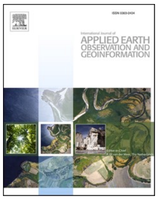
International Journal of Applied Earth Observations and Geoinformation 95 (2021) 102264

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



# Phenology-based sample generation for supervised crop type classification

Mariana Belgiu [a,*], Wietske Bijker [a], Ovidiu Csillik [b], Alfred Stein [a]

[a] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, 7500 AE Enschede, the Netherlands
[b] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University & Research, Wageningen 6708 PB, the Netherlands

ARTICLE INFO

*Keywords:*
Random Forest
Samples cleansing
Training sample
Agricultural mapping
Food security

ABSTRACT

Crop type mapping is relevant to a wide range of food security applications. Supervised classification methods commonly generate these data from satellite image time-series. Yet, their successful implementation is hindered by the lack of training samples. Solutions like transfer learning, development of temporal-spectral signatures of the target classes, re-utilization of existing inventories, or crowdsourcing initiatives are commonly applied to generate samples for thematically coarser classifications. These methods are rarely used for generating crop types samples. In this study, we leverage the phenology information of existing data inventories using Time-Weighted Dynamic Time Warping (TWDTW) to address the problem of automatic crop sample generation in two target areas. Resulting labeled samples are refined using proximity measures obtained from Random Forests (RF). Sentinel-2 time-series are used to obtain phenology information from two study areas. The proposed methodology achieved promising results for classes with a reduced inter-classes similarity such as sugar beets (user's accuracy, UA, of 98% and producer's accuracy, PA, of 100%) or grains (UA of 98% and PA of 90%). The crops with a high inter-classes similarity yielded less satisfactory results. Potatoes, for example, obtained a high PA of 95%, but a UA of only 36% because of the spectral-temporal similarity with maize. The methodology works well for areas with balanced crop samples. Yet, it favors prevalent classes in areas with imbalanced crops at the expense of a low accuracy for the minority crops. Despite these shortcomings, the proposed methodology offers a viable option to generate crop samples in regions with few ground labels.

## 1. Introduction

Agriculture is an important economic sector. For instance, in the European Union (EU) alone, in 2018, 44 million jobs are related to food industries and agriculture, while the EU supported farmers with 58.82 billion euros, of which 44.74 billion euros in income support (European-Union, 2018). Crop mapping and monitoring is required by many EU Member States to verify agricultural subsidies. As a result, EU databases are regularly updated with information on crops grown on agricultural land. In the global South on the other hand, there is a lack of incentives and such databases are non-existent or not regularly updated. Therefore, efficient methods for regularly generating reliable crop type information are required (Weiss et al., 2020).

A broad variety of spaceborne sensors are now providing access to dense time-series data and high spatial and spectral resolution imagery. These data are vital for crop mapping and monitoring (Bégué et al., 2018). Recent scientific and methodological developments enabled crop mapping at local (Belgiu and Csillik, 2018; Csillik et al., 2019; Li and Bijker, 2019), regional (Mohammed et al., 2020; Simoes et al., 2020), or continental scales (Xiong et al., 2017). Important methodological developments in this domain have been based on supervised machine learning methods (Weiss et al., 2020). However, these methods rely on training samples and collecting them through field campaigns is time-consuming and expensive (Maxwell et al., 2018). Therefore, we may benefit from alternative solutions to generate training samples.

Previous research has proposed various ways to generate training samples, including transfer learning methods (Tuia et al., 2009; Tuia et al., 2011), crowdsourcing initiatives (Fritz et al., 2009), making use of the spectral and temporal signatures of the target classes (Malambo and Heatwole, 2020), or leveraging existing inventories to guide the labeling of the new training samples (Huang et al., 2020). Important work was dedicated to generating training samples automatically by inspecting the spectral and temporal signatures of the classes of interest. Malambo and Heatwole (2020), for example, used spectral-temporal trajectories extracted from multi-temporal Landsat images to create training data for burned areas. Globally applicable spectral signatures have been

* Corresponding author.
*E-mail addresses:* m.belgiu@utwente.nl (M. Belgiu), w.bijker@utwente.nl (W. Bijker), ovidiu.csillik@wur.nl (O. Csillik), a.stein@utwente.nl (A. Stein).

https://doi.org/10.1016/j.jag.2020.102264
Received 26 July 2020; Received in revised form 21 October 2020; Accepted 22 October 2020
Available online 10 November 2020

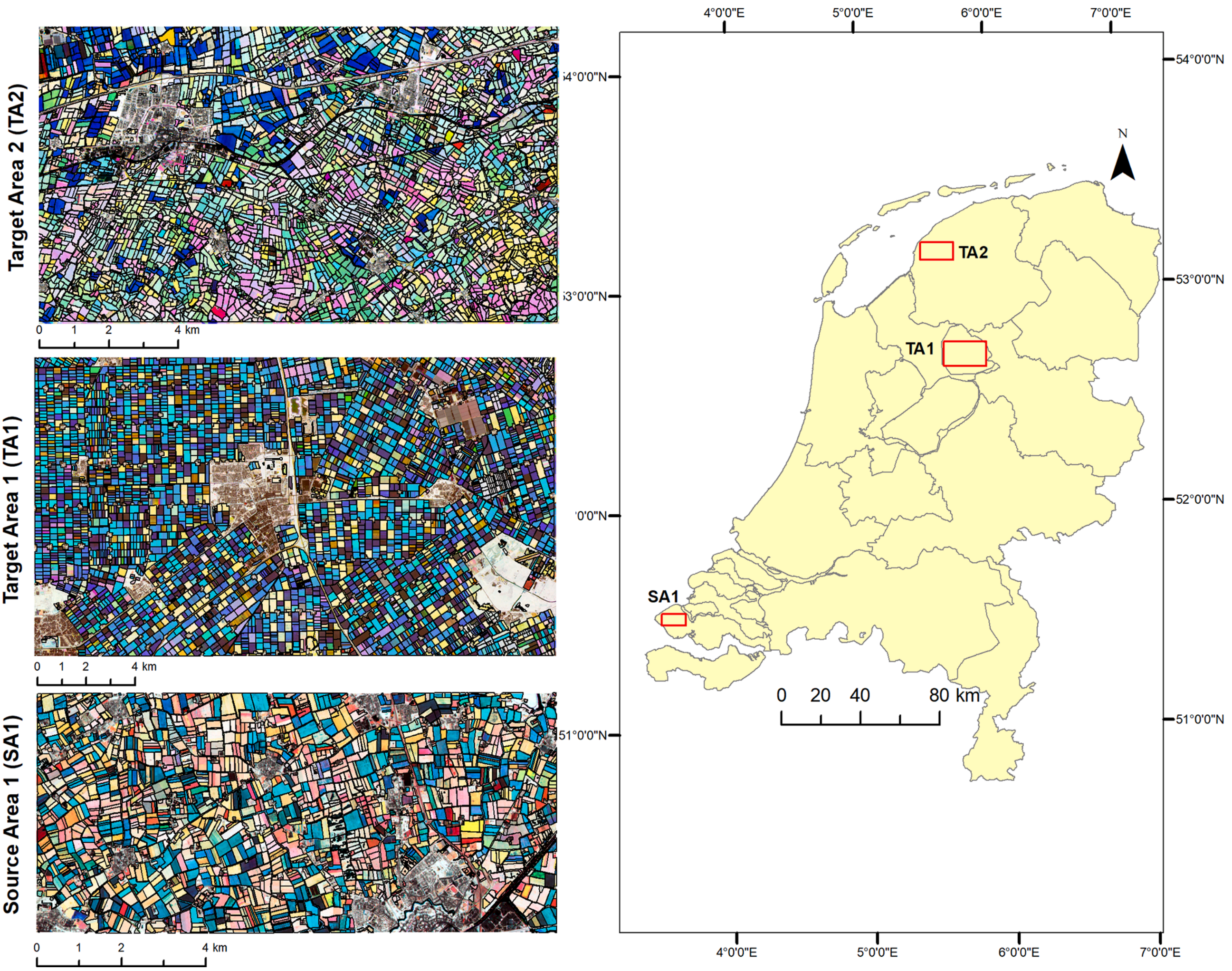


**Fig. 1.** Study areas located in the Netherlands. Source Area 1 (SA1) is located in Zeeland; Target Area 1 (TA1) is located in Flevoland; Target Area 2 (TA2) is located in Friesland. Normalized Difference Vegetation Index (NDVI) layers for May (red channel), June (green channel), and July 2018 (blue channel) are used for the visualization of the study areas, whereas crop fields were superimposed over the NDVI layers. Crop field data were retrieved from the Base Registration Crop Parcels agency in the Netherlands (www.PDOK.nl).

investigated for forest changes monitoring over large areas (Woodcock et al., 2001), forest mapping (Pax-Lenney et al., 2001), tropical forest biomass estimation (Foody et al., 2003), urban areas mapping (Okujeni et al., 2018), sugarcane mapping using generalized spectral libraries generated from multi-annual data (Luciano et al., 2018), and land cover mapping on a regional scale using spectral libraries generated from 8-day MODIS data (Zhang et al., 2018).

An increasing number of studies recommended the utilization of the already available samples for training supervised classifiers (Huang et al., 2020; Radoux et al., 2014). However, the available ground labels are often noisy. In addition, large variability in the phenology of crops caused by varying agricultural practices and different weather conditions prevents the reusability of these labels from one geographic region to another and from one year to another (Belgiu et al., 2020; Wang et al.,

**Table 1**
Weather data over 2018: monthly average temperature (Tav), monthly average of daily maximum temperature (Tmax), monthly average of daily minimum temperature (Tmin), and total monthly precipitation for SA1, T1, and T2. Temperatures are in °C, precipitation in mm, Precipitation-P, based on data from the Royal Meteorological Institute (KNMI) for Vlissingen (SA1), Marknesse (TA1), and Leeuwarden (TA2).

| | Jan. | Feb. | Mar. | Apr. | May | Jun. | Jul. | Aug. | Sept. | Oct. | Nov | Dec. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tav SA1 | 6.20 | 4.60 | 4.48 | 11.29 | 15.38 | 16.77 | 21.00 | 19.25 | 16.29 | 13.53 | 8.23 | 7.06 |
| Tav TA1 | 4.93 | 0.39 | 4.07 | 11.82 | 16.68 | 16.70 | 19.76 | 18.15 | 14.58 | 11.93 | 6.34 | 5.92 |
| Tav TA2 | 4.62 | 0.22 | 3.55 | 10.86 | 15.63 | 15.93 | 19.09 | 18.00 | 14.65 | 12.17 | 6.43 | 6.06 |
| Tmax SA1 | 7.91 | 4.53 | 7.39 | 15.20 | 19.76 | 20.41 | 25.65 | 22.65 | 19.28 | 16.46 | 10.1 | 8.72 |
| Tmax TA1 | 7.13 | 3.63 | 8.12 | 16.96 | 22.77 | 21.24 | 26.51 | 23.43 | 19.25 | 16.54 | 9.21 | 7.85 |
| Tmax TA2 | 6.72 | 3.47 | 7.39 | 15.36 | 21.10 | 19.76 | 24.43 | 22.58 | 18.73 | 16.01 | 9.08 | 8.00 |
| Tmin SA1 | 4.39 | 0.21 | 2.14 | 8.25 | 11.64 | 13.76 | 17.29 | 16.41 | 13.44 | 10.96 | 6.32 | 5.34 |
| Tmin TA1 | 2.69 | −2.04 | 0.39 | 6.63 | 10.47 | 12.04 | 12.19 | 12.56 | 9.72 | 7.26 | 3.26 | 3.83 |
| Tmin TA2 | 2.29 | −2.90 | −0.28 | 6.37 | 9.71 | 12.01 | 13.09 | 13.38 | 10.50 | 7.98 | 3.44 | 3.81 |
| P SA1 | 69.70 | 35.70 | 35.70 | 86.10 | 82.30 | 31.60 | 0.70 | 13.90 | 82.20 | 57.80 | 67.7 | 20.2 |
| P TA1 | 83.00 | 16.70 | 55.60 | 61.40 | 52.60 | 20.30 | 3.50 | 100.90 | 50.60 | 43.00 | 28.9 | 101 |
| P TA2 | 103.30 | 19.00 | 42.80 | 73.50 | 34.20 | 19.60 | 11.00 | 88.40 | 37.00 | 38.40 | 26.3 | 111 |

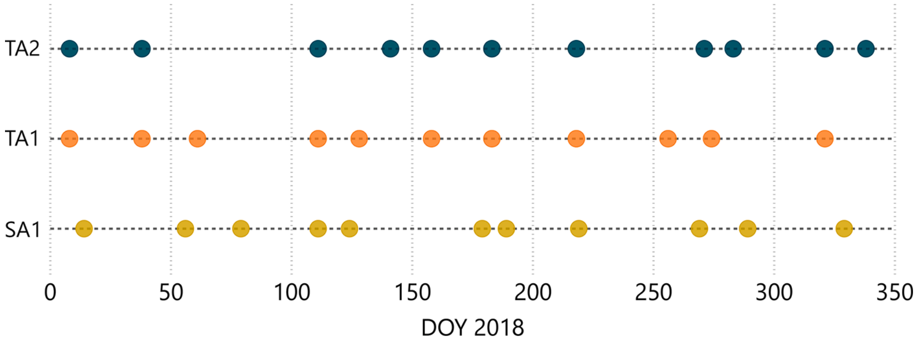


**Fig. 2.** Acquisition dates for Sentinel-2 images used in this study for SA1, TA1, and TA2.

2019). As the performance of supervised classifiers is influenced by the quality of the samples (Frénay and Verleysen, 2013), noisy labels may lead to the decrease of the classifier accuracy while increasing the complexity and training time (Miranda et al., 2009; Zhu and Wu, 2004). Different strategies have been proposed to improve the quality of the samples, including reclassification, or removal of samples, and their combination. Simoes et al. (2020), for example, used self-organizing maps (SOM) to assess the quality of the existing land cover-land use training samples by iteratively computing the probability of the available samples to belong to the created clusters.

Most of the above-mentioned methods have been applied either to one-class classification problems (Malambo and Heatwole, 2020) or to thematically coarser classifications such as land cover classification (Simoes et al., 2020). The application of these methods to generating crop type labels is missing.

Leveraging existing samples is an attractive solution to address the problem of the availability of crop labels for classification (Fowler et al., 2020). Therefore, in this study, we will leverage the phenology information of samples from one study area (referred to as source area) using Time-Weighted Dynamic Time Warping (TWDTW) (Maus et al., 2016) to generate labels in two target areas using Sentinel-2 time-series. Proximity measures between samples, calculated with Random Forests (RF) (Breiman, 2001), are used to further refine the automated labeled samples. Breiman (2001) mentioned that the RF proximity measure is 'one of the most useful tools in RF'. Corcoran et al. (2013) used RF-derived proximities to find outliers in the labeled wetland data by identifying samples with small proximities relative to their classes. Yet, these proximity measures have not been used for further sample refinement purposes. The scientific contributions of this paper are summarized as follows: (i) we evaluate the potential of phenology information computed from Sentinel-2 time-series and TWDTW in a source area to generate samples for two different target areas; (ii) we apply RF-derived proximity measures combined with *k*-means clustering for refining of samples.

## 2. Study area and datasets

All three study areas are situated in the Netherlands, in major agriculture areas. Source Area (SA1) is located in the south-western part of the country in the province of Zeeland. Target Area 1 and 2 (TA1 and TA2) are situated in the northern part of the country, in Flevoland (Noordoostpolder) and Friesland provinces, respectively (Fig. 1). The distance between the SA1 and TA2, the farthest test area, is about 240 km. All three areas are situated on young marine clay soils and located close to the sea, which gives them mild winters and cool summers.

The major crops cultivated in SA1 are beans, cauliflower, grassland, maize, onions, carrots, potatoes for consumption, seed potatoes, sugar beets, and grains. Similar to SA1, grains, grassland, maize, onions, potatoes for consumption and potatoes for seeds, and sugar beets crops are also present in TA1. Apples and pears plantations are missing from SA1, and minor crops (i.e. referred to as other classes in this study) are mainly represented by chicory, winter carrots, lucerne, spinach, and flower bulbs. In TA2, grassland is the main cultivated crop in this study area. The remaining agricultural fields were cultivated with potatoes for consumption and for seeds, maize, sugar beets, and grains (summer wheat and summer barley). Minor crops represent a combination of winter carrots, onions, and flower bulbs. TA2 was selected to assess the impact of the class imbalance problem in labeling and refining training samples.

According to the data of the Royal Meteorological Institute (KNMI), 2018 was an exceptionally dry and warm year. In general, monthly average temperatures were slightly higher in SA1, mainly due to higher minimum temperatures. SA1 also received more rain than TA1 and TA2, especially during April, May, and June, while TA1 and TA2 received more rain during August when many crops were already ripening. The average humidity over the year in all three study areas was higher than 80%. The crop growth and crop management were similar in all three areas, but differences in temperature and precipitation affected the phenological development of the crops. We intentionally selected areas

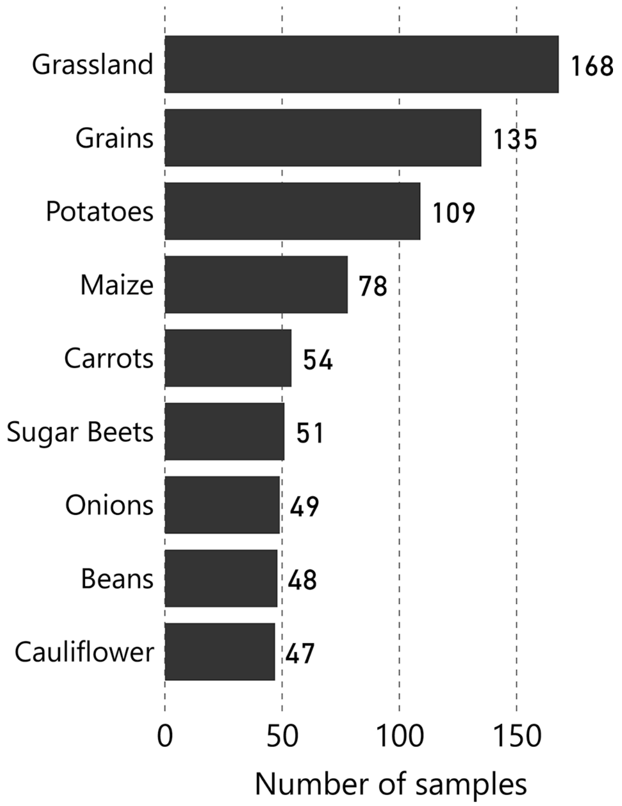


**Fig. 3.** The number of samples in SA1 used to generate the temporal growth of the target crops.

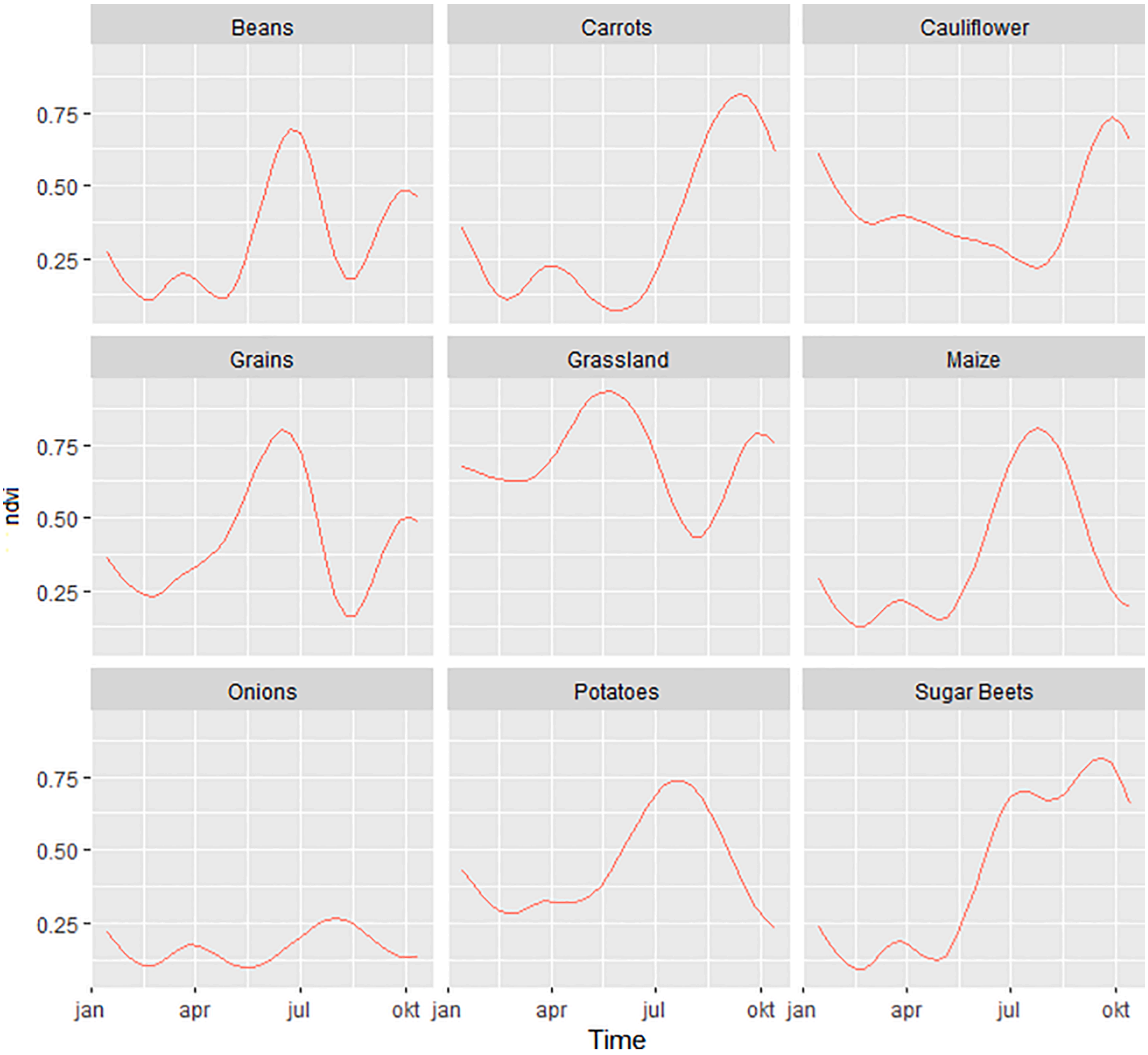


**Fig. 4.** The number of samples per class generated randomly in TA1.

at some distance and with differences in temperature and precipitation to test the robustness of our method to variations in phenological development between different areas (Table 1).

For this study, we downloaded Sentinel-2 Level-2A data representing the Bottom-Of-Atmosphere (BOA) reflectance images. The images were acquired on different dates (Fig. 2).

All three time series consist of images acquired at different dates throughout 2018 (Fig. 2). These differences are not expected to impact the results because previous studies emphasized the capability of TWDTW to perform well when temporal sequences of the target crops are misaligned on the time axis (Maus et al., 2016; Petitjean et al., 2012). We calculated the Normalized Difference Vegetation Index (NDVI) using band 4 and band 8 of Sentinel-2 images using Google Earth Engine (GEE) (Gorelick et al., 2017). Nationwide crop datasets are available for the Netherlands. The Agricultural Area Netherlands (AAN) supplies the parcel boundaries (polygons) of all land used for agriculture, including annual crops, grassland, and perennial crops. Each year, the Base Registration Crop Parcels (BRP) gives the (main) crop cultivated for all parcels in AAN. The farmers supply the information on the crops they cultivate. AAN and BRP are provided by the Ministry of Economic Affairs and Climate via the Public Data on the Map initiative (PDOK) (www.PDOK.nl).

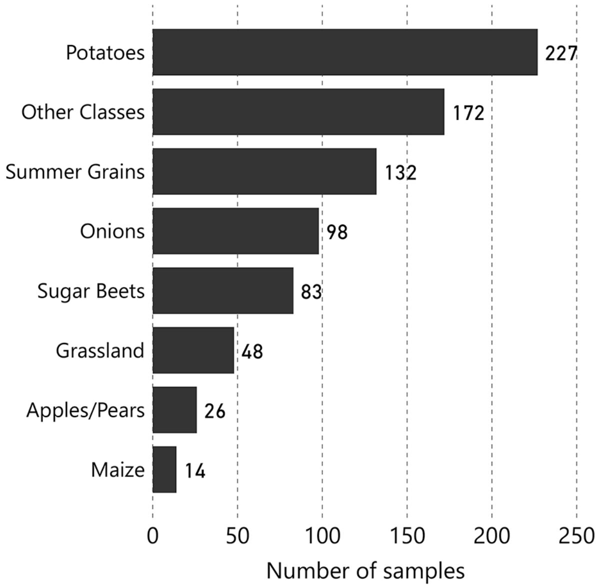


**Fig. 5.** The number of samples per class generated randomly in the TA2.

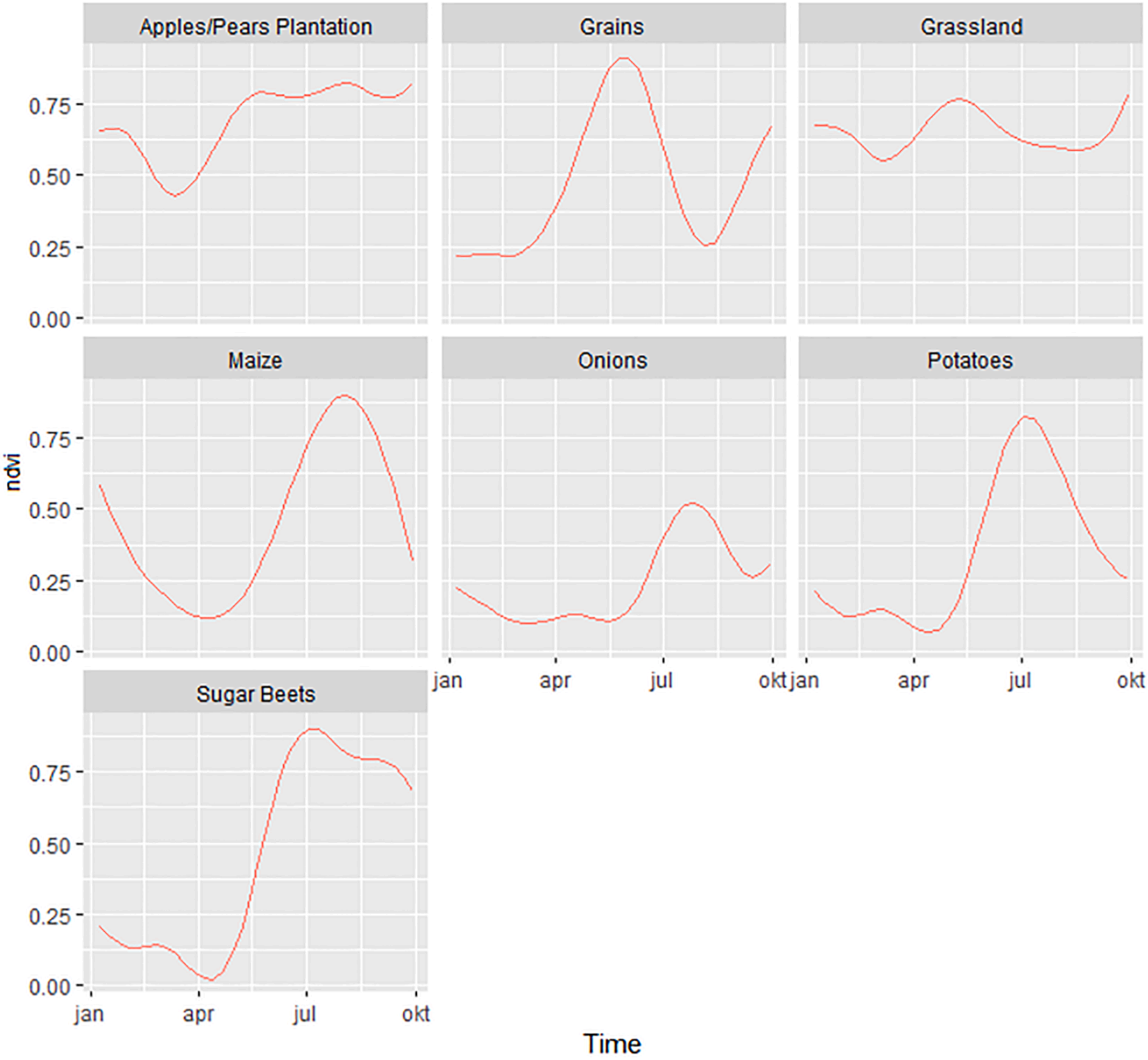


**Fig. 6.** The temporal growth pattern of the crops from SA1 calculated from NDVI computed from eleven Sentinel-2 images.

## 3. Methodology

The methodology proposed in this work comprises the following main steps: (1) generation of the temporal growth pattern of the most representative crops present in the SA1, (2) classification of samples in two target areas (TA1 and TA2) using TWDTW (Maus et al., 2016), and (3) refining classified samples using RF-based proximity measures. The prerequisite for applying this method is to have the same crop types in the source and target areas.

### *3.1. Generation of samples in the source area and the two target areas*

We used the parcel information provided by AAN and BRP to generate 80 samples (pixels) per class using a stratified sampling method. In the next step, we discarded those samples that intersected with a buffer of 20 m around the boundary of the agricultural fields. In this way, we avoided the incorporation of mixed pixels located along the boundaries in the sample set. We merged the potatoes for consumption and the seed potatoes classes into one potato class, as the crop on the field does not show whether it is grown for consumption or seeds and shows a similar temporal profile. Also, the three grassland classes (permanent, temporary, and seeds) were merged into one grassland class. Temporary grassland is grown in alternation with a crop such as maize, but as we focus on a single growing season, this difference is not important and patterns are defined by mowing and grazing. The distribution of the remaining samples in the source area is presented per class in Fig. 3.

In the next step, we generated for each target area 800 randomly distributed samples (Figs. 4 and 5). Note that the grains class present in all three areas consists of summer barley and summer wheat. These two crops were merged because of their spectral-temporal similarity.

### *3.2. Classification of unlabeled samples in target areas*

TWDTW was used to classify the crop samples in TA1 and TA2. Dynamic Time Warping (DTW) is a nonlinear warping algorithm that compares the similarity between two temporal patterns and finds their optimal alignment (Sakoe and Chiba, 1978). It is a time-flexible method ideal to compare two temporal growth patterns of crops because it considers possible temporal distortions of the time series, like amplitude difference, time shifting, shape changes, or noise in the data. The alignment between two sequences is done recursively by comparing each element of a sequence with all other elements of the other sequence, thus obtaining a DTW matrix. The last element of the matrix represents the degree of dissimilarity between the two compared sequences, with values closer to 0 representing very good matching of the

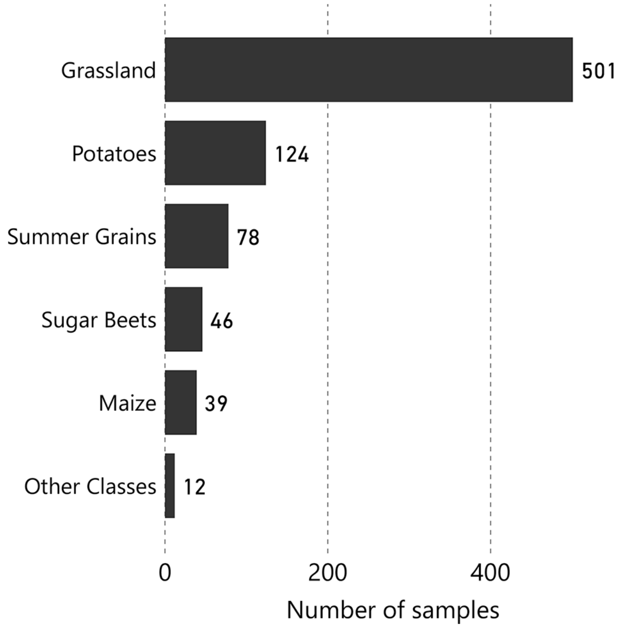


**Fig. 7.** The temporal growth pattern of the crops in TA1 calculated from NDVI computed from eleven Sentinel-2 images.

sequences compared. However, in the case of crop mapping, computing the entire DTW matrix can be computationally intensive and can lead to misleading results, like matching a winter crop with a summer crop. This extreme warping can be avoided by adding a temporal cost when comparing two elements of the sequences, either using a linear or a logistic model (Maus et al., 2016). Therefore, we used the time-weighted DTW (TWDTW) to classify the crop samples in TA1 and TA2 based on the temporal patterns of crops from SA1. We used a logistic TWDTW that was shown to outperform a linear TWDTW (Maus et al., 2016) with a low penalty for time warps smaller than 50 days and a higher penalty otherwise.

### 3.3. *Refining automatically labeled samples using Dynamic time Warping (DTW) similarity values and Random Forest-based proximity measures*

TWDTW-based classified samples were further refined by removing the samples whose similarity with the SA1 samples are one standard deviation above the mean. In this way, we discarded the samples that are less similar to the temporal pattern of the crop class they were assigned to.

RF is a popular classifier used in different remote sensing applications because of its robustness to noise in the training sample set, a reduced number of parameters that require user's input, multi-class handling capacity, and availability of built-in capability to assess the importance of the input variables (Belgiu and Drăguţ, 2016). RF is an ensemble of decision trees created through bootstrapping. The classifier is sensitive to two parameters: *Ntree* parameter defining the number of decision trees to be built and the *Mtry* parameter that refers to the number of variables used to split the tree nodes. Based on previous studies (Belgiu and Drăguţ, 2016), *Ntree* was set to 1000 and *Mtry* was defined as the square root of the total number of variables.

Besides the above-mentioned characteristics, RF allows the calculation of proximity values, i.e. similarity, between samples. Proximities between all samples (including the out-of-bag samples) generated an $N \times N$ matrix, denoting the similarity between each pair of samples. The $(n,m)$ element of the proximity matrix was calculated by the fraction of the number of trees in which the elements $n$ and $m$ ended up in the same terminal node of the decision trees created in the RF model. For example, if $Ntree = 1000$ and a pair of samples ends up in the same terminal node in 300 of the 1000 trees, then the proximity value is 0.3. Therefore, similar samples should be in the same terminal node more often than dissimilar samples (Liaw and Wiener, 2002). The inverted proximity matrix (1-proximity) denoting the distance between samples was mapped into a Cartesian space using multidimensionality scaling (MDS) (Cox and Cox, 2008). MDS is a common method to visualize the proximities between samples in two dimensions (Buja et al., 2008). The proximity of samples is proportional to their similarity.

In the next step, we used *k*-means to cluster the samples using the proximity measures. *K*-means is an unsupervised clustering method that generates *k* number of clusters by iteratively measuring the Euclidean distance between input samples and assigning them to the nearest cluster centroid. It has been commonly used for land cover classification (Gómez et al., 2016) and, more recently, for crop type mapping (Wang et al., 2019). The number of clusters (*k*) was determined using the gap statistic method proposed by Tibshirani et al. (2001). Gap statistic is a data-driven method that compares the within-cluster dispersion with a null reference distribution of the data, i.e. there is only one cluster (Eqs. (1) and (2)).

$$Gap(k) = \frac{1}{B} \times \sum_{B} \log(W_b^*(k)) - \log(W(k)) \quad (1)$$

where $B$ is the number of reference datasets, $W(k)$ is the within-cluster sum of squares (with k cluster) and $W_b^*(k)$ is the within-cluster sum of squares of the reference datasets.

The number of optimal clusters is chosen when:

$$Gap(k) \geq Gap(k+1) - s_{k+1} \quad (2)$$

where $s_{k+1}$ is the estimate of the standard deviation of $\log(W_b^*(k+1))$.

The labeled samples representing the same crop might be clustered into more than one cluster. In this case, the cluster is selected to which the majority of samples are assigned. The remaining samples were distributed over the classes with the majority of missing samples assigned to the other clusters. For example, samples of class A might be clustered into three clusters: two samples as cluster 4, 46 samples as cluster 7, and three samples as cluster 10. Cluster 7, with the majority samples, is retained, whereas the other samples were allocated to the sample sets assigned to clusters 4 and 10, respectively.

On the other hand, samples representing different crops might be assigned to the same cluster. To address this misclassification, we re-ran the RF classification to obtain the proximity measures for the crops assigned to the same cluster as input and applied *k*-means clustering using the resulting proximity values. By using this iterative method, we refined the samples until the majority of samples of one crop class were assigned to a unique cluster, i.e. no other class had its majority samples assigned to this cluster.

### 3.4. *Assessement of the quality of the generated reference data*

The quality of the labeled reference was assessed using the following metrics: overall accuracy (OA), user's accuracy (UA), and producer's accuracy (PA) (Congalton, 1991).

## 4. Results

### 4.1. *Temporal profiles of the investigated crops in all three study areas*

Grains had a similar temporal profile in all three areas, reaching the highest peak in June-July (Figs. 6–8). The phenological pattern of grassland differed across the three study areas since the growth cycle of this class is highly influenced by mowing, grazing, or harvesting of the seeds.

Maize and potatoes crops reached the highest NDVI value in July-August in all study areas. Onions differed between SA1 and TA1 in

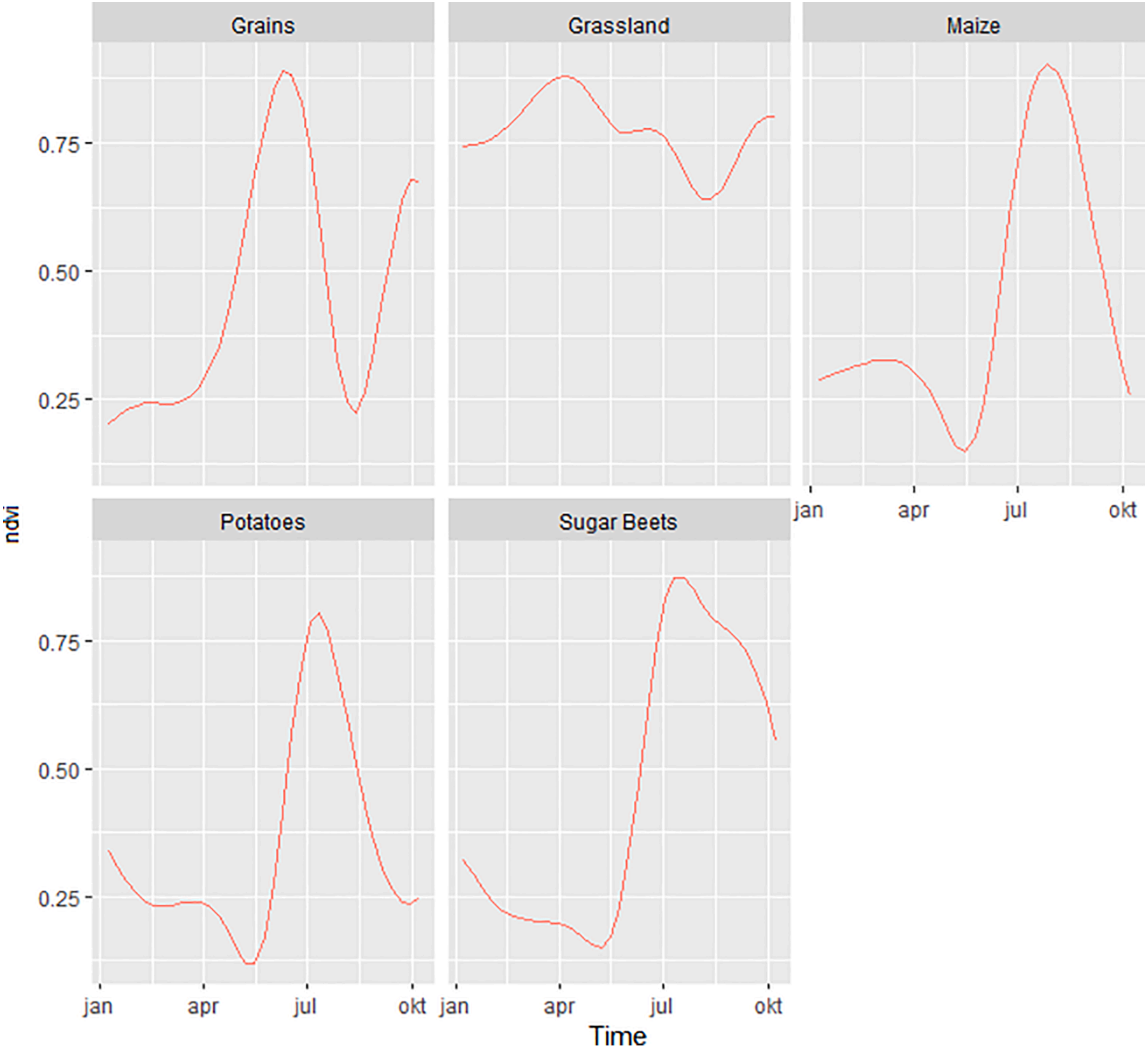


**Fig. 8.** The temporal growth pattern of the crops in TA2 using the NDVI from eleven Sentinel-2 images.

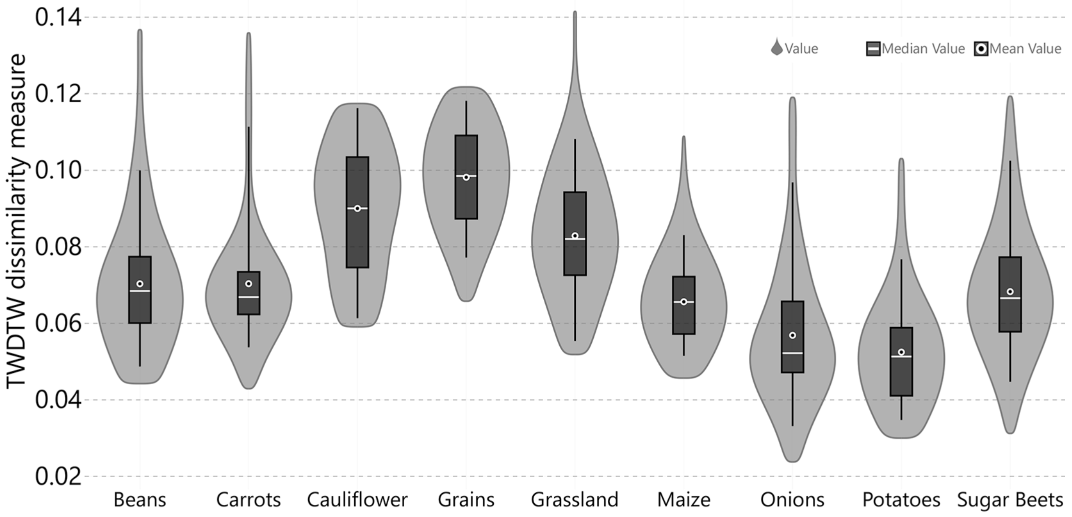


**Fig. 9.** Boxplots based on TWDTW distances between labeled samples from SA1 and unlabeled samples from TA1.

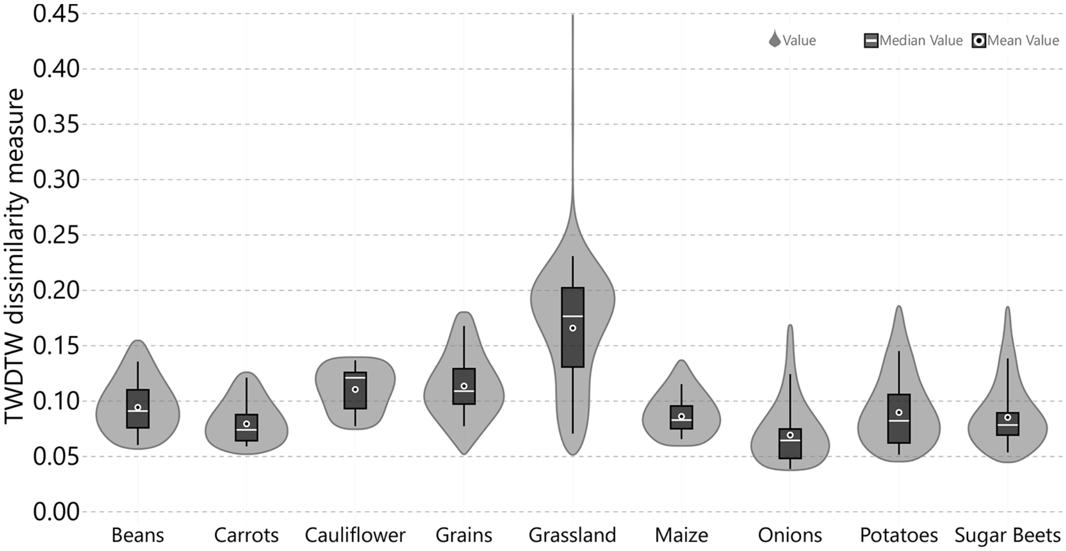


**Fig. 10.** Boxplots based on TWDTW distances between labeled samples from SA1 and unlabeled samples from TA2.

terms of the values of their NDVI-based peaks. Sugarbeets showed two peaks, in July and October, with a slight valley between them. While for SA1 the peak in October was higher than in July, for TA1 and TA2 the peak in July was the highest. The green-up stage of the Apples/Pears Plantation class showed similarity with the temporal profile of grassland.

### *4.2. Spectral-temporal distances between crop types*

The distributions of spectral-temporal distances between labeled samples from SA1 and unlabeled samples from TA1 and TA2 calculated using TWDTW are presented in Figs. 9 and 10. We discarded the outliers with one standard deviation above the mean. This increased the overall accuracy of the classification, as compared to the scenario in which no samples were removed at this stage (Table A.1.1 in Annex 1).

The samples were assigned to different crop types based on the lowest dissimilarity values. The boxplots showed that onions, potatoes, and sugar beets from TA1 and TA2 presented a higher similarity with the same crops in SA1. On the other hand, cauliflower and grains had the highest dissimilarity values in both areas. Cauliflower was one of the crops present in SA1 that was missing in TA1 and TA2. In TA2, the grassland class presented a large dispersion of the dissimilarity values. This might happen because this class consisted of different types of grassland.

### *4.3. Samples classification results obtained by applying the Random Forest-based refinement method*

Although we had nine crops in SA1, the gap statistic method estimated ten clusters for both samples set from TA1 and TA2. Clustering was performed using the distances computed between all training samples that ended up in the same terminal node (i.e. 1-proximity values). Visualizing the *k*-means clusters using the MDS plots revealed a clear overlap between resulting clusters in both target areas (Fig. 11). The samples that were not assigned to any clusters were discarded from the samples set. Note that the MDS plot depicted the arrangement of samples from TA1 and TA2 in two dimensions in which the proximity of samples to one another is proportional to how similar the samples were to each other. It is important to mention that the axes themselves have no real meaning.

The majority of samples of potatoes, maize, onions, and other classes in TA1 were assigned to the same cluster. This happened because of the similarity between their temporal profiles and, hence, the similarity of proximity values that ended up in the same cluster. Therefore, these samples were iteratively re-clustered based on the RF proximity measures until the majority of samples of each class belonged to a cluster different than the clusters other classes were assigned to. In the case of TA2, six classes were assigned to the same cluster because of their similarity in the spectral-temporal domain as identified by TWDTW: grains, potatoes, carrots, maize, cauliflower, and beans.

Beans, cauliflower, and carrots were present in the SA1 but were missing from TA2. All samples assigned to these classes were considered as ‘other classes’ in the confusion matrix. As already reported in the literature, the combined class ‘other classes’ obtained a lower accuracy because of the high intra-class heterogeneity. The apples/pears plantation (orchards) were present in TA1 but were absent from the samples set of SA1. Therefore, these samples were also included in the confusion matrix as ‘other classes’.

We obtained an OA of 69% in TA1. The best-classified classes were sugar beets, potatoes, and grains, which obtained a PA of 100%, 95.12%, and 90.38%, respectively. The UA of potatoes was low (35.77%) because of the high overlap with the maize (24 potatoes samples misclassified as maize) and onions (30 samples misclassified as onions) (Fig. 12 and Table A.1.3).

Onions obtained a rather low PA (57.17%) because of the confusion with the potatoes class. This happened because onions in SA1 looked more like fallow/idle crop, whereas onions in TA1 had a more prominent peak in the temporal profile, being similar to potatoes from SA1. The UA of onions was 82.98%. Grassland achieved a UA of 85.71% and a low PA (62.07%) because of the confusion with the other class. The lowest PA was obtained by maize (7.41%) because of the confusion with potatoes.

The proposed methodology yielded an OA of 75% for TA2. Similar to TA1, we grouped the samples classified as onions, carrots, beans, and cauliflower as other classes. These crops were present in the source area but were missing in this target area. Grassland yielded the best classification results with a PA of 98.46% and UA of 92.75% (Fig. 13 and Table A.1.5). There was a large confusion between maize and potatoes classes, which led to low PA and UA values for these two classes.

Maize, for example, yielded a PA of only 24.16% because of the confusion with potatoes. Both potatoes and other classes were misclassified as maize and, therefore, the UA of maize was only 43.75%. Potatoes had a UA of only 26.06% and a PA of 56.24%. Sugar beets were misclassified as grassland and obtained a PA of 64.51%. The UA of this class was relatively low (76.92%) because a relatively large number of maize, other classes, and potatoes samples were misclassified as sugar beets. The grains class obtained a low PA (65.51%) and UA (50%) because of the high confusion with potatoes, grassland, and other classes. The classification results obtained using TWDTW without RF-based refinement are presented in Annex 1 (Tables A.1.2 and A.1.4).

## 5. Discussion

In this study, we evaluated the effectiveness of TWDTW and RF-derived proximity measures to label and refine crop samples in two target areas situated in the Netherlands. To perform the labeling, we used reference data available from a similar area, referred to as the source area in our study. The proposed methodology is purely data-driven and can be adapted to different areas including developing countries that are confronting with serious food security challenges and, therefore, need access to up-to-date and reliable information on crop types for implementing efficient and sustainable intervention programs.

By measuring the spectral-temporal similarity between phenology of labeled samples from source areas and those unlabeled from target areas, we achieved promising results for classes with a reduced inter-classes similarity, such as sugar beets or grains in TA1 and grassland in TA2. The crops with a high inter-classes similarity yielded less satisfactory results. Potatoes, for example, obtained a PA of 95% in TA1, but a UA of only 36% because of the spectral-temporal similarity with maize. The classification results of potatoes and maize might be improved by using textural features that proved to be an efficient input variable for crop classification (Kwak and Park, 2019).

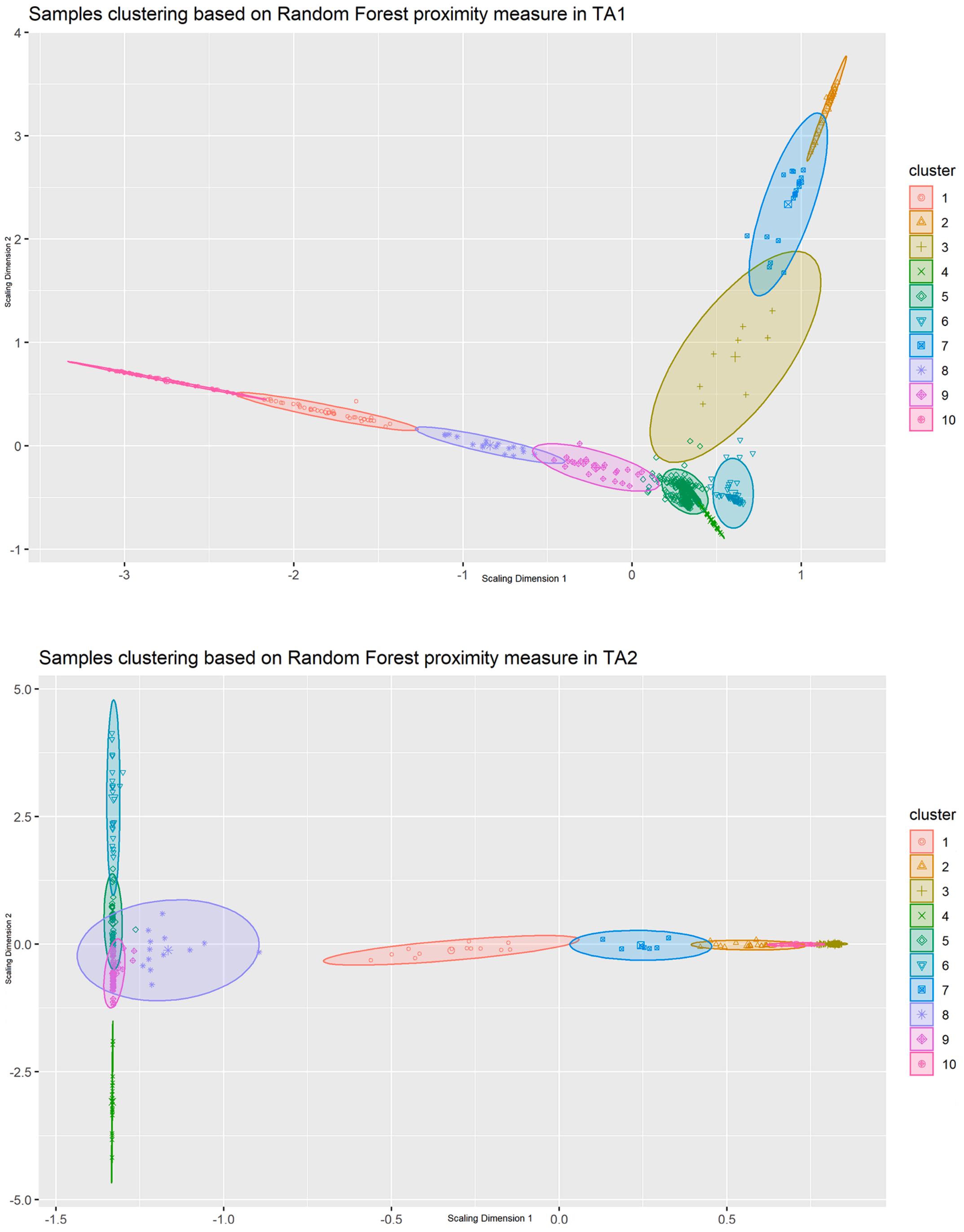


**Fig. 11.** Samples clustering in TA1 and TA2. The clustering was performed using proximity measures as input and by applying *k*-means with 10 predefined clusters.

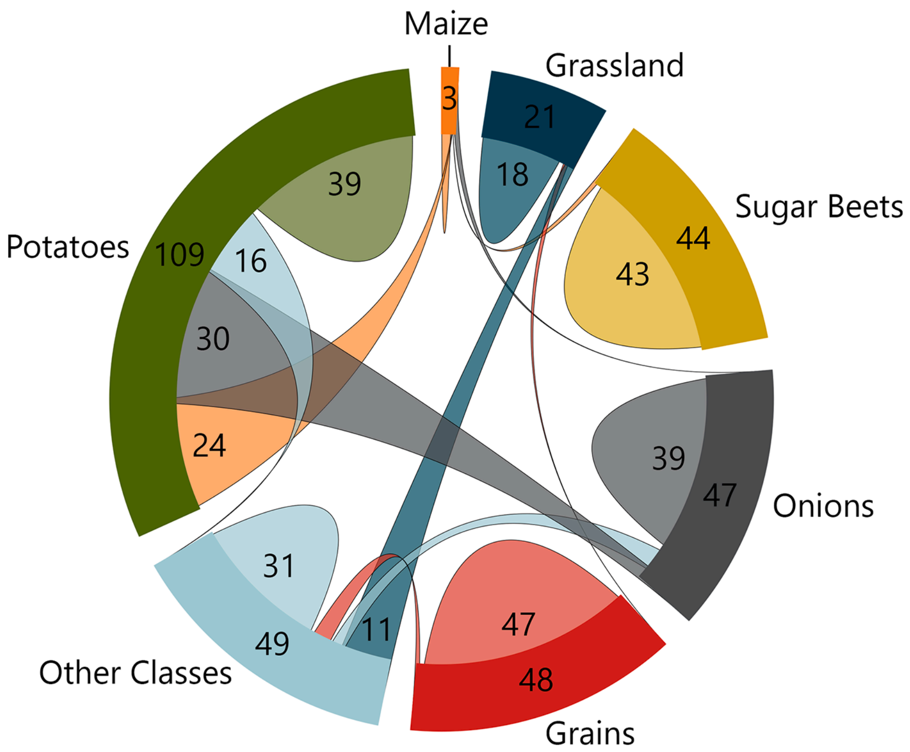


**Fig. 12.** Assessment of the sample classification and refining results in TA1. The correlations between classes read like this: e.g. 30 samples classified as potatoes were actually onions, while 3 samples classified as onions should have been classified as potatoes. 39 samples of potatoes were correctly identified.”.

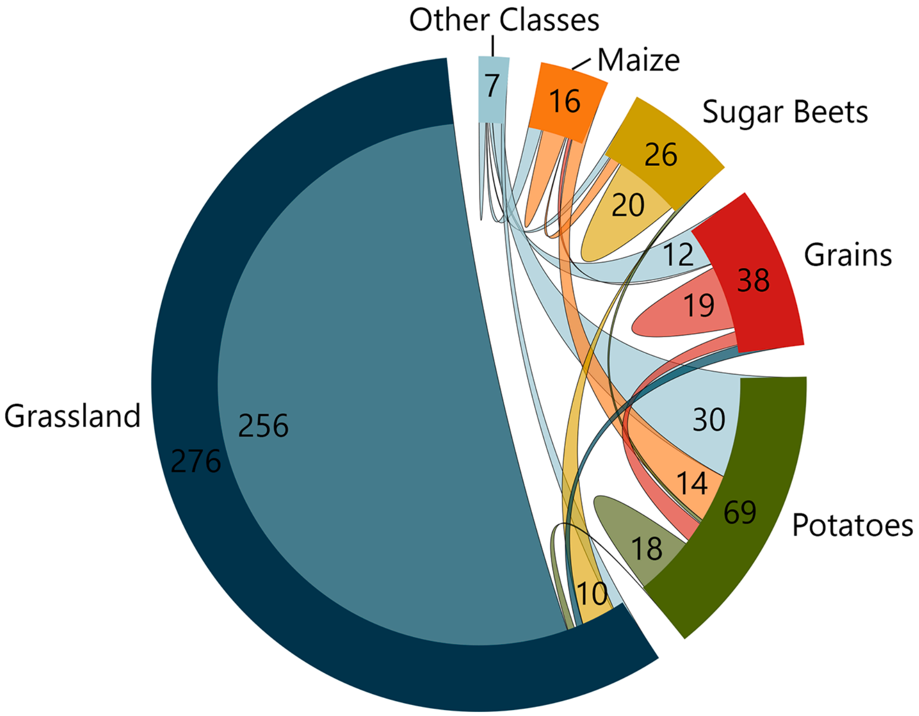


**Fig. 13.** Assessment of the sample classification and refining results in TA2.

**Table A.1.1**
Assessment of the sample classification obtained in TA1 without removing the samples those spectral-temporal similarity with SA1 labels as computed by means of TWDTW is one standard deviation above the mean.

| | Potatoes | Sugar Beets | Grassland | Maize | Grains | Onions | OC | Total | UA (%) |
|---|---|---|---|---|---|---|---|---|---|
| Potatoes | 37 | 9 | 0 | 51 | 1 | 45 | 16 | 159 | 23.27 |
| Sugar Beets | 0 | 53 | 0 | 0 | 0 | 0 | 0 | 53 | 100 |
| Grassland | 0 | 0 | 31 | 0 | 0 | 0 | 0 | 31 | 100 |
| Maize | 0 | 2 | 0 | 4 | 0 | 0 | 0 | 6 | 66.66 |
| Grains | 0 | 7 | 0 | 0 | 84 | 2 | 9 | 102 | 82.35 |
| Onions | 23 | 0 | 0 | 0 | 0 | 58 | 4 | 85 | 68.23 |
| OC | 2 | 4 | 24 | 1 | 2 | 10 | 75 | 118 | 63.55 |
| Total | 62 | 75 | 55 | 56 | 87 | 115 | 104 | 554 | |
| PA (%) | 59.67 | 70.66 | 56.36 | 7.14 | 96.55 | 50.43 | 72.1 | | |
| OA (%) | | | | | | | | | 61.73 |

**Table A.1.2**
Assessment of the sample classification results obtained by applying TWDTW in TA1. UA – User's Accuracy, PA – Producer's Accuracy, OA – Overall Accuracy, OC – Other Classes.

| | Potatoes | Sugar Beets | Grassland | Maize | Grains | Onions | OC | Total | UA (%) |
|---|---|---|---|---|---|---|---|---|---|
| Potatoes | 50 | 0 | 2 | 1 | 1 | 25 | 9 | 88 | 56.81 |
| Sugar Beets | 20 | 80 | 4 | 7 | 17 | 1 | 33 | 162 | 49.38 |
| Grassland 0 | | 0 | 39 | 0 | 11 | 0 | 28 | 78 | 50 |
| Maize | 82 | 3 | 1 | 5 | 0 | 0 | 2 | 93 | 5.34 |
| Grains | 1 | 0 | 0 | 0 | 90 | 0 | 2 | 93 | 96.77 |
| Onions | 50 | 0 | 0 | 0 | 2 | 62 | 17 | 131 | 47.32 |
| OC | 24 | 0 | 2 | 1 | 11 | 10 | 107 | 155 | 69.03 |
| Total | 227 | 83 | 48 | 14 | 132 | 98 | 198 | 800 | 54.15 |
| PA (%) | 22.02 | 96.38 | 81.25 | 35.71 | 68.18 | 63.26 | 54.04 | | |
| OA | | | | | | | | | 54 |

**Table A.1.3**
Assessment of the sample classification results obtained by applying TWDTW and Random Forest refining methodology in TA1.

| | Potatoes | Sugar Beets | Grassland | Maize | Grains | Onions | OC | Total | UA(%) |
|---|---|---|---|---|---|---|---|---|---|
| Potatoes | 39 | 0 | 0 | 24 | 0 | 30 | 16 | 109 | 35.77 |
| Sugar Beets | 0 | 43 | 0 | 1 | 0 | 0 | 0 | 44 | 97.73 |
| Grassland | 0 | 0 | 18 | 0 | 1 | 0 | 2 | 21 | 85.71 |
| Maize | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 3 | 66.66 |
| Grains | 0 | 0 | 0 | 0 | 47 | 0 | 1 | 48 | 97.92 |
| Onions | 3 | 0 | 0 | 0 | 0 | 39 | 5 | 47 | 82.98 |
| OC | 0 | 0 | 11 | 0 | 4 | 3 | 31 | 49 | 63.27 |
| Total | 41 | 43 | 29 | 27 | 52 | 72 | 55 | 319 | |
| PA (%) | 95.12 | 100 | 62.07 | 7.41 | 90.38 | 54.17 | 56.36 | | |
| OA | | | | | | | | | 69 |

**Table A.1.4**
Assessment of the sample classification results obtained by applying TWDTW in TA2.

| | Potatoes | Sugar Beets | Grassland | Maize | Grains | Other class | Total | UA (%) |
|---|---|---|---|---|---|---|---|---|
| Potatoes | 27 | 2 | 4 | 8 | 12 | 3 | 56 | 48.21 |
| Sugar Beets | 18 | 29 | 12 | 1 | 0 | 0 | 60 | 48.33 |
| Grassland | 1 | 1 | 467 | 2 | 6 | 1 | 478 | 97.69 |
| Maize | 32 | 10 | 0 | 19 | 0 | 0 | 61 | 31.14 |
| Grains | 9 | 0 | 5 | 2 | 40 | 2 | 58 | 68.96 |
| OC | 37 | 4 | 12 | 7 | 20 | 6 | 86 | 6.976 |
| Total | 124 | 46 | 501 | 39 | 78 | 12 | 800 | |
| PA (%) | 21.77 | 63.04 | 93.21 | 48.71 | 51.28 | 50 | | |
| OA (%) | | | | | | | | 73.5 |

Some crops showed differences in patterns between the three areas. We could not directly explain this from the difference in temperature or precipitation (see Table 1). Also, over the winter period, differences can be seen. This may be due to different management practices in winter, with or without a second crop or green manure. The phenological pattern of grassland is mainly influenced by mowing, grazing, or harvesting of the seeds, which varies between fields.

The proximity-based clustering results provide useful information on the possible classification errors. For example, some potato samples and the great majority of maize samples were assigned to the same cluster in both target areas. Such cluster overlapping could be automatically flagged as an error and, therefore, the concerned samples would be further investigated through ground truth campaigns.

Our methodology performed well for study areas with a balanced

**Table A.1.5**
Assessment of the sample classification results obtained by applying TWDTW and Random Forest refining methodology TA2.

| | Potatoes | Sugar Beets | Grassland | Maize | Grains | OC | Total | UA (%) |
|---|---|---|---|---|---|---|---|---|
| Potatoes | 18 | 1 | 0 | 14 | 6 | 30 | 69 | 26.09 |
| Sugar Beets | 1 | 20 | 0 | 3 | 0 | 2 | 26 | 76.92 |
| Grassland | 2 | 10 | 256 | 0 | 2 | 6 | 276 | 92.75 |
| Maize | 4 | 0 | 0 | 7 | 1 | 4 | 16 | 43.75 |
| Grains | 4 | 0 | 3 | 0 | 19 | 12 | 38 | 50.00 |
| OC | 3 | 0 | 1 | 0 | 1 | 2 | 7 | 28.57 |
| Total | 32 | 31 | 260 | 24 | 29 | 56 | 432 | |
| PA (%) | 56.25 | 64.51 | 98.46 | 29.16 | 65.51 | 3.57 | | |
| OA (%) | | | | | | | | 75 |

number of crop samples (TA1). Yet, it favors prevalent classes in areas with an imbalanced number of samples at the expense of low accuracy for the marginal crops. In TA2, for example, the major crop, i.e. grassland, achieved the highest accuracy (PA and UA of 98% and 93%, respectively). Misclassification of potatoes, sugar beets, grains, and other classes was much more severe because these crops represent a smaller proportion of the sample sets. These results confirm the findings of previous studies that showed that sample noise removal strategies are biased towards the majority classes (Seiffert et al., 2014; Van Hulse and Khoshgoftaar, 2009). Existing balancing methods such the synthetic minority oversampling technique (SMOTE) might be used to reduce the errors caused by class imbalance (Waldner et al., 2019). Furthermore, the sample-refining strategies reduced considerably the number of samples. For example, we removed 481 samples from TA1 and 368 samples for TA2 (see Tables A.1.2–A.1.5 in Annex 1). We see this is as one of the disadvantages of these strategies as already reported in the literature (Miranda et al., 2009).

RF is a fast and efficient supervised classifier that has been used successfully in different remote sensing applications (Belgiu and Drăguţ, 2016), including classification of crops from time-series datasets (do Nascimento Bendini et al., 2019). This study took advantage of the proximity measures capability of RF to refine the sample set. It can also be used as a proxy for the presence of several subclasses within the same generic class, e.g. different grassland, potatoes, or maize types (Touw et al., 2012).

Previous studies used DTW and its variations for land use-land cover mapping (Maus et al., 2016; Petitjean et al., 2012) or crop mapping using either pixels (Li and Bijker, 2019) or objects (Belgiu and Csillik, 2018; Csillik et al., 2019) as the smallest units of analysis. Xue et al. (2014) used DTW to assess the quality of land cover samples identified through visual interpretation. We used DTW because of its recognized capability to address the classification challenges caused by the presence of gaps and shifts in the available time-series (Petitjean et al., 2012). All three time-series available for SA1, TA1, and TA2 consist of images acquired at different dates (Fig. 2).

Before applying RF-based proximity measures, we removed the outliers from the distribution of the spectral-temporal dissimilarity values. Previous studies excluded the outliers from spectral signatures using probabilistic iterative trimming such as the $\chi^2$ test on Mahalanobis distance between the samples and computed distribution (Desclée et al., 2006; Radoux et al., 2014). We also evaluated the performance of the proposed methodology without removing outliers from the training sample set. Yet, the results were less satisfactory (see Table A.1.1). These results contradict the findings in Zhu et al. (2016) that concluded that there is no need to remove the outliers from the training sample set.

The dependence on existing samples in the source area might limit the application of this method in areas lacking these data (Malambo and Heatwole, 2020). To address this limitation, samples from previous years (Huang et al., 2020). Although phenological development will vary over the years because of different weather conditions, we expect our method will be able to deal with these inter-annual variations, which we plan to test as part of future work. Since the spectral-temporal characteristics of the crops from different countries could vary significantly, samples of crops cultivated in the same agro-ecological zones might be a promising solution to this challenge as shown by Li et al. (2020) in a study dedicated to land cover mapping in Africa.

For future work, we will focus on the generation of spectral-temporal libraries for crops in the Netherlands and use them to regularly update crop data. The completeness of these libraries is important to avoid situations when not all crops from target areas are represented in the source area sample set. In addition, phenology-based labeled training samples will be further refined using spatial filters such as morphological filters defined for different window sizes (Radoux et al., 2014) and tested on robustness against inter-annual variability. If successful, the method can be expanded to countries in the global South, where agricultural fields are often smaller and more heterogeneous and, therefore, more challenging to map.

## 6. Conclusions

High-quality crop samples are essential for accurate supervised classification. This paper proposed an automatic generation of labeled crop datasets in two target areas by leveraging the spectral-temporal characteristics of crops from a similar area. Spectral-temporal dissimilarities between labeled samples from the source area and those unlabeled from target areas were measured using TWDTW. The quality of the automated labeled samples was improved by computing the proximity values between all samples using the RF classifier. Proposed method worked well for sugar beets and grain crops. Yet, it obtained less satisfactory results for the potatoes and maize crops because of their similarity in the spectral-temporal domain. Further investigations and improvements of the proposed methodology are required in areas with an imbalance in the area or number of fields per crop, where minority crops obtained less satisfactory results.

## Supplementary materials

The R code used for refining labeled samples and sample data are available at https://github.com/mbelgiu/LabelingTrainingSamples as supplementary material.

## CRediT authorship contribution statement

**Mariana Belgiu:** Conceptualization, Methodology, Validation, Writing - original draft, Writing - review & editing. **Wietske Bijker:** Conceptualization, Writing - review & editing. **Ovidiu Csillik:** Conceptualization, Writing - review & editing, Visualization. **Alfred Stein:** Writing - review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We thank the three anonymous reviewers for their constructive comments.

## Appendix A

See Tables A.1.1–A.1.5.

## References

Bégué, A., Arvor, D., Bellon, B., Betbeder, J., de Abelleyra, D., Ferraz, P.D.R., Lebourgeois, V., Lelong, C., Simões, M., Verón, R.S., 2018. Remote sensing and cropping practices: a review. Remote Sensing 10, 99.

Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sens. Environ. 204, 509–523.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31.

Belgiu, M., Zhou, Y., Marshall, M., Stein, A., 2020. Dynamic Time Warping for crops mapping. Int. Arch. Photogram., Remote Sensing Spatial Inform. Sci. 43, 947–951.

Breiman, L., 2001. Random forest. Mach. Learn. 45.

Buja, A., Swayne, D.F., Littman, M.L., Dean, N., Hofmann, H., Chen, L., 2008. Data visualization with multidimensional scaling. J. Comput. Graph. Stat. 17, 444–472.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37, 35–46.

Corcoran, J., Knight, J., Gallant, A., 2013. Influence of multi-source and multi-temporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in Northern Minnesota. Remote Sensing 5, 3212–3238.

Cox, M.A., Cox, T.F., 2008. Multidimensional scaling. Handbook of data visualization. Springer, pp. 315–347.

Csillik, O., Belgiu, M., Asner, P.G., Kelly, M., 2019. Object-based time-constrained dynamic time warping classification of crops using Sentinel-2. Remote Sensing 11.

Desclée, B., Bogaert, P., Defourny, P., 2006. Forest change detection by statistical object-based method. Remote Sens. Environ. 102, 1–11.

da Nascimento Bendini, H., Garcia Fonseca, L.M., Schwieder, M., Sehn Körting, T., Rufin, P., Del Arco Sanches, I., Leitão, P.J., Hostert, P., 2019. Detailed agricultural land classification in the Brazilian cerrado based on phenological information from dense satellite image time series. Int. J. Appl. Earth Obs. Geoinf. 82, 101872.

European-Union, 2018. Agriculture. In.

Foody, G.M., Boyd, D.S., Cutler, M.E., 2003. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. Remote Sens. Environ. 85, 463–474.

Fowler, J., Waldner, F., Hochman, Z., 2020. All pixels are useful, but some are more useful: efficient in situ data collection for crop-type mapping using sequential exploration methods. Int. J. Appl. Earth Obs. Geoinf. 91, 102114.

Frénay, B., Verleysen, M., 2013. Classification in the presence of label noise: a survey. IEEE Trans. Neural Networks Learn. Syst. 25, 845–869.

Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F., Obersteiner, M., 2009. Geo-Wiki. Org: the use of crowdsourcing to improve global land cover. Remote Sensing 1, 345–354.

Gómez, C., White, J.C., Wulder, M.A., 2016. Optical remotely sensed time series data for land cover classification: a review. ISPRS J. Photogramm. Remote Sens. 116, 55–72.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

Huang, H., Wang, J., Liu, C., Liang, L., Li, C., Gong, P., 2020. The migration of training samples towards dynamic global land cover mapping. ISPRS J. Photogramm. Remote Sens. 161, 27–36.

Kwak, G.-H., Park, N.-W., 2019. Impact of texture information on crop classification with machine learning and UAV images. Appl. Sci. 9, 643.

Li, M., Bijker, W., 2019. Vegetable classification in Indonesia using dynamic time warping of sentinel-1A dual polarization SAR time series. Int. J. Appl. Earth Obs. Geoinf. 78, 268–280.

Li, Q., Qiu, C., Ma, L., Schmitt, M., Zhu, X.X., 2020. Mapping the land cover of Africa at 10 m resolution from multi-source remote sensing data with google earth engine. Remote Sensing 12, 602.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news 2, 18–22.

Luciano, A.C.D.S., Picoli, M.C.A., Rocha, J.V., Franco, H.C.J., Sanches, G.M., Leal, M.R.L. V., le Maire, G., 2018. Generalized space-time classifiers for monitoring sugarcane areas in Brazil. Remote Sens. Environ. 215, 438–451.

Malambo, L., Heatwole, C.D., 2020. Automated training sample definition for seasonal burned area mapping. ISPRS J. Photogramm. Remote Sens. 160, 107–123.

Maus, V.G.C., Cartaxo, R., Sanchez, A., Ramos, F.M., Queiroz, G.R.D., 2016. A time-weighted dynamic time warping method for land-use and land-cover mapping. IEEE J. Selected Top. Appl. Earth Observ. Remote Sensing 1–11.

Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. Int. J. Remote Sens. 39, 2784–2817.

Miranda, A.L., Garcia, L.P.F., Carvalho, A.C., Lorena, A.C., 2009. Use of classification algorithms in noise detection and elimination. In: International Conference on Hybrid Artificial Intelligence Systems. Springer, pp. 417–424.

Mohammed, I., Marshall, M., de Bie, K., Estes, L., Nelson, A., 2020. A blended census and multiscale remote sensing approach to probabilistic cropland mapping in complex landscapes. ISPRS J. Photogramm. Remote Sens. 161, 233–245.

Okujeni, A., Canters, F., Cooper, S.D., Degerickx, J., Heiden, U., Hostert, P., Priem, F., Roberts, D.A., Somers, B., van der Linden, S., 2018. Generalizing machine learning regression models using multi-site spectral libraries for mapping vegetation-impervious-soil fractions across multiple cities. Remote Sens. Environ. 216, 482–496.

Pax-Lenney, M., Woodcock, C.E., Macomber, S.A., Gopal, S., Song, C., 2001. Forest mapping with a generalized classifier and Landsat TM data. Remote Sens. Environ. 77, 241–250.

Petitjean, F., Inglada, J., Gançarski, P., 2012. Satellite image time series analysis under time warping. IEEE Trans. Geosci. Remote Sens. 50, 3081–3095.

Radoux, J., Lamarche, C., Van Bogaert, E., Bontemps, S., Brockmann, C., Defourny, P., 2014. Automated training sample extraction for global land cover mapping. Remote Sensing 6, 3965–3987.

Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoust. Speech Signal Process. 26, 43–49.

Seiffert, C., Khoshgoftaar, T.M., Van Hulse, J., Folleco, A., 2014. An empirical study of the classification performance of learners on imbalanced and noisy software quality data. Inf. Sci. 259, 571–595.

Simoes, R., Picoli, M.C.A., Camara, G., Maciel, A., Santos, L., Andrade, P.R., Sánchez, A., Ferreira, K., Carvalho, A., 2020. Land use and cover maps for Mato Grosso State in Brazil from 2001 to 2017. Sci. Data 7, 34.

Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc.: Series B (Statistical Methodology) 63, 411–423.

Touw, W.G., Bayjanov, J.R., Overmars, L., Backus, L., Boekhorst, J., Wels, M., van Hijum, S.A., 2012. Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle? Briefings Bioinform., bbs034.

Tuia, D., Ratle, F., Pacifici, F., Kanevski, M.F., Emery, W.J., 2009. Active learning methods for remote sensing image classification. IEEE Trans. Geosci. Remote Sens. 47, 2218–2232.

Tuia, D., Volpi, M., Copa, L., Kanevski, M., Munoz-Mari, J., 2011. A survey of active learning algorithms for supervised remote sensing image classification. IEEE J. Sel. Top. Signal Process. 5, 606–617.

Van Hulse, J., Khoshgoftaar, T., 2009. Knowledge discovery from imbalanced and noisy data. Data Knowl. Eng. 68, 1513–1542.

Waldner, F., Chen, Y., Lawes, R., Hochman, Z., 2019. Needle in a haystack: Mapping rare and infrequent crops using satellite imagery and data balancing methods. Remote Sens. Environ. 233, 111375.

Wang, S., Azzari, G., Lobell, D.B., 2019. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. Remote Sens. Environ. 222, 303–317.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: a meta-review. Remote Sens. Environ. 236, 111402.

Woodcock, C.E., Macomber, S.A., Pax-Lenney, M., Cohen, W.B., 2001. Monitoring large areas for forest change using Landsat: generalization across space, time and Landsat sensors. Remote Sens. Environ. 78, 194–203.

Xiong, J., Thenkabail, P.S., Gumma, M.K., Teluguntla, P., Poehnelt, J., Congalton, R.G., Yadav, K., Thau, D., 2017. Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. ISPRS J. Photogramm. Remote Sens. 126, 225–244.

Xue, Z., Du, P., Feng, L., 2014. Phenology-driven land cover classification and trend analysis based on long-term remote sensing image series. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 1142–1156.

Zhang, X., Liu, L., Wang, Y., Hu, Y., Zhang, B., 2018. A SPECLib-based operational classification approach: a preliminary test on China land cover mapping at 30 m. Int. J. Appl. Earth Obs. Geoinf. 71, 83–94.

Zhu, X., Wu, X., 2004. Class noise vs. attribute noise: a quantitative study. Artif. Intell. Rev. 22, 177–210.

Zhu, Z., Gallant, A.L., Woodcock, C.E., Pengra, B., Olofsson, P., Loveland, T.R., Jin, S., Dahal, D., Yang, L., Auch, R.F., 2016. Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative. ISPRS J. Photogrammetry Remote Sensing 122, 206–221.