Mapping crop types in complex farming areas using SAR imagery with dynamic time warping

Getachew Workineh Gella*, Wietske Bijker, Mariana Belgiu

University of Twente Faculty of Geo-information Science and Earth Observation, Enschede, Netherlands

ARTICLE INFO

Keywords:
Crop type mapping
Decision level fusion
Sentinel-1
TerraSAR-X
Time Weighted Dynamic Time Warping

ABSTRACT

Crop type information is essential for many practical applications, yet its mapping is often constrained by inherent characteristics of most farming areas, such as fragmentation and small farm plots, changes in crop morphology across the season, and cloud cover. This study investigated whether these limitations could be overcome by using time-series of Synthetic Aperture Radar (SAR) and crop phenological information combined with different Dynamic Time Warping implementation strategies. Focusing on a fragmented landscape with small farm plots in the Netherlands, we used Sentinel-1 dual polarimetry (VV + VH) and TerraSAR-X single polarimetry (HH) images and the Dutch Basic Registration of Crop Plots (BRP) dataset for training and validation. Image pre-processing was followed by the generation of radar vegetation indices and polarimetric decomposition. Crop-specific responses to incident radar signal were analyzed, as well as the accuracy of the crop classification by Time-Weighted Dynamic Time Warping (twDTW) using either backscatter bands only or backscatter bands in combination with derived indices and polarimetric decomposition features. In addition, two further modified Dynamic Time Warping strategies, namely Variable Time Weight Dynamic Time Warping (vtwDTW) and Angular Metric for Shape Similarity (AMSS), were tested for their performance. Furthermore, we investigated the accuracy performance of a decision level fusion of TerraSAR-X and Sentinel-1 classification outputs. Results show that even in a fragmented landscape with relatively small plots (around 0.08 ha), crop types can be successfully mapped by using decision level fusion of the twDTW results of both sensors, reaching an accuracy of 77.1%. When using twDTW on Sentinel-1 (VV + VH) only, including Ratio, MRVI, and DPSVI indices, overall accuracy reached 69.5%; without those indices, accuracy was slightly lower (67.5%). Merging six different grain crops with similar leaf geometry into winter grains and summer grains improved classification accuracy to 80.6%. Our findings demonstrate that twDTW on SAR imagery allows to map crop types in fragmented landscapes with relatively small farm plots, offering potential for crop type mapping in areas with smallholder farming.

1. Introduction

Accurate information on crop production is essential for many applications. For example, annual information on crop types and areas grown is needed for decision making in food production and food security (Samberg et al., 2016), characterization of cropping intensity (Jain et al., 2013), soil and water resources research, especially erosion modelling (Panagos et al., 2015), generation of cost-effective information about agricultural production (Carfagna and Gallego, 2005) and estimation of crop damage for insurance payout (ESA, 2017).

Mapping this information based on single-date satellite imagery can be unsatisfactory, especially in complex farming areas (e.g. Delrue et al., 2013). One of the persisting bottlenecks is crop-specific seasonality, requiring time series observation of satellite imagery to characterize the different phenological phases of crop development. In the last few years, the domain of optical remote sensing has made good progress towards time-series analysis, including the formulation of automatic information retrieval procedures in smallholder farming areas (Stratoulias et al., 2017), parcel boundary delineation in complex farming systems (Persello et al., 2019), extraction of cropland extent (Olipham et al., 2019; Useya et al., 2019; Mohammed et al., 2020), crop types mapping (Belguj and Gällik, 2018) and parcel level classification of crops (Aguilar et al., 2018; Sonobe et al., 2018).

Despite the progress made, using optical satellites to account for crop
development across the growing season is still constrained by the persistent presence of clouds. For example, in the Netherlands, the probability of getting cloud-free images from an optical satellite with daily observation frequency is only 20% of the days in a year (van der Wal et al., 2013). In contrast, Synthetic Aperture Radar (SAR) satellites such as Sentinel-1 and TerraSAR-X are not hindered by cloud cover and can supply regular observations for monitoring of crop development (Khabbazan et al., 2019). Hence, some studies have tried to classify crop types from TerraSAR-X dual polarimetry (Hütt and Waldhoff, 2018; Sonobe, 2019) and Sentinel-1 dual polarimetry imagery (Danilla et al., 2017). However, a great deal of work is still needed to properly include crop phenological development into a classification scheme.

While the use of SAR sensors allows regular, all-weather observation for crop monitoring, other challenges remain which emanate from the inherent characteristics of complex farming areas. These include the presence of small farm plots, irregular planting dates, and spatial complexity resulting from fragmentation, intercropping, and complex terrain. As a result, at a given time, similar crops can exhibit different backscatter while different crops may exhibit similar backscatter. This inter-crop similarity and intra-crop heterogeneity can only be resolved in the temporal domain by using time-series imagery to account for the evolvement of scattering mechanisms. Many studies have attempted to address this issue using time series of SAR data with classifiers such as Random Forest (RF) (Hütt and Waldhoff, 2018), Support Vector Machine (SVM) (Kang et al., 2018), and Convolutional Neural Networks (CNN) (Danilla et al., 2017; Sonobe, 2019). However, although these classifiers can incorporate multi-temporal observations across the growing season as a stack of time-series images, their implementation strategy does not allow to leverage phenological information into a classification scheme, i.e. shuffling the temporal order of the images in a stack does not have any implications for crop classification performance.

To solve this problem, Dynamic Time Warping (DTW) is a more suitable classifier (Sakoe and Chiba, 1978), as it allows to account for phase shifts in backscatter and reflectance characteristics as a function of crop development. Using DTW, Petitjean et al. (2012) were able to classify land cover features from irregularly observed Satellite Image Time Series (SITS). Following the work of Jeong et al. (2011), which improved the original version of DTW by including time weights, Maus et al. (2016) classified land cover features from SITS using time-weighted Dynamic Time Warping (twDTW). Belgiu and Csillik (2018) and Csillik et al. (2019) applied twDTW to map crop types from optical imagery using objects as the smallest unit of analysis. Li and Bijker (2019) classified short-cycle vegetables with a high degree of accuracy from SAR imagery using twDTW combined with SPRING strategy for subsequence searching.

Building on the significant progress in DTW-based crop mapping made by these previous studies, we identified four additional issues that need further investigation. Firstly, while past studies mainly focus on homogenous farming landscapes with large plots (Belgiu et al., 2020; Csillik et al., 2019), the performance of DTW classifiers using SAR imagery of different spatial resolution and polarization (Sentinel-1 versus TerraSAR-X) has not yet been investigated for a fragmented area with smaller farm plots. Secondly, the few studies that did focus on fragmented areas with smallholder farming (Aguilar et al., 2018; Stratoulias et al., 2017) did not give due emphasis to incorporation of phenology into the classification scheme, while, in a given landscape, crop phenology can be complex due to varying planting and harvesting dates within the same crop. To incorporate this phenological information into the classification procedure requires an investigation of crop-specific time lags in the computation of linear or logistic time weight (vtwDTW), replacing the fixed weight used in twDTW. Thirdly, while previous studies mainly focused on similarity measures based on Euclidean distance, an angular distance measure could be useful for time series of radar imagery, as this measure is considered robust for temporal variation of observations (Teke and Yardımcı, 2019). And finally, although some studies have already been done on fusing optical images...
of different resolution (Zhao et al., 2019) and fusing optical images with SAR images (Waske and van der Linden, 2008), there is a lack of studies on fusing Sentinel-1 with TerraSAR-X images to benefit from the different spatial and temporal resolution and polarimetry of these all-weather sensing systems.

To address these issues, the current study mapped crop types in a complex farming area with small plots sizes, using time-series of Sentinel-1 and TerraSAR-X imagery, phenological information, and time-weighted DTW with both Euclidean and angular distance measures. In addition, we tested the use of a variable time lag in the computation of time weight in DTW to incorporate variations in crop phenology in the classification scheme. Finally, we devised a novel strategy for fusing medium resolution Sentinel-1 and high resolution TerraSAR-X data to increase mapping accuracy by using the potential of both sensing systems.

2. Materials and methods

2.1. Study site

Our study site was situated at 6° 45′ 39″ to 7° 3′ 29″ Longitude and 52° 15′ 5″ to 52° 21′ 33″ Latitude, in the eastern Netherlands (Over IJssel province) (Fig. 1).

Although the Netherlands is well-known for its well-planned agricultural landscape resulting from past land consolidation (van den Noort, 1987; van den Brink and Molema, 2008), the selected study site was characterized by irregular, relatively small plots (~0.08 ha) and fragmented holdings where crops dominate the landscape in a mixed pattern. In the Netherlands, main crop production takes place from the middle of February to the end of September (Fig. 2). The average total annual precipitation is 832 mm, while the coldest month is January with a mean monthly temperature of 2.3 °C (KNMI, 2019).

2.2. Data

For this study we used time-series of dual polarimetry Sentinel-1 SAR (VV and VH) data, obtained from the European Space Agency (ESA), and single polarimetry TerraSAR-X (HH) data, obtained from AIRBUS Defence and Space (see Table 1 for metadata). Both products were in Slant Range Single Look Complex (SLC) data format, which includes both the phase and amplitude information needed for polarimetric decomposition. Detailed descriptions of the Sentinel-1 and TerraSAR-X missions are provided by Torres et al. (2012) and Roth (2003), respectively.

For training and validation, we used the Netherlands’ Basic Registration of Crop Plots (BRP) for the year 2018. This registration system records the primary crops cultivated in a given year on all parcels registered in the Agricultural Area Netherlands (AAN). AAN was initially built on RGB air photos of 2008, with a resolution of 25 cm, and is regularly updated by users. The BRP is updated yearly, based on information provided by farmers, with a sounding date of May 15. AAN and BRP are made available via the ‘Public Data on the Map’ initiative of the Dutch Ministry of Economic Affairs and Climate via (https://www.pdok.nl/geo-services/-/article/basisregistratie-gewaspecrelen-brp-9464039d91ac261a857ee92e9f215250).

For terrain correction of SAR images, we used a 1-arc-second Surface Radar Topographic Mission (SRTM) Digital Elevation Model (DEM) from the United States Geological Survey (USGS) (https://urs.earthdata.nasa.gov).

2.3. Data preprocessing

First, orbit correction was applied to the Sentinel-1 data. As these images were sensed with Terrain Observation and Progressive Scan (TOPS) in successive sub-swaths and bursts within each sub-swath (De Zan and Guarnieri, 2006), burst lines were removed. As our study site was situated within two successive sub-swaths, these were separated and processed individually. After TOPS split operation, radiometric calibration (Miranda and Meadows, 2015) was done to reduce radiometric irregularities. Next, individually processed sub-swaths were merged to create a wider sub-scene. To reduce impacts of local terrain on backscatter, Range Doppler Terrain Correction was done using SRTM DEM. Given the fact that our study site was very small and covered only a small number of bursts in a sub-swath, we could assume the impact of incidence angle variation on backscatter to be negligible and not needing the incidence angle correction conducted for larger geographic regions such as in Arias et al. (2020). After terrain correction, an improved Lee Sigma filter with a window size of 7 by 7 pixels and a sigma of 0.9 (Lee et al., 2009) was used for speckle reduction. Next, images were converted to a decibel scale for both polarizations (VV and VH) as described in Miranda and Meadows (2015). Three secondary features were computed: the ratio between VH and VV bands, the Modified Radar Vegetation Index (MRVI) (Czuchlewski et al., 2003), and the Dual Polarization SAR Vegetation Index (DPSVI) (Periasamy,
To integrate the changing scattering mechanisms related to crop development into the classification scheme, we added polarimetric decomposition features Entropy, Anisotropy, and Angle ($H-A-a$) (Nasirzadeh dizaji et al., 2019). As for pre-processing the TerraSAR-X HH images, the obtained SLC data were radiometrically corrected and multi-looked. Similar to the Sentinel-1 SAR images, speckle was filtered from the TerraSAR-X images using a refined Lee sigma filter with similar parameters and terrain correction was applied using SRTM DEM.

Fig. 3. Temporal backscatter profiles of dominant crops imaged by (A) Sentinel-1 VV; (B) Sentinel-1 VH, and (C) TerraSAR-X HH polarimetry.
Finally, we tailored the BRP dataset to be used for training and validation, by removing grassland and horticultural crops from the dataset before starting the sampling procedure. The horticultural crops had too few fields per crop for adequate sampling.

2.4. Sampling strategy and creation of temporal backscatter profiles

Training samples were selected from the BRP dataset reference polygons in a two-stage sampling strategy. In the first stage, by using crop types as strata, 20 polygons were picked randomly from each stratum. For a crop with fewer than 20 polygons, all parcels were included in the training sample. The specified sample size for this study was mainly decided by considering the uniformity of existing terrain and climatic conditions in the study site. Evidently, random sampling cannot guarantee spatial heterogeneity and is vulnerable to spatial autocorrelation (Dobbie et al., 2008). Therefore, to control spatial autocorrelation during random selection, a neighbourhood constraint was imposed for each polygon. This was done by computing the minimum distance to the nearest neighbour polygon and then, within each stratum, selecting the first 20 polygons with the largest minimum distance. It should be noted that within a single polygon, there may also be significant variations in the backscatter characteristics resulting from differences in surface and subsurface soil moisture and crop development in response to variations in soil fertility. To account for this variation, we picked random pixel samples from each selected sample polygon, in proportion to the polygon areal extent, to serve as final training pixels. To validate classification output, we selected the centroid points of all polygons from the whole parcel data set. This yielded a dense number of points (Table 2) that were fairly well distributed across the test site.

After the selection of training samples from reference polygons, values for all observables (VV, VH, HH, derived indices, and polarimetric decomposed features) were extracted from the processed SITS per crop. These values were then aggregated into a single time series per crop per observable using the median value. The median statistic was selected because of its robustness regarding outliers, especially for skewed distribution in a series. To suppress noise in the temporal domain and create a smooth temporal profile representative of crop development, a Savitzky-Golay filter (Savitzky and Golay, 1964) was applied. This filter effectively suppresses noise in the temporal domain of time-series satellite observations (Chen et al. 2004) while preserving the time series structure, especially the positions of local minima and maxima, and smoothing unnecessary fluctuations. The generated temporal profiles provided clues on the shape-based separability of the different crops from the SAR signal (Section 3.1) and hence the classification accuracy (Section 2.5 and 3.2).

Fig. 4. Crop type maps resulting from twDTW on (A) Sentinel VV + VH, (B) TerraSAR-X HH, (C) Sentinel (VV + VH) + Indices and (D) Sentinel-1 (VV + VH) + Indices + Decomposed; (E) Reference polygons from BRP. (see Supplementary material 3 for full extent images).
2.5. Classification using twDTW

The first procedure in twDTW based classification (Maus et al., 2016) is the computation of the distance matrix between a data sequence at a given pixel location in SITS and a temporal profile of a specific crop (query sequence). Accordingly, for a data sequence with length $M$ and specific crop temporal profile with length $N$, we computed a distance matrix $D$ with a size of $M \times N$ using Euclidean distance as provided in Petitjean et al. (2012):

$$D[i, j] = \sum_{i=1}^{M} \left( SITS_{r,c}[i] - TPF_j \right)^2$$  \hspace{1cm} (1)

where $D[i, j]$ is the distance at row $i$ and column $j$ of the distance matrix $D$, $SITS_{r,c}[i]$ is a value at row $r$ and column $c$ in SITS at time $i$, $TPF_j$ is the temporal profile of a specific crop at a time $j$, and $w_{ij}$ is a logistic weight that accounts for the absolute differences between times $i$ and $j$, which was computed as:

$$w_{ij} = \frac{1}{1 + e^{-\alpha|i-j|\beta}}$$  \hspace{1cm} (2)

where $i$ is the time of observation in $SITS_{r,c}$ at the $i^{th}$ position in a time series, $j$ is the time of observation in the temporal profile at the $j^{th}$ position, $\beta$ is the maximum time lag for the warping window to search for a match, which is constant for all crops, and $\alpha$ is a user-defined constant to control the steepness of the slope. While previous studies fine-tuned $\alpha$ and $\beta$ values by using a set of user-defined values (Maus et al., 2016; Belgiu and Csillik, 2018), we implemented a grid search strategy where the classifier picks an optimal combination of values after recursively computing overall accuracy on the fly from a series of provided $\alpha$ and $\beta$ values. The warping path ($P$) of two sequences was then considered as a set of data points that define the mapping between the two signals $SITS_{r,c}$ and $TPF_j$ as:

$$P = p_1, p_2, p_3, \ldots p_k$$  \hspace{1cm} (3)

The $k^{th}$ element of the warping path can be defined as $p_k = (i, j)_k$ where $i$ and $j$ are an index of the data points in $SITS_{r,c}$ and $TPF_j$, respectively, that are part of the warping path. This warping path is subject to constraints of boundary condition, continuity, and monotonicity (Berndt and Clifford, 1994). As there are potentially many paths in a distance matrix $D$ which satisfy these constraints, the optimum alignment is generated by following a path that minimizes the sum of distances:

$$DTW \left( [SITS_{r,c}], TPF_j \right) = \min \left\{ \sum_{k=1}^{k} p_k \right\}$$  \hspace{1cm} (4)
Fig. 6. Crop type maps resulting from decision level fusion of Sentinel-1 and TerraSAR-X outputs, using (A) twDTW, (B) vtwDTW, and (C) AMSS. (D) Reference polygons.

Fig. 7. Crop type maps resulting from merging various crops into summer grains and winter grains, using (A) twDTW, (B) vtwDTW, and (C) AMSS. (D) Reference polygons.
Table 2
Training and validation points.

| S.N. | Crop       | Training polygons | Validation points |
|------|------------|-------------------|-------------------|
| 1    | Maize      | 20                | 1072              |
| 2    | Potato     | 20                | 51                |
| 3    | Rye        | 20                | 41                |
| 4    | Summer barley | 20          | 35                |
| 5    | Summer wheat | 13            | 13                |
| 6    | Triticale  | 20                | 35                |
| 7    | Winter barley | 7              | 7                 |
| 8    | Winter wheat | 16            | 16                |
| Total|            | 136               | 1270              |

Table 3
Accuracy of crop type classification using twDTW with different inputs.

| Input           | Sentinel-1 (VV + VH) | TerraSAR-X HH | Sentinel-1 (VV + VH + Indices) | Decomposed |
|-----------------|----------------------|---------------|-------------------------------|------------|
|                  | UA       | PA       | UA   | PA       | UA       | PA       | UA   | PA       | UA       | PA       |
| Maize            | 97.7     | 73.6     | 91.5 | 27.1     | 96.5     | 77.1     | 94.6 | 40.9     |          |          |
| Potato           | 20.8     | 72.6     | 5.2  | 62.8     | 15.6     | 43.1     | 4.9  | 15.7     |          |          |
| Rye             | 4.6      | 24.4     | 5.8  | 17.1     | 8.1      | 7.4      | 7.4  | 12.2     |          |          |
| Summer barley   | 8.2      | 17.1     | 13.8 | 31.4     | 11.3     | 20.0     | 0.0  | 0.0      |          |          |
| Summer wheat    | 11.9     | 14.3     | 14.1 | 21.5     | 15.4     | 9.5      | 15.4 | 23.1     |          |          |
| Triticale       | 11.9     | 14.3     | 8.1  | 17.1     | 9.7      | 7.4      | 7.4  | 12.2     |          |          |
| Winter barley   | 8.0      | 57.1     | 31.6 | 42.9     | 9.7      | 57.1     | 11.4 | 3.3      |          |          |
| Winter wheat    | 17.6     | 81.3     | 28.3 | 81.3     | 23.5     | 75.0     | 2.9  | 37.5     |          |          |
| Overall accuracy | 67.5     | 28.0     | 69.5 | 36.6 |

As presented by Sakoe and Chiba (1978) and Berndt and Clifford (1994), keeping the minimal values (valley points) and fulfilling search constraints can be achieved by using a dynamic programming approach. Therefore, we generated a cumulative distance cost matrix $C$ with size $M$ by $N$ where rows at index $i$ and column $j$ were recursively computed by summing a distance matrix $D_{ij}$ and the minimum of surrounding cumulative distance cost values as:

$$ C_{ij} = D_{ij} + \min\{D_{i-1,j}, D_{i,j-1}, D_{i-1,j-1}\} $$

which was further subjected to the following constraints:

$$ C_{ij} = \begin{cases} 
D_{ij} & \text{if } j = 1 \\
D_{ij} + C_{i-1,j-1} & \text{if } 1 < j < M \\
D_{ij} + C_{i-1,j} & \text{if } 1 < i < N 
\end{cases} $$

In a cumulative distance cost matrix, the optimal similarity or dissimilarity between two different time series observations is a value at the position $C[M,N]$. To label a pixel with a specific crop type, warping distances were computed for each crop recursively. Following a 1-Near-est Neighbour (1-NN) classification approach (Hastie et al., 2009), each pixel was assigned the crop type label that had the lowest twDTW distance, compared to other crops.

The classifier was tested using Sentinel-1 (VV and VH), Sentinel-1 backscatter, and TerraSAR-X (HH) backscatter coefficients. It was also rerun by incorporating derived secondary products and indices (Ratio, MRVI, DPSVI) and decomposed polarimetric features (H, A, $\alpha$) to test the effects of changing input observables on accuracy performance.

2.6. Classification using vtwDTW and AMSS

To account for variations, within the same crop type, in planting date and duration in the field, twDTW was tested with a crop-specific variable time lag in what hereafter will be named variable time weight Dynamic Time Warping (vtwDTW). The implementation of vtwDTW is almost similar to twDTW except for the computation of the logistic weight, changing Eq. (2) into:

$$ w_{ij} = \frac{1}{1 + e^{-w_{ij}/\beta_k^*}} $$

where $\beta_k$ is the maximum time lag for crop $k$ in the temporal profile, where crop-specific time lags were provided as independent input. We generated these time-lags based on the agricultural calendar for dominant crops (Fig. 2), where short-duration crops were given smaller time lags while longer-duration crops were given larger time lags. Based on values investigated by Jeong et al. (2011), the $v$ value was set to 0.1.

Both twDTW and vtwDTW were implemented by using Euclidean distance as a distance measure. As Euclidean distance is sensitive to outliers (Shirkhorshidi et al., 2015), we also tested Angular Metric for Shape Similarity (AMSS) using an angular distance measure, in this case, a cosine distance. Overall implementation of the classifier followed Nakamura et al. (2013) who used it for classification of various time series datasets. Pseudocodes for the sampling procedure, temporal profile creation, and algorithmic implementations of twDTW, vtwDTW, and AMSS are provided in Supplementary Material 1.

2.7. Decision level fusion of Sentinel-1 and TerraSAR-X outputs

To benefit from both Sentinel-1 (dual-polarized medium spatial resolution) and TerraSAR-X images (single polarization high spatial resolution), we implemented a decision level fusion (Zhang et al., 2009) on the individual classification outputs from Sentinel-1 and TerraSAR-X. These individual outputs showed different user and producer accuracies for the various crop types. To prioritize a crop type which is better separated in either TerraSAR-X or Sentinel-1, user and producer accuracies from each classification output for each crop were converted into a single accuracy metric as:

$$ A_{k,u} = w^*UA_{k,u} + w^*PA_{k,u} $$

$$ A_{k,p} = w^*UA_{k,p} + w^*PA_{k,p} $$

where $A_{k,u}$ and $A_{k,p}$ are normalized accuracy for a specific crop ($c$) for classifications from Sentinel-1 ($s$) and TerraSAR-X ($t$) images respectively, $UA$ and $PA$ are user and producer accuracies respectively, and $w$ is a weight for each accuracy, to which 0.5 was assigned to give equal weight to UA and PA. Then, the classification output from Sentinel-1 was resampled to TerraSAR-X pixel spacing using the Nearest Neighbour (NN) approach. We chose the NN resampling approach because of its suitability for resampling categorical and nominal data such as land cover classes, which cannot be interpolated. It is evident that NN resampling approach has limitations with regard to the positional uncertainty of linear features such as roads and rivers, creating a jagged pattern (Verbyla, 2002). Since crop polygons do not have a linear shape and individual polygons are covered by one single crop type, we assumed the positional error after resampling is negligible. For

Table 4
Accuracy of crop type classification using vtwDTW and AMSS with different inputs.

| Classifier | Input | Sentinel-1 VV + VH | TerraSAR-X HH | Sentinel-1 VV + VH | TerraSAR-X HH |
|-----------|-------|-------------------|---------------|-------------------|---------------|
| vtwDTW    | UA    | PA                | UA            | PA                | UA            | PA            |
| AMSS      | UA    | PA                | UA            | PA                | UA            | PA            |
| Maize     | 98.6  | 60.0              | 91.6          | 26.4              | 92.3          | 37.9          | 88.9          | 9.0          |
| Potato    | 13.0  | 80.4              | 5.4           | 66.7              | 7.1           | 56.9          | 4.2           | 39.2         |
| Rye       | 12.5  | 2.5               | 14.6          | 14.6              | 12.2          | 4.6           | 2.4           | 4.6          |
| Summer barley | 9.7    | 42.9             | 22.0          | 37.1              | 16.7          | 8.6           | 3.9           | 11.4         |
| Summer wheat | 14.3  | 15.4             | 4.7           | 23.1              | 2.4           | 23.1          | 2.0           | 46.2         |
| Triticale | 15.4  | 11.4             | 7.6           | 25.7              | 9.1           | 11.4          | 3.3           | 11.4         |
| Winter barley | 11.8  | 57.1             | 20.0          | 42.9              | 3.5           | 57.1          | 8.3           | 28.6         |
| Winter wheat | 18.2  | 75.0             | 26.1          | 75.0              | 0.0           | 0.0           | 1.7           | 12.5         |
| Overall accuracy | 56.9  | 28.6             | 35.8          | 10.6              |               |               |               |              |
verification we followed Porwal and Katiyar (2014) by doing a visual inspection of the fused classified image overlaid with the reference polygons of the field boundaries (see Fig. 6). To preserve the spatial detail in TerraSAR-X, we opted for resampling the Sentinel-1 output. Resampling the coarser-resolution Sentinel output provided the opportunity to align this image to the finer resolution TerraSAR-X image, making the combined images suitable for the pixel-wise overlay needed for decision level fusion.

Based on the computed single accuracy metric and co-registered images, the following decision rule was established to create a new crop type map:

\[
E[i,j] = \begin{cases} 
S[i,j] \text{, or } T[i,j], & i \neq A, \, j \neq A, \\
S[i,j], & A_i > A_j, \\
T[i,j], & A_i < A_j,
\end{cases}
\]  

(9)

where \( E[i,j] \) is the crop type label for crop \( c \) in a map classified by decision level fusion, \( S[i,j] \) is the crop type label for crop \( c \) in Sentinel-1 and \( T[i,j] \) is its TerraSAR-X counterpart, at the \( i^{th} \) row and \( j^{th} \) column of the pixel location. At each specific pixel location, the classifier picked the crop type with the highest normalized accuracy.

Crops with similar leaf dimensions that grow in a similar season can cause confusion in the classification (Bagrill and Herrmann, 2011; Hütt and Waldhoff, 2018). To test the accuracy implications of grouping such similar crops together, the models were re-run after merging the various grain crops into summer grains and winter grains respectively. Finally, we calculated the user, producer, and overall accuracies using validation points, following procedures presented by Congalton (1991).

3. Results

3.1. Crop-specific temporal backscatter profiles

The main crops grown in the study site showed different backscatter patterns, which also varied between Sentinel-1 and TerraSAR-X output (Fig. 3). For the co-polarized Sentinel-1 VV polarization band (Fig. 3A), the backscatter coefficients of all crops decreased up to the early weeks of May. After that, the backscatter coefficients for maize and potato abruptly increased until the end of June, stabilized during July and then started to drop around the beginning of August. Backscatter for summer wheat, winter wheat, summer barley, and winter barley showed a smaller increase in August. Backscatter of potato was higher than that of maize. From the beginning of May, backscatter of summer wheat and winter wheat continued to decrease until July and then increased again, while backscatter for winter barley showed a smaller increase in August. Backscatter of summer barley increased from May until October. Except for minor differences, triticale and rye followed a similar pattern of decreasing backscatter until August, followed by stabilization or a slight increase.

In the TerraSAR-X co-polarized HH band (Fig. 3C), all crops showed a similar backscatter pattern until the end of April. From May, we observed three clusters of crops with similar patterns: 1) maize and potato, 2) summer wheat, summer barley, triticale and rye and 3) winter wheat and winter barley.

3.2. Crop type mapping using twDTW

The accuracy of crop type mapping using twDTW varied with input datasets (Table 3). Overall accuracy was highest (69.5%) when using Sentinel-1 input combining backscatter and generated indices (Ratio, MRVI, and DPSVI) but this was only slightly higher than when using Sentinel-1 backscatter alone (67.5%). Adding polarimetric decomposition features (H-A-\( c \)) decreased the overall accuracy to 36.6%, and also significantly reduced the user and producer accuracies for potato. The lowest overall accuracy (29%) was achieved with the TerraSAR-X single polarimetry (HH) dataset. Large differences were observed between user and producer accuracies for the same crop (see Supplementary material 2 for confusion matrices).

These different accuracy levels were reflected in the crop classification maps (Fig. 4). Maps produced with Sentinel-1 backscatter coefficients only (Fig. 4A) or with added indices (Fig. 4C) resulted in outputs comparable to the reference crop polygons (Fig. 4E). However, the schemes produced with TerraSAR-X single polarimetry (Fig. 4B) and Sentinel-1 with decomposed features (Fig. 4D) yielded a poorer result, in which potato was overpredicted while maize was underpredicted. Confusion between maize and potato was very high when using TerraSAR-X HH (Fig. 4B).

3.3. Crop type mapping using vtwDTW and AMSS

Compared to twDTW (Section 3.2), the alternative classifiers vtwDTW and AMSS yielded lower accuracies. Using Sentinel-1 data, vtwDTW yielded an overall classification accuracy of 56.9%, while AMSS yielded only 35.8% (Table 4). For the TerraSAR-X data, accuracies achieved with vtwDTW and AMSS were even lower (28.6% and 10.6%, respectively). Fig. 5 shows a subset of the maps resulting from vtwDTW and AMSS of Sentinel-1 and TerraSAR-X data. The low accuracies achieved with TerraSAR-X data (Fig. 5B and 5D) were mainly due to maize crops being misclassified as potato crops. As indicated on the classification maps, the best result was achieved with vtwDTW on Sentinel-1 data (Fig. 5A), producing a crop type map comparable to reference crop polygons. AMSS provided a more spatially confused crop type map, with larger within-field variation. In particular, accuracy results were significantly reduced for winter grains and summer grains (Fig. 5C-D).

3.4. Crop type mapping using decision level fusion of Sentinel-1 and TerraSAR-X

Integration of Sentinel-1 and TerraSAR-X achieved better classification results than using either of them alone (Table 5). Decision level fusion of outputs from twDTW yielded an overall classification accuracy of 77.1%, while fusion of outputs from vtwDTW and AMSS achieved accuracies of 69.7% and 51.8%, respectively. The user and producer accuracies also improved. For example, for maize, twDTW on the separate outputs of Sentinel-1 and TerraSAR-X yielded producer accuracies of 73.6% and 27.1%, respectively (Table 3), while using the fused outputs increased this accuracy to 84.5% (Table 5). Overall, decision level fusion twDTW provided more homogeneous polygons, especially for maize, which better resembled the reference polygons (Fig. 6).

| Classifier | Decision level twDTW | Decision level vtwDTW | Decision level AMSS |
|------------|------------------------|------------------------|---------------------|
| Accuracy   | UA | PA | UA | PA | UA | PA |
| Maize      | 94.5 | 84.5 | 95.3 | 75.7 | 88.5 | 58.0 |
| Potato     | 23.8 | 70.6 | 15.6 | 80.4 | 5.4 | 25.5 |
| Rye        | 71.4 | 12.2 | 50.0 | 12.2 | 6.4 | 19.5 |
| Summer barley | 47.4 | 25.7 | 52.9 | 25.7 | 7.8 | 17.1 |
| Summer wheat | 10.8 | 30.8 | 11.8 | 30.8 | 7.1 | 15.4 |
| Triticale  | 25.0 | 17.1 | 12.1 | 11.4 | 10.6 | 14.3 |
| Winter barley | 11.3 | 85.7 | 12.2 | 71.4 | 5.0 | 28.6 |
| Winter wheat | 35.0 | 43.8 | 28.6 | 37.5 | 0.0 | 0.0 |
| Overall accuracy | 77.1 | 69.7 | 51.8 |

Table 5

Accuracy of crop type classification resulting from decision level fusion of Sentinel-1 and TerraSAR-X outputs.
time series of radar images with different spatial resolution and polarization between potato and maize (Fig. 7). Compared to the results achieved for randomly oriented and dense potato leaves (Khabbazan et al., 2019) and individual winter grains (Tables 3-5), the user and producer accuracies for the merged winter grains were significantly higher, reaching 59.6% and 62.6%, respectively.

The merging of grain crops using the backscatter dataset of Sentinel-1. The merging of grain crops with similar leaf morphology, geometry, and seasonal growth pattern improved crop classification accuracy (Table 6). The highest overall accuracy of 80.6% was achieved using twDTW, while the lowest accuracy, using AMSS, was still as high as 54.9%. Merging the grain classes also reduced within-polygon mixing of crops, especially the confusion between potato and maize (Fig. 7). Compared to the results achieved for individual winter grains (Tables 3-5), the user and producer accuracies for the merged winter grains were significantly higher, reaching 59.6% and 62.6%, respectively.

### 3.5. Crop type mapping using merged classes

Based on the observed backscatter profiles (Fig. 3) and confusion matrices (Supplementary material 2), we merged all grains grown in winter into the class ‘winter grains’ and those grown in summer into the class ‘summer grains’, while leaving potato and maize as individual classes. Next, we re-ran all classifiers (twDTW, vtwDTW, and AMSS) using the backscatter dataset of Sentinel-1. The merging of grain crops with similar leaf morphology, geometry, and seasonal growth pattern improved crop classification accuracy (Table 6). The highest overall accuracy of 80.6% was achieved using twDTW, while the lowest accuracy, using AMSS, was still as high as 54.9%. Merging the grain classes also reduced within-polygon mixing of crops, especially the confusion between potato and maize (Fig. 7). Compared to the results achieved for individual winter grains (Tables 3-5), the user and producer accuracies for the merged winter grains were significantly higher, reaching 59.6% and 62.6%, respectively.

### 4. Discussion

The central aim of this study was to assess the accuracy performance of different DTW classifiers in a fragmented farming area with relatively small plot sizes. Specifically, we investigated whether crop classification in areas of such complexity could be improved by 1) using twDTW on time series of radar images with different spatial resolution and polarization (Sentinel-1 versus TerraSAR-X), including decision level fusion of the two radar outputs; 2) using crop-specific logistic weights (vtwDTW) to account for variable planting and harvesting dates within the same crop type (time lags); 3) exploring angular distance (AMSS) as an alternative similarity measure to Euclidean distance; and 4) merging crops with similar leaf morphology and seasonality to reduce confusion between classes.

In general, the classification accuracy of the DTW classifiers tested in this study hinges on the separability of the temporal backscatter profiles of the various crops. The profiles measured in our study (Fig. 3) show that, in all observables (VV, VH, and especially HH), maize and potato showed a similar profile shape which only differed in amplitude. Before these crops emerge, backscatter decreases due to changes in surface roughness as a result of tillage (McNairn and Brisco, 2004); after sprouting, backscatter increases, levels, and then drops, reflecting the scattering responses synchronized with changing Leaf Area Index (LAI) (Macelloni et al., 2001), crop fraction of vegetation cover (Liao et al., 2018), crop height (Nasirzadehdizaji et al., 2019), and finally, harvesting (Harfenmeister et al., 2019; Shang et al., 2020). The backscatter peak from potato was higher than from maize, which is likely due to the randomly oriented and dense potato leaves (Khabbazan et al. 2019) and the greater effect of surface soil moisture in the less dense maize canopy (El Hajj et al., 2019). In contrast to maize and potato, the backscatter profiles of the various summer grains showed a continuous decrease throughout most of the season. This finding agrees with Macelloni et al. (2001), who found that the backscatter from crops with narrow dimension and vertically oriented leaves, such as wheat and barley, starts to drop as soon as their Leaf Area Index (LAI) increases. At the end of the growing season, backscatter from both summer and winter grains is influenced by the presence of crop residues and surface moisture after harvest (McNairn et al., 2001), subsequent tillage operations (McNairn and Brisco, 2004), weed regrowth, and planting of secondary crops. These secondary crops are not registered in the dataset we used for training and validation, but they may have influenced the temporal profiles of the primary crops and hence the accuracy of the classification.

The shape similarity of these temporal profiles was particularly strong in the profiles generated with TerraSAR-X single polarimetry (HH), despite the higher spatial resolution of this sensor (Fig. 3C). As a result, confusion between maize and potato and between the various grain crops was high, which negatively affected the accuracy performance of all classifiers (twDTW, vtwDTW, AMSS) when using TerraSAR-X data. We expect that the utilization of dual polarimetry TerraSAR-X (HH + VH) would have achieved better accuracy for crop type mapping (Sonobe, 2019), but this option was not available for our study area.

Of the three classifiers tested, the highest classification accuracy was achieved with twDTW on Sentinel-1 (VV and VH) data including Ratio, MRVH, and DPSVI indices (69.5%); without these indices, accuracy was slightly lower (67.5%) (Table 3). Including polarimetric decomposition features (H-A-\(\alpha\)) did not improve classification with twDTW, which confirmed findings by Li and Bijker (2019). Since Sentinel-1 is not fully polarimetric, it cannot effectively characterize surface and volumetric scattering from the crop canopy (Ji and Wu, 2015). Furthermore, crops that have the geometry and leaf morphology similar to wheat, barley, and triticale in one cluster and maize and potato in another cluster, respond similarly to the incident SAR signal. As crop height, LAI, and vegetation cover reach their peak, crops exhibit similar scattering characteristics for an extended period. It can be argued that polarimetric decomposition features (from fully polarimetric sensors) would be more relevant to identify crop emergence or to distinguish cropland from other covers. Kumar et al. (2020) noted that compact polarimetric SAR data from Radasat-2 with is \(\Omega\) decomposition could be used to characterize growth stages of rice, cotton, and sugarcane crops.

Although we found the crop classification accuracy achieved with twDTW on Sentinel-1 (VV + VH) promising, it was lower than the accuracies found by other studies using time series of optical images (Belgiu et al., 2020; Belgiu and Csillik, 2018; Csillik et al., 2019). However, the test sites of these studies were less fragmented and had larger plot sizes than our test area. As noted in recent studies (Arias et al., 2020; Orynbaikzyz et al., 2020), the existence of small and fragmented plots has a significant impact on classification accuracy for crop type mapping, as fragmentation causes an increase in the number of mixed pixels with different crops in one pixel. Among studies using SAR instead of optical imagery, Li and Bijker (2019) found that twDTW on Sentinel-1 data performed well for the classification of short-cycle vegetable crops, reaching an accuracy of 80% in a relatively complex farming area in Indonesia. In a less complex farming area in the Netherlands, Danilla et al. (2017), using SAR data with a machine and deep learning models followed by post-classification smoothing, reported accuracies comparable to our study (60.28%, 59.33%, 64.12%, and 70.85% for SVM, RF, CNN, and CNN + MRF, respectively). Like in our study, these authors also noted a strong confusion between crops with similar physical structure. In our case, we did not implement a post-classification smoothing procedure, to avoid removal of small farm plots.

Contrary to our expectations, introducing a variable time lag for logistic weight calculation (vtwDTW) to account for crop-specific variation in planting date resulted in slightly higher accuracy than using a fixed time lag (twDTW) (compare Tables 3 and 4). Most likely this was due to the less precise crop calendar used for this study, which was generated from the literature and not from local field observations. In addition, the high confusion between crops with similar temporal profiles may have played a role. Furthermore, our test site had distinct seasons and mechanized agriculture, which likely resulted in crops being planted in a relatively short time span at the start of the season, decreasing the time lag between early and late sowing dates of the same

| Classifier   | twDTW | vtwDTW | AMSS |
|--------------|-------|--------|------|
| Crop type    | PA    | UA     | PA   | UA   | PA   | UA   |
| Maize        | 84.5  | 94.5   | 75.7 | 95.3 | 58.0 | 88.5 |
| Potato       | 70.6  | 23.8   | 80.4 | 15.6 | 25.5 | 5.4  |
| Summer grains| 39.6  | 33.9   | 33.3 | 31.4 | 29.2 | 13.3 |
| Winter grains| 62.6  | 59.6   | 57.6 | 54.3 | 48.5 | 21.8 |
| Overall      | 80.6  | 72.8   | 54.9 |

Table 6: Accuracy of crop type classification after merging crops with similar leaf structure and growing season, using Sentinel-1 and TerraSAR-X data.
crop. It would be worthwhile to test vtwDTW in an area with larger time lags and less pronounced seasons.

Like for vtwDTW, accuracies achieved with AMSS were lower than for twDTW, whether used on Sentinel-1 or TerraSAR-X data (compare Tables 3 and 4). Previous studies on AMSS for time series classification have reported mixed results. For example, Choi and Kim (2018) concluded that AMSS was more robust than other DTW classifiers for gesture recognition, while Nakamura et al. (2013) reported that the performance of AMSS depended on the dataset. The latter finding calls for investigation of the SAR or optical datasets to establish a concrete knowledge base on the predictive potential of the AMSS model for crop classification. In the remote sensing domain, Teke and Yardımcı (2019) implemented AMSS for crop type mapping from harmonized Landsat 8 and Sentinel-2 time-series and reported higher accuracies than our study. The difference between their findings and ours is most likely due to variations in input images and training samples, as well as differences in plot sizes and complexity of the study area. For example, Teke and Yardımcı (2019) studied a less complex area with only two crops: maize and cotton, while our study included more crops and small farm plots. Part of the low performance of AMSS lies in the strong confusion between maize and potato. We assume this could be attributed to a strong scattering similarity of the two crops as they are both broadleafed crops. Moreover, both crops contain different varieties with minor differences in sowing date and growing period. This may all constrain the model’s learning existing temporal patterns using angular metrics for shape similarity.

For all classifiers tested (twDTW, vtwDTW, and AMSS) decision level fusion of Sentinel-1 and TerraSAR-X outputs yielded better results than when using the classifiers on the individual datasets (compare Tables 3, 4 and 5). This has two main reasons. First, for a given pixel, decision level fusion assigns the crop type with better separability based on the standardized accuracy metric from the confusion matrix. Second, the classifiers benefit from the synergy of Sentinel-1 dual polarimetry and TerraSAR-X high spatial resolution. While studies on fusing SAR data of different spatial resolution and polarimetry are lacking, studies on fusing SAR with optical data have found decision level fusion effective to increase accuracy performance for crop type mapping (e.g., Waske and van der Linden, 2008). Our decision level fusion of SAR data achieved higher accuracy (77.1%) than what Aguilar et al. (2018) found for complex farming areas with the voting of ensemble classifiers using time series of optical Worldview-2 imagery (75.9%). Results of Orynbaikzyzy et al. (2020), using decision level fusion of optical and SAR images and merging of similar crops, also reported comparable results to ours. In addition, as indicated in Fig. 6, results of decision level fusion of NN-resampled output of Sentinel-1 and output of TerraSAR-X have smooth boundaries, in agreement with the reference polygons, which confirms good positional accuracy. Provided each polygon contains a single crop and polygons are not linear features, distortions in positional accuracy of pixels as a result of NN-resampling were negligible.

Finally, using the Sentinel-1 backscatter dataset, we found that merging the six different grain crops into two groups, i.e. summer grains and winter grains, significantly improved the accuracy of crop classification (Table 6). The merging of grains reduced within-polygon confusion of crops, not only for the grains but also for maize and potato (Fig. 7), which increased the overall accuracy of all classifiers. The highest overall accuracy of 80.6% was achieved with twDTW, while the lowest accuracy, using AMSS, was still as high as 54.9%. Hätt and Wilde (2018) achieved similar results from dual polarimetric TerraSAR-X data with an RF classifier. Following a hierarchical approach, Bargiel and Herrmann (2011) also reported significant accuracy improvements in classification when grouping crops with similar leaf geometry. In a study that fused Sentinel-1 and Sentinel-2 datasets for crop mapping, the grouping of cereals and legume classes also significantly increased the classification accuracy (Orynbaikzyzy et al., 2020).
Shirkhorshidi, A.S., Aghabozorgi, S., Ying Wah, T., 2015. A Comparison study on similarity and dissimilarity measures in clustering continuous data. PLoS One 10. https://doi.org/10.1371/journal.pone.0144059.

Sonobe, R., 2019. Parcel-based crop classification using multi-temporal TerraSAR-X dual polarimetric data. Remote Sens. 11 https://doi.org/10.3390/rs11101148.

Sonobe, R., Yanaya, Y., Tani, H., Wang, X., Kobayashi, N., Mochizuki, K., 2018. Crop classification from Sentinel-2-derived vegetation indices using ensemble learning. J. Appl. Remote Sens. 12, 1. https://doi.org/10.1117/1.jrs.12.026019.

Stratoulias, D., Tolpekin, V., de By, R.A., Zurita-Milla, R., Retson, V., Bijker, W., Hasan, M.A., Vermote, E., 2017. A workflow for automated satellite image processing: From raw VHSR data to object-based spectral information for smallholder agriculture. Remote Sens. https://doi.org/10.3390/rs9101048.

Teke, M., Yardimci, Y., 2019. Multi-Year Vector Dynamic Time Warping Based Crop Mapping.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B.O., Flourney, N., Brown, M., Traver, I.N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L’Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., Rostan, F., 2012. GMES Sentinel-1 mission. Remote Sens. Environ. 120, 9–24. https://doi.org/10.1016/j.rse.2011.05.028.

van den Brink, A., Molema, M., 2008. The origins of Dutch rural planning: A study of the early history of land consolidation in the Netherlands. Plan. Perspect. 23, 427–453. https://doi.org/10.1080/02665430802319005.

van den Noort, P.C., 1987. Land consolidation in the Netherlands. Land use policy 4, 11–13. https://doi.org/10.1016/0264-8377(87)90004-4.

van der Wal, T., Abma, B., Viguria, A., Prévinaire, E., Zarco-Tejada, P.J., Serruya, P., Van Valkengoed, E., Van Der Voet, P., 2013. Fieldcopter: Unmanned aerial systems for crop monitoring services, in: Precision Agriculture 2013 - Papers Presented at the 9th European Conference on Precision Agriculture, ECPA 2013. pp. 169–175.

Verbyla, D.L., 2002. Practical GIS Analysts, Practical GIS Analysts. CRC Press, London. https://doi.org/10.4324/97802032217931.

Waske, B., van der Linden, S., 2008. Classifying Multilevel Imagery From SAR and Optical Sensors by Decision Fusion. IEEE Trans. Geosci. Remote Sens. 46, 1457–1466. https://doi.org/10.1109/TGRS.2008.916089.

Zhang, D., Song, F., Xu, Y., Liang, Z., 2009. Decision Level Fusion. In: Advanced Pattern Recognition Technologies with Applications to Biometrics. Information Science Reference, New York, pp. 328–334.

Zhao, L., Shi, Y., Liu, B., Hovis, C., Duan, Y., Shi, Z., 2019. Finer classification of crops by fusing UAV images and sentinel-2A data. Remote Sens. 11 https://doi.org/10.3390/rs11243012.