The first dataset created, from now on referred to as "Dataset 1", consists of 32 different agricultural land use classes. During the confection of this dataset, we focused on different crops that were grown in a "Single Cropping" approach. The rotation of crops across different years often leads to better yields due to soil fertility improvements [27], while also reducing the external dependency on agrochemicals [28], [29]. Following the data manual documentation from the California Land Use Survey, we first filtered the ground truth data removing all the classes that did not represent an agricultural field. From a list of 47 different classes, we removed the ones that contained equal or less than 10 polygons in total. Thus, the final dataset is composed of the 32 best-represented classes. Class imbalances are present in our dataset, with the highest number of pixels for "Mixed Pasture" class, and the lowest number of pixels for "Bush berries" class. For each pixel, a multivariate time-series was created using 6 different spectral bands (Near Infrared, Red, Blue, Green, SWIR-1 and SWIR-2). Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) were also calculated. Table 1 presents the total amount of polygons and the total amount of pixels sampled. Figure 1 presents a corn time series profile using NIR, Red, Green, SWIR-1 and SWIR-2 bands.

### Table 1. Dataset 1 Metadata

| Code | Label | Polygons No. | Pixels No. |
|---|---|---|---|
| 0 | Miscellaneous grain and hay | 26 | 2246 |
| 1 | Safflower | 65 | 7051 |
| 2 | Corn | 774 | 90804 |
| 3 | Grain Sorghum | 17 | 2065 |
| 4 | Sudan | 35 | 4587 |
| 5 | Beans | 25 | 2775 |
| 6 | Sunflowers | 14 | 1285 |
| 7 | Alfalfa | 523 | 62949 |
| 8 | Clover | 38 | 4576 |
| 9 | Mixed Pasture | 1338 | 101675 |
| 10 | Native Pasture | 26 | 478 |
| 11 | Miscellaneous grasses | 13 | 914 |
| 12 | Melons, squash, and cucumbers | 50 | 2678 |
| 13 | Potatoes | 16 | 2854 |
| 14 | Tomatoes | 102 | 18203 |
| 15 | Flowers, nursery and Christmas tree farms | 41 | 2485 |
| 16 | Mixed 4+ | 221 | 2758 |
| 17 | Miscellaneous truck | 113 | 1099 |
| 18 | Bush berries | 11 | 118 |
| 19 | Strawberries | 97 | 492 |
| 20 | Apples | 13 | 1031 |
| 21 | Cherries | 85 | 5258 |
| 22 | Peaches and nectarines | 11 | 135 |
| 23 | Pears | 418 | 23511 |
| 24 | Miscellaneous deciduous | 111 | 1760 |
| 25 | Mixed deciduous | 89 | 527 |
| 26 | Almonds | 17 | 1371 |
| 27 | Walnuts | 57 | 4812 |
| 28 | Pistachios | 13 | 2339 |
| 29 | Olives | 46 | 691 |
| 30 | Eucalyptus | 77 | 5169 |
| 31 | Idle 1 | 233 | 11489 |
| | **Total** | **4715** | **370185** |



**Figure 1. Corn Time Series Profile**

The second dataset created, from now on referred to as "Dataset 2", characterizes two agricultural practices: single cropping and double cropping, within the same year. Double cropping practice is an important sustainable practice that aims at reducing the fallow periods of the land, exploiting solar energy to enhance the quality of the soil and preventing soil erosion [30]. During the dataset confection, we first filtered the ground truth data removing classes that were not representing agricultural fields. Then we identified the fields where double cropping practice and single cropping practice took place during the year. Table 2 presents the total amount of polygons and the total amount of pixels sampled. In this case, as the double-cropping class was under-represented, we selected the total amount of double cropping pixels as the limit to be sampled from the single cropping polygons. The sampling was done randomly, maximizing the diversity of single cropping polygons and not exceeding the limit of 29596 pixels. In the end, the total number of pixels sampled was 29596 for both classes.

### Table 2. Dataset 2 Metadata

| Code | Label | Polygons No. | Pixels No. |
|---|---|---|---|
| 0 | Double | 256 | 29596 |
| 1 | Single | 6784 | 29596 |
| **Total** | | **7040** | **59192** |

# 4. Approach

## 4.1. Methodology

Our methodology focuses on experimentation aimed at comparing the performance of a popular algorithm used in remote sensing classification with new approaches in temporal deep learning classification. We study the performance impact of data availability constraints, spectral indices calculation, as well as the impact of architectural designs, main parameters settings, and dimensionality reduction techniques.

In terms of research infrastructure setup, the models for all the experiments were trained using the Azure cloud infrastructure provided by Microsoft AI for Earth grant program. The virtual machine uses an NVIDIA Tesla K80 GPU card. Each dataset used was partitioned in three sets: 60% training set, 30% testing set and 10% validation set. The training set was used to train each classifier. The validation set was used to monitor loss and accuracy avoiding overfitting, and the testing set was used to evaluate the final results. This partitioning was randomly done 30 times, and the same random partition configuration was used for all the experiments involving a given dataset. In the case of Random Forest (RF), this set was discarded. M. Rubwurm et al. [31] defined two principles that these types of subsets should follow. First, they need to be independent of each other. Secondly, the class distributions in all the datasets should be sufficiently similar. To respect the independence principle, the division of data was done at the polygon level, meaning that no pixels from the same polygon are in the training, testing or validation set at the same time.

In all the experiments performed we applied the Overall Accuracy as the evaluation criterion, calculated as the number of the corrected classified pixels divided by the number of the total pixels. Results are then presented as an average across all 30 shuffle configurations and multiplied by 100 ± one standard deviation.

Because in practice, Neural Networks parameters usually exceed the number of data samples, they can potentially fit any training data. This leads to an overfitting problem, a model that performs well in the data that has already seen but does not generalize well with unseen data [32]. During our experiments with neural networks, we implemented an early stopping technique to mitigate this problem. Because stopping training too early may reduce variance but increment bias and stopping too late may reduce bias but increment variance [33], we utilized the validation set accuracy to stop the learning when the validation loss increases or the validation accuracy decreases over a number of epochs, in our case set to zero. Data normalization attempts to give all attributes an equal weight and avoid the dependence on the choice of measurement units [34]. Z-score normalization has been widely adopted in machine learning time series problems [35], subtracting the mean and dividing by the standard deviation for each time series. This approach has been used in remotely sensed time series, normalizing each feature, where each timestamp is considered a different feature. Pelletier et al. [36] indicated that this approach leads to a loss of significance in the magnitude, an important aspect for vegetation mapping, where certain classes have higher spectral values than others. Authors used a min-max normalization (a subtraction of the minimum then a division by the range maximum minus the minimum), but instead of using the minimum and the maximum values for each feature, they proposed to use a 2% and 98% percentile, respectively. This decision was based on the fact that this type of normalization is very sensitive to extreme values. We adopted this normalization approach to preparing our datasets.

## 4.2. Models and Architectures

### 4.2.1. Random Forest

Random Forest (RF) is an ensemble classification method, which means that uses not only one, but many classifiers. It consists of a combination of tree predictors where each one depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [37]. Random Forest uses a combination of *k* binary CART trees (Classification and Regression Trees). These trees are built without pruning, and at each node, a subset of randomly selected variables is used as input reducing the computational complexity of the algorithm and the correlation between the trees. One common value for splitting each RF node is the square root of the number of input variables (denoted by *m*) [38]. This recursive process is repeated on each derived subset until a maximum depth (*max_depth*) is reached or when the number of samples at a node is less than a certain threshold (*min_samples*) [39]. Random Forest has been successfully used in Remote Sensing for different classification tasks. Schmidt et al. [40] used several machine learning techniques to create maps of cropping activity for the period 1987-2015 using Landsat imagery. In this study, Random Forest performed better when compared with SVM, multinomial logistic regression, and decision-tree classifier. Tian et al. [41] used Random Forest to map wetland landcover using multiple sources of remotely sensed data. Random Forest accuracy surpassed SVM and Artificial Neural Networks by more than 10%. Chan et al. [42] compared the performance of Random Forest and Adaboost, both tree-

based ensemble methods, to classify ecotopes using hyperspectral data. Their results show that both algorithms perform similarly in terms of accuracy, outperforming neural network classifier. However, Random Forest results were more stable. Due to the good classifications results and the capacity to handle high dimensionality data, RF has settled as a popular algorithm in the remote sensing domain [38], and we use it as a gold standard to compare the performance of the deep learning architectures studied.
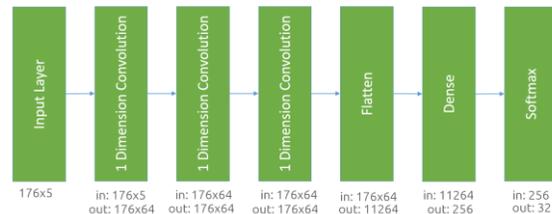
### 4.2.2. One-dimensional CNN

Deep learning methods are characterized by Neural Networks built using more than two hidden layers. The composition of simple but non-linear modules allows DNNs to learn raw data representation at many levels. Starting from the raw input, each level transforms the representation into a more abstract level. In this way, many complex functions can be learned. CNNs are DNNs where one or more convolutional layers are used. Convolution can be seen as applying and sliding a filter over different dimensions of the data representation. An over-simplified supervised learning process consists of modifying the adjustable parameters (or weights) of the network architecture to minimize an error function. This error function can be thought of as a representation of the distance between the output score produced by a given input, and the desired pattern of scores [13]. This optimization is performed by the learning algorithm by computing a gradient vector, that represents for each weight, the positive or negative impact on the error function when the weight value is slightly increased.

Scarce studies have addressed the use of deep learning applied to the temporal and spectral domain of remotely sensed imagery. Liheng et al. [11] have exploited the intrinsic characteristics of time-series data to describe seasonal patterns and sequential relationships for classifying summer crops. They developed different deep neural network architectures and used Enhanced Vegetation Index (EVI) calculated from Landsat Level 2 product imagery bands and ground in-situ data from California Department of Water Resources. Their results, based on an architecture that includes three one-dimensional convolution layers and an Inception Module (IM), outperformed traditional algorithms for land use classification including XGBoost, Random Forest, Support Vector Machine and Long-short Term Memory (LSTM) network. Pelletier et al. [36] proposed a temporal convolutional neural network constructed with three convolutional layers, a dense layer and finally, a SoftMax layer. Different from [11], the authors of this study used three spectral bands of

the available satellite imagery. Results show that the proposed architecture outperformed Random Forest algorithm by 2 to 3 % and based on the evidence gathered they point out the importance of using both spectral and temporal dimensions when computing the convolutions. Cai et al. [22] developed a deep learning architecture to train a model able to classify corn and soybean fields. They used a combination of Landsat-5, Landsat-7 and Landsat-8 satellite images time-series covering a period of sixteen years. They approach consisted of averaging the six spectral bands used in the study at the field level, thus reducing missing data and improving availability. They report an overall accuracy of 97%. Inspired by these studies, we define two main general architectures to study. Architecture 1 is a one-dimensional CNN proposed by Pelletier et al. [36] based on stacking different numbers of Convolutional Layers. Figure 2 depicts a general view of the architecture for three convolutional layers. For simplicity, we excluded from the diagram the Batch Normalization, Activation, and Dropout layers. This sequence is followed after each 1 Dimension Convolution and after the Dense layer, as well. Table 3 presents a list of parameters and values used for the network configuration.

#### Table 3. Architecture Parameters

| Batch size | 32 | Beta_1 | 0.9 | l2 rate | 1.00E-06 |
|---|---|---|---|---|---|
| Epochs | 20 | Beta_2 | 0.999 | Dropout rate | 0.5 |
| Optimizer | Adam | Loss function | categorical_crossentropy | Kernel/filter size | 5 |
| Learning rate | 0.001 | Kernel regulirizer | l2 | Activation | ReLu |



176x5  in: 176x5 out: 176x64  in: 176x64 out: 176x64  in: 176x64 out: 176x64  in: 176x64 out: 11264  in: 11264 out: 256  in: 256 out: 32

**Figure 2. General Architecture 1**

Architecture 2 (Figure 3) is based on stacking IM and Convolutional layers. As explained by Szegedy et al. [43], augmenting the size of a Neural Network is one of the simplest ways to improve its performance. However, increasing the width and the depth of a neural network derives in a larger number of parameters and the increase in computational resources. A large number of parameters makes the neural network more prone to overfitting [44]. To overcome these problems, Szegedy et al. [43] proposed to move from fully connected neural networks to more sparse ones. Based on this idea, they presented the Inception architecture based on IMs, characterized by the use of skip connections. To tackle the computational resources inefficiency and based on the embedding's benefits, they added 1x1 convolutions before the 3x3 and 5x5 expensive convolutions to compute dimensionality reduction, in the context of two-

dimensional images classification. We adapted the Inception Module with dimensionality reduction to one-dimensional convolutions as depicted in Figure 4. We used strides of 1 for convolutions and max-pooling layers, 64 filters for convolutions and a window of size 3 for the max-pooling layer. The dropout rate after the Filter Concatenation was set to 0.4.
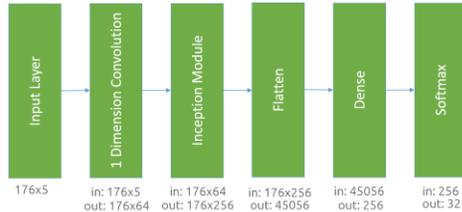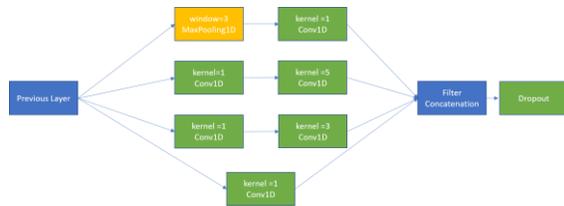


**Figure 3. General Architecture 2**



**Figure 4. 1D Inception Module with Dimensionality Reduction**

# 5. Evaluation

## 5.1. Single Date Classification

To measure the benefits of using multitemporal spectral data, we performed a classification task using single date acquisition data at the pixel level. We used RF to classify Dataset 1 using single date spectral information from NIR, Red, Green, SWIR-1, and SWIR-2. We trained 30 RF classifiers following the 30 different dataset split configuration and we averaged the results. We did the same for all 176 acquisition dates of our dataset. Accuracy results over time are presented in Figure 5.



**Figure 5. Single Date RF Classification**

The highest accuracy achieved by the RF classifier on a single date was $64.95 \pm 0.82$. This measure corresponds to measurement 96 of our dataset. If we

translate this back to real dates, we can see that date 96 corresponds to the 26th of July (day of the year 206). It can also be observed that 80% of the best 15 single date classification are between the days 20th of May and 27th of July, days of the year 140 and 207, respectively.

## 5.2. RF vs One-dimensional CNNs

To compare the performance of Random Forest and one-dimensional CNNs algorithms described in Section 4.2 we conducted a series of experiments using Dataset 1 and Dataset 2 presented in Section 3.4. Using Dataset 1 and five spectral bands we evaluated Architecture 1 using a different number of Convolutional Layers and a batch size of 128. Table 4 presents the results for one, two, three, four, and five convolutional layers stacked one after each other.

**Table 4. Conv1D-based CNN Results**

|  | 1 x Conv1D | 2 x Conv1D | 3 x Conv1D | 4 x Conv1D | 5 x Conv1D |
|---|---|---|---|---|---|
| **Accuracy** | 85.57 ± 1.11 | 85.78 ± 1.15 | 85.77 ± 0.99 | 85.55 ± 1.51 | 85.27 ± 1.24 |
| **Test Loss** | 0.64 ± 0.08 | 0.61 ± 0.08 | 0.64 ± 0.08 | 0.66 ± 0.07 | 0.67 ± 0.08 |

Dataset 1 was used to test different configurations of Inception Modules and Convolutional Layers using a batch size of 256. Table 5 presents the accuracy and loss results for a single Naïve Inception Module without dimensionality reduction (NI), a single IM with dimensionality reduction (IM), an IM followed by a single one-dimensional convolution layer (IM + Conv1D), a Conv1D followed by a single IM (Conv1D + IM), an IM followed by two one-dimensional convolution layers (IM + 2 x Conv1D) and we also present the results for a 3 fully connected layers architecture (3 x FC), for comparison.

**Table 5. Inception Modules and Convolutional Layers Results**

|  | Accuracy | Test Loss |
|---|---|---|
| **NI** | 85.54 ± 1.17 | 0.66 ± 0.07 |
| **IM** | 85.85 ± 1.11 | 0.67 ± 0.09 |
| **IM + Conv1D** | 86.14 ± 1.34 | 0.62 ± 0.06 |
| **Conv1D + IM** | 85.02 ± 1.33 | 0.65 ± 0.07 |
| **IM + 2 x Conv1D** | 85.99 ± 1.01 | 0.62 ± 0.07 |
| **3 x FC (256 units)** | 82.2 ± 1.69 | 0.73 ± 0.07 |
| **3 x FC (1024 units)** | 82.88 ± 1.4 | 0.82 ± 0.1 |

Rodriguez-Galiano et al. [45] demonstrated that the number of trees (k) in RF is directly proportional to the classifiers' accuracy up to the number of 100 trees. Once this value is reached, the generalization error converges. Pelletier et al. [39] studied different values of k, ranging from 50 to 400 and also concluded that this value can be set to 100 without a major accuracy loss. The *m* parameter value suggested by the literature is the square

root of *p*, where *p* is the number of features [46]. However, small values of m have shown very good performance due to the reduction correlation among individual trees [108]. Finally, the values for *max_depth* and *min_samples* have been less explored in literature. Pelletier et al. [39] used a *max_depth* of 25, and a min_samples of 10 or 25, and showed that the accuracy impact of these parameters' selection are low. Therefore, we studied the impact of parameter k (Table 6) on Dataset 1 using RF sklearn python library with parameters $m=\sqrt{p}$, *where p=n_features; max_depth = None, nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples; min_samples_split = 2.*

### Table 6. RF Parameter k Evaluation

| k=10 | k=100 | k=500 |
|------|-------|-------|
| 82.79 ± 1.19 | 85.2 ± 1.18 | 85.41 ± 1.16 |

In Table 7 we compared the performance of the best configurations from Architecture 1 and Architecture 2 with RF using a batch size of 256 and a k equals to 100, while in Table 8 we compared both architectures performances using different batch sizes and five spectral bands.

### Table 7. Algorithms and Spectral Bands Combinations Comparison

| Bands | 3 x Conv1D | IM + Conv1D | RF (k=100) |
|-------|-----------|-------------|------------|
| **R** | 81.67 ± 1.35 | 81.96 ± 1.17 | 82.29 ± 1.28 |
| **G** | 80.75 ± 1.38 | 81.36 ± 1.24 | 81.72 ± 1.25 |
| **NIR** | 80.85 ± 1.44 | 81.34 ± 1.49 | 81.15 ± 1.24 |
| **SWIR-1** | 81.27 ± 1.58 | 81.12 ± 1.01 | 80.82 ± 1.3 |
| **SWIR-2** | 81.35 ± 1.62 | 81.46 ± 1.49 | 81.24 ± 1.28 |
| **G-R-S1-S2** | 84.43 ± 1.19 | 85.72 ± 1.18 | 84.36 ± 1.2 |
| **NIR-G-R-S1** | 84.87 ± 1.47 | 85.74 ± 1.16 | 85.07 ± 1.16 |
| **NIR-G-R-S2** | 85.22 ± 1.38 | 85.63 ± 1.22 | 85.04 ± 1.19 |
| **NIR-G-S1-S2** | 85.35 ± 0.97 | 85.72 ± 1.14 | 85.1 ± 1.18 |
| **NIR-R-S1-S2** | 85.28 ± 1.18 | 85.96 ± 1.18 | 84.85 ± 1.18 |
| **ALL** | 85.37 ± 1.16 | 86.14 ± 1.34 | 85.2 ± 1.18 |

### Table 8. Batch Sizes Comparison

| Batch Size | Measure | IM + Conv1D | 3 x Conv1D |
|------------|---------|-------------|------------|
| **32** | Accuracy | 86.39 ± 1.13 | 86.02 ± 1.23 |
| | Test Loss | 0.66 ± 0.09 | 0.66 ± 0.08 |
| **128** | Accuracy | 86.1 ± 1.21 | 85.77 ± 0.99 |
| | Test Loss | 0.63 ± 0.07 | 0.64 ± 0.08 |
| **256** | Accuracy | 86.14 ± 1.34 | 85.37 ± 1.16 |
| | Test Loss | 0.62 ± 0.06 | 0.81 ± 0.09 |

We classified Dataset 2 using the best configuration for Architecture 1 and Architecture 2, and RF. The results are presented in Table 9.

### Table 9. Dataset 2 Evaluation

| | 3xConv1D | | IM+Conv1D | | RF (k=100) |
|-------|----------|----------|-----------|----------|------------|
| **Batch Size** | 32 | 256 | 32 | 256 | - |
| **Acc.** | 94.19 ± 1.65 | 93.76 ± 1.68 | 94.38 ± 1.65 | 93.96 ± 1.5 | 93.02 ± 1.92 |

## 5.3. RF Dimensionality Reduction

Although Random Forest classifier has proven to be robust to the use of high data dimensionality, some authors have suggested that prior features filtering can improve the classifier accuracy [38]. Many methods can be used to remove redundant, noisy and irrelevant features, such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), or Minimum Noise Fraction (MNF) analysis [47], among others. However, Random Forest built-in feature importance identification method can be successfully used to filter most relevant features [48]. In this experiment, we analysed the performance of RF algorithm over Dataset 1 and five spectral bands when reducing the dimensionality of the data using five different thresholds of features importance: 10%, 25%, 50%, 75% and 100%, as shown in Table 10. We implemented the features filtering using the embedded feature selection method of the RF classifier.

### Table 10. Dimensionality Reduction

| Percentage | Dimensions | Accuracy |
|------------|-----------|----------|
| 10% | 88 | 84.04 ± 1.06 |
| 25 % | 220 | 85.21 ± 1.18 |
| 50 % | 440 | 85.47 ± 1.14 |
| 75 % | 660 | 85.33 ± 1.19 |
| 100 % | 880 | 85.2 ± 1.18 |

We first trained the algorithm with the original dimensions, getting a list of features importance. This list was then used to filter the most important features using the values described before. This process was repeated for the thirty different partitioning configurations and the results were then averaged.

## 5.4. Vegetation Indices vs. Raw Data

Vegetation Indices (VIs) derived from remotely sensed spectral data are quite simple and effective algorithms for quantitative and qualitative evaluations of vegetation cover, vigour, and growth dynamics, among other applications [49]. One of the most used indices calculated from multispectral information is the Normalized Difference Vegetation Index (NDVI) [50], which is based on the difference between the maximum absorption of radiation in the red band due to chlorophyll pigments, and the maximum reflectance in the near-infrared band as a result of leaf cellular structure [51]. The NDVI is susceptible to many errors and uncertainty over variable atmospheric and canopy background conditions [52]. Liu et al. [52] proposed a modified

NDVI equation based on a feedback-based approach that incorporates both background adjustment and atmospheric resistance concepts into the NDVI, the Enhanced Vegetation Index (EVI). The EVI has thus been considered a modified NDVI with improved sensitivity to high biomass regions and improved vegetation monitoring capability and a reduction in atmospheric influences [53].

### Table 11. Spectral Indices

|  | 3xConv1D | IM+Conv1D | RF (k=10) | RF (k=100) |
|---|---|---|---|---|
| NDVI | 83.4 ± 1.17 | 83.02 ± 1.29 | 81.34 ± 1.34 | 83.59 ± 1.39 |
| NIR-R | 85.19 ± 0.98 | 84.95 ± 1.2 | 81.98 ± 1.32 | 84.42 ± 1.22 |
| EVI | 83.25 ± 1.13 | 83.15 ± 1.19 | 80.69 ± 1.42 | 82.97 ± 1.36 |
| NIR-R-B | 85.19 ± 0.98 | 85.24 ± 1.11 | 82.11 ± 1.3 | 84.53 ± 1.22 |

In this experiment, we compared the performance of RF and 1D-CNN for classifying Dataset 1 using the calculated NDVI and EVI indices as well as the results of performing this classification using only the raw bands' data involved in each index. Thus, for NDVI we compared the performance with NIR and Red band, and for EVI we used NIR, Red and Blue bands. Table 11 presents the results obtained.

## 6. Discussion

In this study, we investigated and compared different deep learning approaches to classify agricultural land use and practices exploiting both the temporal and spectral dimensions of EO data. In our experiments, we ensured that all the results are of most statistical significance by running each experiment thirty times with different dataset splitting configuration. That required substantial computational resources, covered by Microsoft Azure grant, secured explicitly for this research. Therefore, we claim the high relevance of our results over smaller experimental works performed to the date. We showed that classification errors were largely reduced when introducing temporal information at the pixel level. The best accuracy obtained by RF with single date information was almost 65%, whereas more than 85% accuracy was obtained when temporal information was included. Architecture 1 performs better when utilizing two and three convolutional layers. We chose three layers to ensure better stability of the results. We show that the addition of dimensionality reduction in the IM positively impacts the performance. For the different configurations' options, we tried for Architecture 2, placing the IM in the first place seems to have a significant, positive impact on the performance. The results were already better when using the IM alone, and are improved with the addition of a convolutional layer. Switching the order of these two, or adding an extra convolutional layer does not improve accuracy. When comparing the results of the three fully connected layers architecture with the rest, we can see the importance of using convolutions as features extractors in the network. Even though a big part of our datasets consists of synthetic data created during the linear interpolation process to fill the gaps, RF seems to handle high dimensionality of data efficiently. The best result was obtained when using the best 50% features, but the difference is not significant compared with the performance when using the entire features set. Table 7 shows that no significant performance differences exist between RF and CNNs when utilizing single spectral bands information. However, when combining them, CNNs, and specially IM+Conv1D, outperformed RF significantly, with specific combinations having differences over 1%. Moreover, when comparing to the values obtained with RF using k equals to 500, or when applying dimensionality reduction, RF could not reach CNN's performance. Parameter k has a big impact on RF performance. However, for values over 100, the difference is not significant while training cost increases accordingly. Spectral indices calculation does not improve classification performance, as shown in Table 11. Furthermore, using raw spectral data, improves the accuracy in all the algorithms, with CNN's taking further advantage of them than RF. The selection of batch sizes has an impact on the model's performance. We presented in Table 8 the results of training Architecture 1 and Architecture 2 optimizations with three different batch sizes: 32, 128, and 256. We showed that both architectures performed better when using a batch size of 32 and both exhibit a similar test loss for the small batch size, with IM+Conv1D outperforming. The major limitation of the work relates to transferability of the models created since our datasets are specific to California and the mentioned satellites. The results presented demonstrate the efficiency of the algorithms described when compared to state-of-the-art approaches and showcase the potential of these algorithms for mapping sustainable agricultural practices. Therefore, this work makes a substantial step forward in addressing the adoption of deep learning for agriculture in the remote sensing domain. Comparing our experiments with other related studies, ours have made use of publicly available satellite imagery, making a transfer learning approach that would make the process of fitting the models for other geographical locations, viable. While some studies have only focused on the classification of a few agricultural types, we have trained a superior single model that can classify 32 agricultural classes with more than 86% accuracy. None of the studies analyzed before had classified pure temporal characteristics, as we did with Dataset 2 with more than 94% accuracy. Future work includes performing two and three-dimensional convolutions over the spectral, temporal and spatial dimensions.

# 7. Conclusions

In this paper, we evaluated the performance of different one-dimensional CNNs architectures and Random Forest for agricultural land use and agricultural practices classification over more than 400 thousand multispectral remotely-sensed time series. To answer our first research question, we demonstrated that inclusion of temporal information to RF-based classification boosted the accuracy from 65% to 85% and allowed more than 86% accuracy for CNN-based architecture over Dataset 1. That result also answers our second research question by showing that deep learning architecture (CNN) indeed can outperform the state of the art algorithm in the domain (RF). In particular, to answer our third question, we showed that specific one-dimensional CNNs significantly outperformed the established RF-based approach when utilizing multispectral data, even when applying RF dimensionality reduction. Finally, to answer question four, our results show that the calculation of spectral indices does not improve classification performance.

# 8. References

[1] M. E. Schipanski *et al.*, "A framework for evaluating ecosystem services provided by cover crops in agroecosystems," *Agric. Syst.*, vol. 125, no. March, pp. 12–22, 2014, doi: 10.1016/j.agsy.2013.11.004.

[2] J. De Baerdemaeker, *Precision agriculture technology and robotics for good agricultural practices*, vol. 1, no. PART 1. IFAC, 2013.

[3] M. Tassopoulou, N. Verde, G. Mallinis, C. Georgiadis, D. Kaimaris, and P. Patias, "Demonstrating the potential of remote sensing to support sustainable development goals implementation: case studies over Greece," no. June 2019, p. 42, 2019, doi: 10.1117/12.2533634.

[4] A. Bégué *et al.*, "Agricultural Systems Studies using Remote Sensing," 2019.

[5] A. Bégué *et al.*, "Remote sensing and cropping practices: A review," *Remote Sens.*, vol. 10, no. 1, pp. 1–32, 2018, doi: 10.3390/rs10010099.

[6] C. Gómez, J. C. White, and M. A. Wulder, "Optical remotely sensed time series data for land cover classification: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 55–72, 2016, doi: 10.1016/j.isprsjprs.2016.03.008.

[7] X. X. Zhu *et al.*, "Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, 2017, doi: 10.1109/MGRS.2017.2762307.

[8] D. Lloyd, "A phenological classification of terrestrial vegetation cover using shortwave vegetation index imagery," *Int. J. Remote Sens.*, vol. 11, no. 12, pp. 2269–2279, 1990, doi: 10.1080/01431169008955174.

[9] M. A. Friedl, C. E. Brodley, and A. H. Strahler, "Maximizing land cover classification accuracies produced

by decision trees at continental to global scales," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2 II, pp. 969–977, 1999, doi: 10.1109/36.752215.

[10] V. Simonneaux, B. Duchemin, D. Helson, S. Er-Raki, A. Olioso, and A. G. Chehbouni, "The use of high-resolution image time series for crop classification and evapotranspiration estimate over an irrigated area in central Morocco," *Int. J. Remote Sens.*, vol. 29, no. 1, pp. 95–116, 2008, doi: 10.1080/01431160701250390.

[11] L. Zhong, L. Hu, and H. Zhou, "Deep learning based multi-temporal crop classification," *Remote Sens. Environ.*, vol. 221, no. March 2018, pp. 430–443, 2019, doi: 10.1016/j.rse.2018.11.032.

[12] W. J. Clancey, "The epistemology of a rule-based expert system -a framework for explanation," *Artif. Intell.*, vol. 20, no. 3, pp. 215–251, 1983, doi: 10.1016/0004-3702(83)90008-5.

[13] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[14] D. Marmanis, K. Schindler, J. D. Wegner, S. Galliani, M. Datcu, and U. Stilla, "Classification with an edge: Improving semantic image segmentation with boundary detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 135, pp. 158–172, 2018, doi: 10.1016/j.isprsjprs.2017.11.009.

[15] P. Sidike *et al.*, "dPEN: deep Progressively Expanded Network for mapping heterogeneous agricultural landscape using WorldView-3 satellite imagery," *Remote Sens. Environ.*, vol. 221, no. December 2018, pp. 756–772, 2019, doi: 10.1016/j.rse.2018.11.031.

[16] W. Li and B. Zhou, "Recognizing terrain features on terrestrial surface using a deep learning model -- An example with crater detection."

[17] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep Convolutional Neural Networks for Hyperspectral Image Classification," *J. Sensors*, 2015.

[18] D. Guidici and M. L. Clark, "One-dimensional convolutional neural network land-cover classification of multi-seasonal hyperspectral imagery in the San Francisco Bay Area, California," *Remote Sens.*, vol. 9, no. 6, 2017, doi: 10.3390/rs9060629.

[19] Y. Li, H. Zhang, and Q. Shen, "Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, 2017, doi: 10.3390/rs9010067.

[20] "Landsat Science Products." https://www.usgs.gov/land-resources/nli/landsat/landsat-science-products (accessed Jun. 14, 2019).

[21] N. Flood, "Continuity of reflectance data between landsat-7 ETM+ and landsat-8 OLI, for both top-of-atmosphere and surface reflectance: A study in the australian landscape," *Remote Sens.*, vol. 6, no. 9, pp. 7952–7970, 2014, doi: 10.3390/rs6097952.

[22] Y. Cai *et al.*, "A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach," *Remote Sens. Environ.*, vol. 210, no. April 2017, pp. 35–47, 2018, doi: 10.1016/j.rse.2018.02.045.

[23] N. Zerrouki and D. Bouchaffra, "Pixel-based or object-based: Which approach is more appropriate for remote sensing image classification?," *Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern.*, vol. 2014-Janua, no. January, pp. 864–869, 2014, doi:

10.1109/SMC.2014.6974020.

[24] A. M. Dean and G. M. Smith, "An evaluation of per-parcel land cover mapping using maximum likelihood class probabilities," *Int. J. Remote Sens.*, vol. 24, no. 14, pp. 2905–2920, 2003, doi: 10.1080/01431160210155910.

[25] Z. Zhou, S. Li, and Y. Shao, "Object-oriented crops classification for remote sensing images based on convolutional neural network," *Image Signal Process. Remote Sens. XXIV*, vol. 1078922, no. October 2018, p. 78, 2018, doi: 10.1117/12.2317448.

[26] A. García Pereira, A. Ojo, E. Curry, and L. Porwol, "Data Acquisition and Processing for GeoAI Models to Support Sustainable Agricultural Practices," in *Proceedings of the 53rd Hawaii International Conference on System Sciences 2020 (HICSS 2020)*, 2020, pp. 922–931, doi: http://hdl.handle.net/10125/63854.

[27] B. G. Ellis, W. L. Hargrove, F. J. Pierce, and C. W. Rice, "Crop Rotation and Its Impact on Efficiency of Water and Nitrogen Use," no. 12542, 2013, doi: 10.2134/asaspecpub51.c3.

[28] S. J. Sarandón and Claudia Cecilia Flores, *Agroecología: bases teóricas para el diseño y manejo de Agroecosistemas sustentables*. Universidad Nacional de La Plata – Editorial de la Universidad de La Plata, 2014.

[29] T. Dias, A. Dukes, and P. M. Antunes, "Accounting for soil biotic effects on soil health and crop productivity in the design of crop rotations," *J. Sci. Food Agric.*, vol. 95, no. 3, pp. 447–454, 2014, doi: 10.1002/jsfa.6565.

[30] P. Pinto, M. E. Fernández Long, and G. Piñeiro, "Including cover crops during fallow periods for increasing ecosystem services: Is it possible in croplands of Southern South America?," *Agric. Ecosyst. Environ.*, vol. 248, no. February, pp. 48–57, 2017, doi: 10.1016/j.agee.2017.07.028.

[31] M. Rubwurm and M. Korner, "Temporal Vegetation Modelling Using Long Short-Term Memory Networks for Crop Identification from Medium-Resolution Multi-spectral Satellite Images," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2017-July, pp. 1496–1504, 2017, doi: 10.1109/CVPRW.2017.193.

[32] C. Zhang, B. Recht, S. Bengio, M. Hardt, and O. Vinyals, "Understanding deep learning requires rethinking generalization," *5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc.*, 2019.

[33] Y. Yao, L. Rosasco, and A. Caponnetto, "On Early Stopping In Gradient Descent Boosting," pp. 1–22, 2005.

[34] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*. 2012.

[35] A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh, "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances," *Data Min. Knowl. Discov.*, vol. 31, no. 3, pp. 606–660, 2017, doi: 10.1007/s10618-016-0483-9.

[36] C. Pelletier, G. I. Webb, and F. Petitjean, "Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series," no. March, 2018, doi: 10.3390/rs11050523.

[37] Y. L. Pavlov, "Random forests," *Random For.*, pp. 1–122, 2019, doi: 10.1201/9780367816377-11.

[38] M. Belgiu and L. Drăgu, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31,

2016, doi: 10.1016/j.isprsjprs.2016.01.011.

[39] C. Pelletier, S. Valero, J. Inglada, N. Champion, and G. Dedieu, "Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas," *Remote Sens. Environ.*, vol. 187, pp. 156–168, 2016, doi: 10.1016/j.rse.2016.10.010.

[40] M. Schmidt, M. Pringle, R. Devadas, R. Denham, and D. Tindall, "A framework for large-area mapping of past and present cropping activity using seasonal landsat images and time series metrics," *Remote Sens.*, vol. 8, no. 4, 2016, doi: 10.3390/rs8040312.

[41] S. Tian, X. Zhang, J. Tian, and Q. Sun, "Random forest classification of wetland landcovers from multi-sensor data in the arid region of Xinjiang, China," *Remote Sens.*, vol. 8, no. 11, pp. 1–14, 2016, doi: 10.3390/rs8110954.

[42] J. C. W. Chan and D. Paelinckx, "Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2999–3011, 2008, doi: 10.1016/j.rse.2008.02.011.

[43] C. Szegedy *et al.*, "Going deeper with convolutions," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 1–9, 2015, doi: 10.1109/CVPR.2015.7298594.

[44] S. Lawrence, C. L. Giles, and A. C. Tsoi, "Lessons in Neural Network Training : Overfitting May be Harder than Expected," *Network*, pp. 540–545, 1997.

[45] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 67, no. 1, pp. 93–104, 2012, doi: 10.1016/j.isprsjprs.2011.11.002.

[46] A. Liaw and M. Wiener, "Classification and Regression by randomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[47] C. Zhang, "Combining hyperspectral and lidar data for vegetation mapping in the Florida everglades," *Photogramm. Eng. Remote Sensing*, vol. 80, no. 8, pp. 733–743, 2014, doi: 10.14358/PERS.80.8.733.

[48] M. Weinmann, B. Jutzi, S. Hinz, and C. Mallet, "Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 286–304, 2015, doi: 10.1016/j.isprsjprs.2015.01.016.

[49] J. Xue and B. Su, "Significant remote sensing vegetation indices: A review of developments and applications," *J. Sensors*, vol. 2017, 2017, doi: 10.1155/2017/1353691.

[50] A. Karnieli *et al.*, "Use of NDVI and land surface temperature for drought assessment: Merits and limitations," *J. Clim.*, vol. 23, no. 3, pp. 618–633, 2010, doi: 10.1175/2009JCLI2900.1.

[51] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.*, vol. 8, no. 2, pp. 127–150, 1979.

[52] H. Q. Liu and A. Huete, "Feedback based modification of the NDVI to minimize canopy background and atmospheric noise," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 2, pp. 457–465, 1995, doi: 10.1109/36.377946.

[53] A. R. Huete, K. Didan, and W. Van Leeuwen, "Modis Vegetation Index," *Veg. Index Phenol. Lab*, vol. 3, no. 1, p. 129, 1999.