*Manuscript*

# Early crop-type mapping under climate anomalies

**Michael Marszalek [1]\*, Maximilian Lösch [1], Marco Körner [2] and Urs Schmidhalter [1]**

[1]  Chair of Plant Nutrition, Technical University of Munich, Emil-Ramann-Straße 2, 85354 Freising, Germany; schmidhalter@wzw.tum.de
[2]  Chair of Remote Sensing Technology, Technical University of Munich, Arcisstraße 21, 80333 München, Germany; marco.koerner@tum.de
\*  Correspondence: marszalek@wzw.tum.de

**Abstract:** Crop-type mapping is an important intermediate step for cost-effective crop management at the field level, as an overview of all fields with a particular crop type can be used for monitoring or yield forecasting, for instance. Our study used a data set with 2400 fields and corresponding satellite observations from the federal state of Bavaria, Germany. The study classified corn, winter wheat, winter barley, sugar beet, potato, and winter rapeseed as the main crops grown in Upper Bavaria. We additionally experimented with a rejection class "Other", which summarised further crop types. Corresponding Sentinel-2 data included the normalised difference vegetation index (NDVI) and raw bands from 2016 to 2018 for each selected field. The influence of raw bands compared to NDVI was analysed and the classification algorithms, i.e. support vector machine (SVM) and random forest (RF), were compared. The study showed that the use of an index should be critically questioned and that raw bands provided a wider spectral bandwidth, which significantly improved the mapping of crop types. The results underline the use of RF with raw bands and achieved overall accuracies (OA) of up to 92%. We also predicted crop types in an unknown year with significantly different weather conditions and several months before the end of the growing season. Thus, the influence of climate anomalies and the accuracy depending on the time of prediction were assessed. The crop types of a test site and year without labels could be determined with an OA of up to 86%. The results demonstrate the usefulness of the proof-of-concept and its readiness for use in real applications.

**Keywords:** Precision farming; Early crop-type mapping; Sentinel-2; Random Forest; SVM

---

## 1. Introduction

Agricultural productivity needs to be increased and its environmental impact must be reduced to meet the future demand for food and achieve sustainable development goals (SDGs). The recognition of crop types and their spatial distribution is becoming increasingly important, as many new methodological developments are based on their knowledge [1]. Knowledge of the crop type is essential because it can be used to derive targeted and crop type-specific management zones and management decisions, to make crop type-specific forecasts of crop development, and to make advanced yield prediction on a field- and region-specific basis. In the context of cross-compliance measures, knowledge of the crop grown and documentation of its areas is also crucial, as this is also the basis for support payments and sustainability can be better assessed through knowledge of the crop types. Through "green direct payments" (or "greening"), the European Union (EU) supports farmers who are trying to meet environmental and climate goals. One of the required measures is crop diversification, which must be controlled. Efficient crop mapping will promote this process. Thus, efficient agricultural land classification gives suppliers, policy makers and governments an up-to-date overview of the land under cultivation.

Remote sensing provides multispectral data, and especially since the launch of the Sentinel mission, the spatial and temporal availability of data has been continuously increasing. With its various satellites and sensors, the Sentinel mission enables innovative applications in different domains, such

as agriculture or geology. The spatial and temporal resolutions are suitable for precision agriculture and its applications. This study focuses on the capabilities of Sentinel-2A and 2B, provides an in-depth look at crop mapping, and addresses critical issues for implementing practical crop-type mapping. The lack of benchmarks and various metrics makes it difficult to compare performance, which is why we have published some of our data for research purposes [2].

Data from the Moderate Resolution Imaging Spectroradiometer (MODIS) are generally used for various applications at the county level [3,4] but do not meet the spatial resolution requirements for field-level mapping. Crop-type mapping is usually based on radar or optical data from Sentinel-1, Sentinel-2 or Landsat-8, respectively. Some researchers fused optical and radar data and achieved accuracies of up to 82% [5]. The authors pointed out that the fusion of Sentinel-1 and Sentinel-2 improved the results [5,6] because cloud cover limits the availability of Sentinel-2 data. Our experience has shown that in regions with moderate cloud cover, as is the case in this study, very high accuracies can be achieved without fusion. Past research studies also differed in temporal resolution and frequency. For instance, Saini et al. [7] tested single-date images from Sentinel-2 using support vector machines (SVMs) and Random Forest (RF) and showed that the OA's of the classifications achieved by RF and SVM using one Sentinel-2 image were 84.22% and 81.85%, respectively [7]. While some studies used only one image for classification [7–13], other researchers have preferred multi-temporal information [14–18]. A long short-term memory neural net (LSTM) achieved an overall accuracy of up to 89.7% with Sentinel-2 time series [19,20]. Another study based on support vector regression (SVR) and artificial neural network (ANN) achieved classification accuracies of up to 92% for broad-leaved forests [18]. In general, the multi-temporal approach captures better spectral characteristics over the growing season and improves crop types distinction. Methods such as random forest, neural networks, maximum likelihood or SVMs have been frequently applied [17,21–23].

In most studies, indices such as the NDVI were used to derive plant-specific differences [3,5,17, 18,24–26]. Belgiu et al. [17] described an object-based approach combined with NDVI time series. The authors investigated three different test areas and achieved up to 96.2% accuracy for one area, while the other two areas could not be classified with comparable accuracy. To obtain higher spectral differentiation, we evaluated the use of raw bands in our study and compared it with an index-based approach. Immitzer et al. [8] highlighted the importance of the red-edge and shortwave infrared (SWIR) bands for crop-type classification. Our approach benefits from the usage of all raw bands in combination with an object-based approach where a mean value is determined for each acquisition and each field to reduce the complexity while still reflecting the temporal patterns. Object-based approaches and image segmentation simplify crop-type classifications and provide time series for each field [27,28]. Nevertheless, sometimes pixel-based approaches have to be chosen to reveal the spatial differences in the field.

Previous work rarely addressed the prediction of crop types several weeks before harvest or for a year without labels, and in particular, the impact of different environmental conditions on crop-type identification was rarely addressed. Unusual weather conditions due to climate change or agricultural management practices affect the temporal patterns of time series [17,29,30]. In many cases, ground truth data are insufficient or not available in a timely manner, and it remains a particular challenge to verify crop types in an unknown area or in subsequent years. Hao et al. [24] predicted crop types in absence of ground truth data. They achieved classification accuracy up to 87.13% using MODIS and Landsat NDVI time series. In another work, the use of Sentinel-2 data allowed the correct determination of cotton, spring, and summer corn 6 to 8 weeks prior to harvest [31], while winter wheat could be well classified by using EVI and NDVI time series around harvest. The use of a cost function combined with a stopping rule for crop classification was also investigated and indicated that, depending on the crop type, satellite observations through June or July were essential to successful classifications [32,33]. To address these issues, Google Earth Engine (GEE) was used in this work. It provides an efficient environment for satellite data processing and enables the implementation of various satellite-based applications and research activities [34–36]. GEE allows the reproducibility

of research studies without the need to develop own processing pipelines. Our data set differs from previous studies regarding the number of fields and by inspecting a total of 2400 fields grown in three years with multiple crops, as well as referencing Sentinel-2 imagery with ground observations. The data set covers the years 2016 to 2018. Plant growth in 2018 was affected by a climate anomaly with significant drought and higher temperatures that impacted plant growth.

The objectives of this research are to find out whether (i) index-based classification is suitable for mapping crop types, (ii) raw bands help to improve classification, (iii) how well crop types can be classified in a year with widely differing weather conditions, and (iv) whether the classification accuracy depends on the phenological phase?

## 2. Materials and methods

### 2.1. Study site

The study area is Upper Bavaria in southern Germany, with a size of approximately 17529 square kilometers, of which about 7875 square kilometers make up agricultural land [37]. The main crop types are corn, winter wheat, barley, rapeseed, sugar beet, and potato. Of these, corn, winter wheat, and barley are the predominant crops. In general, Bavaria has an annual mean temperature of 8.6 °C and an annual mean precipitation of 811 mm. Figure 1 shows the typical crop periods with the sowing and harvest months. Figure 2 illustrates the average yearly maximum and minimum temperatures, monthly precipitation, and the test site. We focused on the vegetation period 2016–2018 and summarised the most important weather parameters in Table 1. For example, the average temperature during the March to July growing season in 2016 was 12.3 °C, while in 2018 it reached 13.7 °C. The rainfall amount as well as the ET0 values (FAO-56) [38,39] are based on the mean sum per field. While the rainfall is similar over the years, the evapotranspiration has steadily increased. This is partly due to rising temperatures and indicates an increased water demand. Unfavourable weather conditions did not only affect crop yields or reflectance values. These anomalies also affected crop-type mapping in 2018. Table 1 also shows an overview of the fields for crop-type classification. A total of 2400 samples were used for the experiments, with all 7 classes evenly distributed. 2099 fields were randomly chosen from different regions in Upper Bavaria to match the whole administrative district. In addition, a validation set of 301 neighboring fields from one region and for the year 2018 was created to apply the trained models to a coherent area and generate images for visual assessment. The Bavarian State Ministry of Agriculture and Forestry (StMELF) provided the underlying ground-truth data with all fields and corresponding crop types. The study focuses on Upper Bavaria and the crops considered are corn, winter wheat, winter barley, sugar beet, potato, winter rapeseed, and an additional rejection class called "Other", which covers all other crops, such as grassland or summer barley.

**Table 1.** Overview of the samples per year with corresponding climatic conditions. The coherent area is shown in Figure 2 a), while all other fields (samples) in Upper Bavaria were randomly selected.

|       | Crop samples | Coherent area | Temperature (°C) | Rain (mm) | ET0 (mm) |
|-------|-------------|---------------|------------------|-----------|----------|
| 2016: | 699         | 0             | 12.3             | 308.5     | 447.5    |
| 2017: | 700         | 0             | 13.0             | 292.9     | 501.9    |
| 2018: | 700         | 301           | 13.7             | 315.5     | 511.4    |

| Crop Type | Jan | | Feb | | Mar | | Apr | | Mai | | Jun | | Jul | | Aug | | Sep | | Okt | | Nov | | Dez | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Corn | | | | | | | | | | | | | | | | | | | | | | | | |
| Winter wheat | | | | | | | | | | | | | | | | | | | | | | | | |
| Winter barley | | | | | | | | | | | | | | | | | | | | | | | | |
| Winter rapeseed | | | | | | | | | | | | | | | | | | | | | | | | |
| Potato | | | | | | | | | | | | | | | | | | | | | | | | |
| Sugar beet | | | | | | | | | | | | | | | | | | | | | | | | |

Sowing ▢   Growth ▮   Harvest ▢
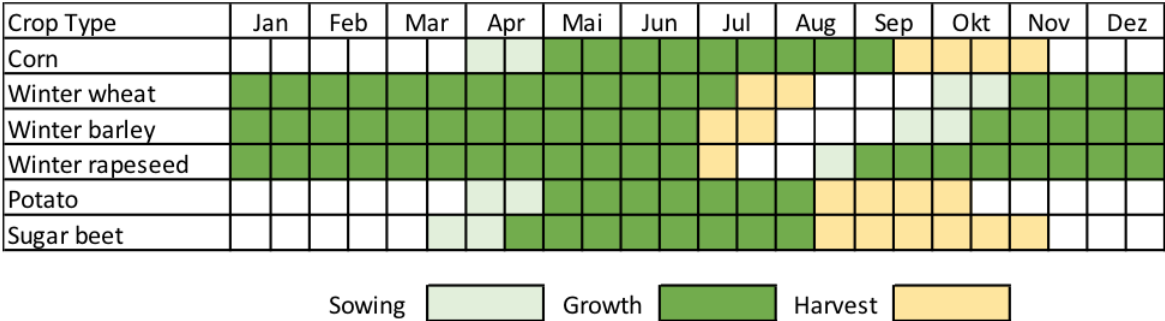
**Figure 1.** Crop phenology of crop types considered. Sowing is marked in lime, green stands for the vegetation period, and yellow covers the harvest period in Bavaria.
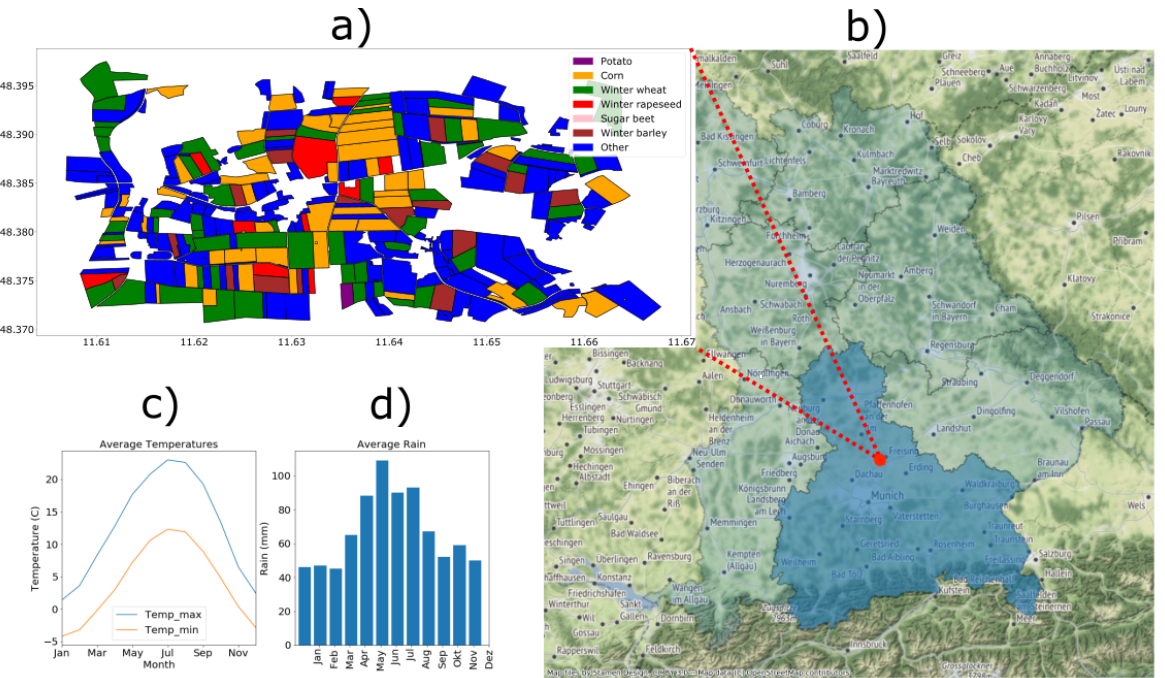
**Figure 2.** Climate overview of Upper Bavaria and the test site. Figure a) shows a test site with decimal latitude and longitude coordinates (WGS 84). Figure b) shows Upper Bavaria and Bavaria as an overlay. Figure c) and d) show the average maximum and minimum temperatures and the monthly precipitation.

*2.2. Sentinel-2*

The Sentinel-2 mission, with its two twin satellites Sentinel-2A and Sentinel-2B, monitors the Earth with a resolution of up to 10 m/pixel and a revisit time of 2–5 days depending on latitude. The availability of useful data is nevertheless limited due to clouds. Each identical satellite is equipped with a multispectral sensor, that covers 13 spectral bands and has a swath width of 290 km. Sentinel-2A was launched in June 2015 and Sentinel-2B was launched in March 2017. Sentinel-2 provides visible and near-infrared (NIR) spectral bands (B2, B3, B4, B8) with a resolution of 10 m/pixel and SWIR bands (B11, B12) with a resolution of 20 m/pixel. Band 1 is for aerosol, while bands 9 and 10 are suitable for detecting water vapour and cirrus clouds. These bands provide a resolution of 60 m/pixel for atmospheric corrections and do not need higher resolutions. Table 2 summarises the properties. In this study, Sentinel's Level-1C product was downloaded from Google Earth Engine, which stores terabytes of up-to-date satellite data and provides an efficient and fast way to access this information through a JavaScript or Python interface [35,40]. Raw bands and NDVI time series were downloaded for each test field and year.

**Table 2.** Overview of raw bands of Sentinel-2 [41] and applied NDVI.

| Band or Index | Central Wavelength (nm) or Formula | Description |
|---|---|---|
| Band 1 | 443.9 (S2A) /442.3 (S2B) | Aerosol (60 m) |
| Band 2 | 496.6 (S2A) /492.1 (S2B) | Blue (10 m) |
| Band 3 | 560 (S2A) /559 (S2B) | Green (10 m) |
| Band 4 | 664.5 (S2A) / 665 (S2B) | Red (10 m) |
| Band 5 | 703.9 (S2A) / 703.8 (S2B) | Red Edge (20 m) |
| Band 6 | 740.2 (S2A) / 739.1 (S2B) | Red Edge (20 m) |
| Band 7 | 782.5 (S2A) / 779.7 (S2B) | Red Edge (20 m) |
| Band 8 | 835.1 (S2A) / 833 (S2B) | NIR (10 m) |
| Band 8A | 864.8 (S2A) / 864 (S2B) | NIR (20 m) |
| Band 9 | 945 (S2A) / 943.2 (S2B) | Water vapour (60 m) |
| Band 10 | 1373.5 (S2A) / 1376.9 (S2B) | Cirrus (60 m) |
| Band 11 | 1613.7 (S2A) / 1610.4 (S2B) | SWIR (20 m) |
| Band 12 | 2202.4 (S2A) / 2185.7 (S2B) | SWIR (20 m) |
| NDVI | $\dfrac{(B8 - B4)}{(B8 + B4)}$ | Ratio of red to NIR |

Furthermore, Google Earth Engine provides Level-2A (L2A) images, which are bottom-of-atmosphere images, based on the European Space Agency's Sen2Cor open source processor [42]. This processor handles all the necessary atmospheric corrections and scene classification tasks. Google Earth Engine provides Level-2A images starting from March, 2017. Former data needs to be processed with Sen2Cor. Current methods for generating surface reflectance products are compared in the scope of the Atmospheric Correction Intercomparison Experiment (ACIX) and the research community [43,44]. We decided to use the Level-1C product for this research study

because L2A is not mandatory to show that an index is not the appropriate choice for crop-type detection, and to investigate predictive ability. Nevertheless, in a previous work, we observed that L2A improves the quality of images but minimally improves the prediction accuracy for some applications. This has also been observed in other research studies [20]. Figure 3 show the temporal patterns for every crop type. We compared the minimum and maximum values, the standard deviation, and the mean value to determine a significant feature. Time, as a feature, uncovers differences as well. The mean values separate the crop types very well and are used in our experiments for the raw bands as well as the NDVI. We computed the NDVI using Google Earth Engine and downloaded all the results into a corresponding file for further analysis. The NDVI was developed to monitor biomass or plant health and has been widely used for various applications. For each field, the pixels were aggregated to a mean value to generate a time series for each field. The satellite images were taken from February to August. All data was downloaded using a cloud meta filter of 20%. The CLOUDY_PIXEL_PERCENTAGE parameter considers cloudy pixels for a Sentinel-2 scene but does not recognise all cloudy pixels at the field level.
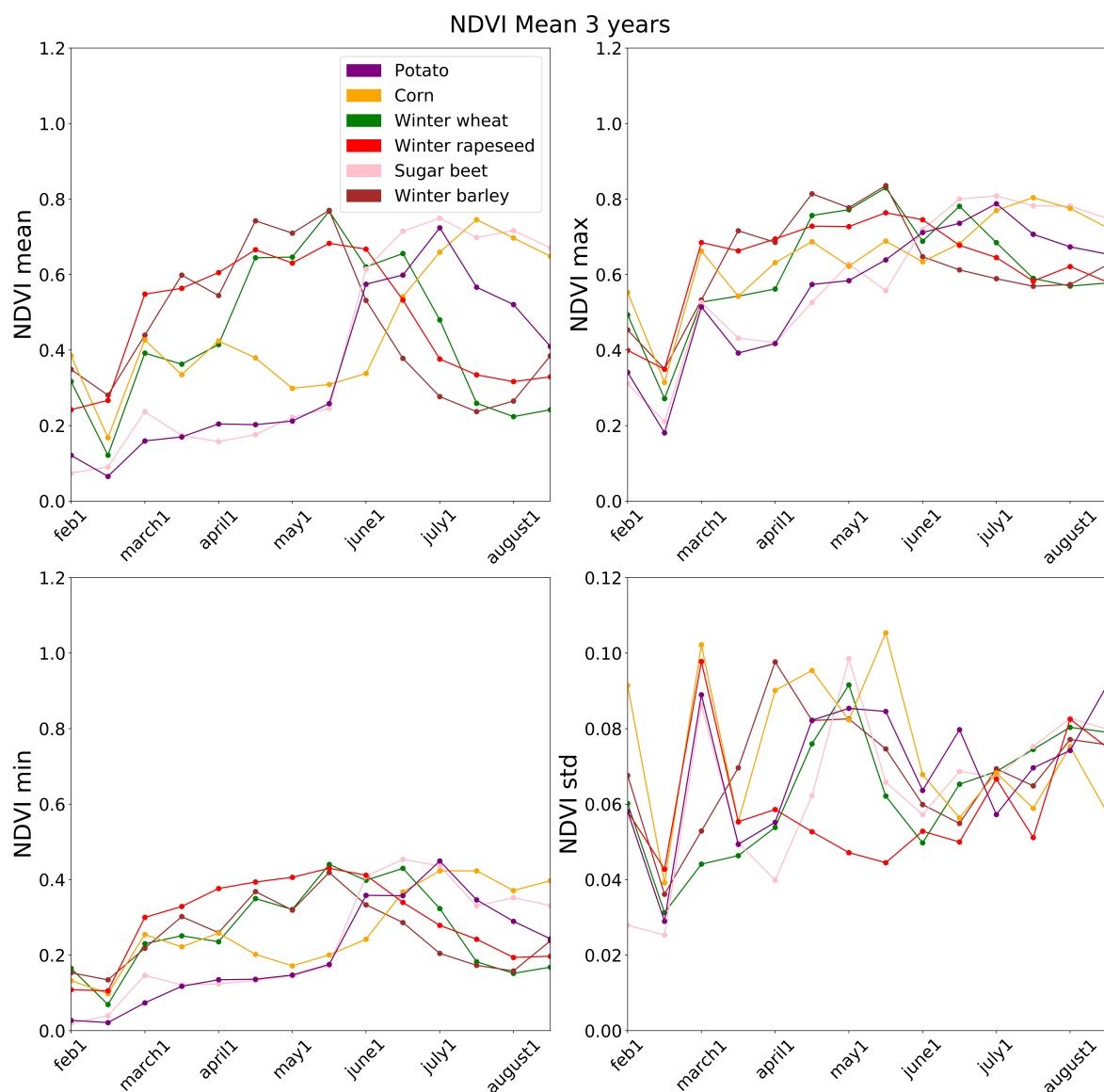
**Figure 3.** Temporal patterns of mean NDVI, min NDVI, max NDVI and standard deviation NDVI values that vary with time.

*2.3. Temporal Interpolation*

Another important task of the study is temporal interpolation. Working with time series entails several problems, such as gaps, different time steps, and noisy measurements. Lepot et all. [45] reviewed interpolation methods, such as nearest-neighbour and linear interpolation. Sentinel images are available every 3 to 5 days for our study sites but cloud coverage reduces the amount of suitable data, which needs to be re-sampled and interpolated. To cover the phenological development, we decided upon a 2-weeks resolution of our time series data. Polynomial interpolation performs slightly better, but does not approximate the real reflectance values when satellite data are not available for the beginning of the vegetation period. Linear interpolation shows stable performance in all the cases and is, therefore, used.

*2.4. Metrics*

The performance evaluation of crop-type mapping is based on overall accuracy (OA), producer's accuracy (PA), user's accuracy (UA) and Kappa statistics [46]. The PA and UA values are presented for each crop type while the OA and Kappa values evaluate the global performance of our trained model. All the metrics are part of scikit-learn [47]. The OA represents the number of correctly classified crops divided by all the reference crops. The Kappa coefficient is a measure in the range 0–1 and quantifies prediction and ground truth. For instance, a value of 1 indicates a perfect match between the prediction and ground truth. The PA shows how well a classification matches with the ground truth while the UA refers to the reliability of the classification.

*2.5. Crop-type processing*

Figure 4 illustrates the general processing chain with its data inputs, pre-processing, training, and final classification. OpenStreetMap was used to remove irrelevant information, such as roads, water bodies, and forests. Furthermore, we applied the standard deviation of NDVI as a filter to remove pixels, which did not change over time. An agricultural area usually changes during a vegetation period. The training and testing were carried out with official field boundaries and crop types from StMELF. The SVM and RF were trained with mean values (NDVI/raw bands) and tested with an independent data set [48–50]. Another experiment focused on the prediction for the year 2018 using a model trained with data from 2016 and 2017. This represents a challenging step, mainly because of the target year being divergent.
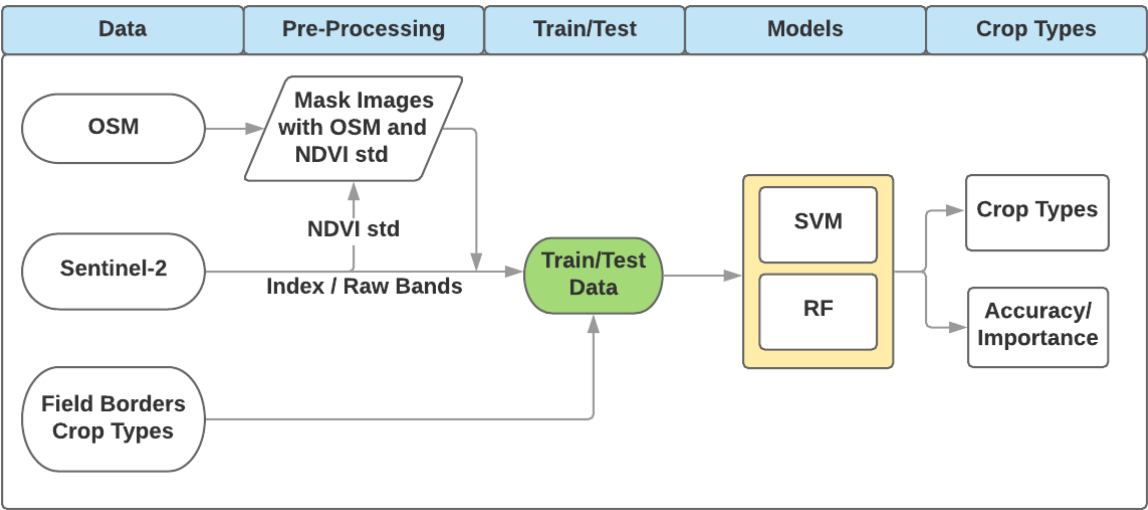
**Figure 4.** Overview of the processing steps with their input data, pre-processing and models.

## 3. Results

### 3.1. Crop-type classification with SVM and RF

We compared several interpolation methods in combination with SVMs and RF. Cleaning and interpolating the data set improves the performance of a classifier. We decided upon linear interpolation to fill in the missing values and a temporal resolution of 2 weeks from February to August. Our SVM and RF models were trained with NDVI and all the bands for 2016, 2017, and 2018. NDVI is a good proxy but the usage of all bands outperforms an index-based approach in our experiments. The verification is based on a random train/test split with a test size of 25% and 5 splits for K-Fold cross validation. The entire data set consists of 2099 samples, including 699 samples for 2016, 700 for 2017, and 700 for 2018. The additional coherent test site is visualised in Figure 6. The yearly crop data set is uniformly distributed and includes around 100 samples for every class. RF was parametrised with 1000 decision trees in the forest and SVM applied 10 for the "C" parameter and a linear kernel. Table 3 compares the performance of the classifiers using NDVI and all the bands. It clearly simplifies the collected results and shows that RF performs better in all cases, which is why only the confusion matrices for RF have been presented in Table 4 and Table 5. Nevertheless, SVM also classified the crop types very well. In general, the use of all bands improves the OA for SVM and RF. Table 4 summarises the results of RF using NDVI while Table 5 lists the performance of RF with raw bands and without the rejection class. The greatest influence on accuracy is provided by the raw bands. This is important because they cover a wider spectral range than an index alone. Thus, possible important insights are not lost. Table 5 confirms the performance without the rejection class.

**Table 3.** Overview of the results with SVM and RF and various features.

|       | NDVI | Raw bands | Bands without "Other" |
|-------|------|-----------|-----------------------|
| SVM:  | 70%  | 87%       | 90%                   |
| RF:   | 75%  | 88%       | 92%                   |

The performance of this model is very promising and is analysed in terms of feature importance (Figure 5). It can be seen that bands B3, B5, B6, B7, B8, B8A, B11 and B12 are the most important

features, and mainly from May to August. The results could be increased to 94% using hyperparameter optimisation, but the optimised values were omitted for this publication to keep the configuration simpler. On the one hand, the rejection class worsens the results, on the other hand, it includes all agricultural land. The class "Other" includes various types of crops, such as different types of grasslands. It introduced an error, especially in regions with crops not covered in our model. The best model was additionally applied to the coherent test site of 301 fields and for the year 2018. When training the classifier, this area was not included, which allowed us to test the performance of our model in a real scenario. Figure 6 shows the recorded results from StMELF which act as ground truth data, followed by the classification results. The OA of 96% was achieved without the rejection class and it can be seen that the classifier had problems to distinguish barley and rapeseed. This was a bit surprising since rapeseed is clearly different from other crops in spring.
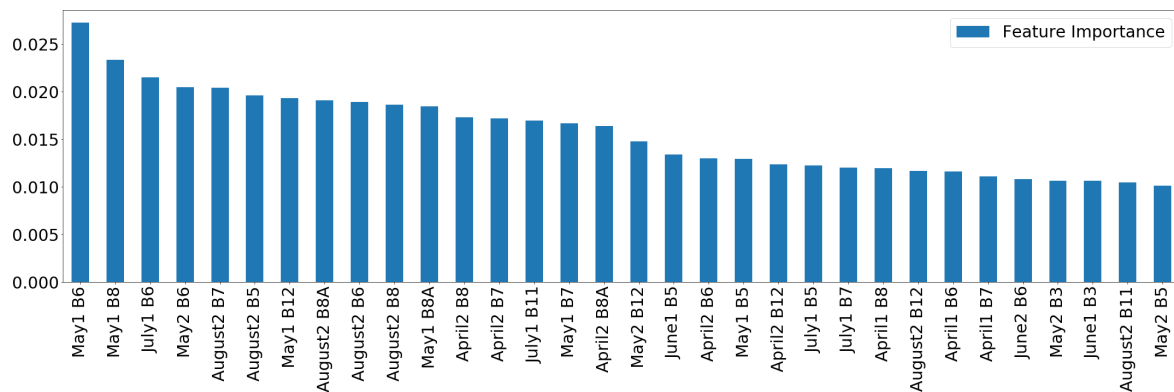
**Figure 5.** Sorted features of model for 2016-2018 and without class "Other"

**Table 4.** Confusion matrix with NDVI and RF.

|  | Other | Winter wheat (WW) | Winter barley (WB) | Winter rapeseed (WR) | Maize | Potato | Sugar beet (SB) | Total | UA (%) | PA (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| **Other** | 54 | 0 | 0 | 5 | 2 | 1 | 0 | 62 | 87.1 | 69.23 |
| **WW** | 4 | 54 | 4 | 5 | 1 | 0 | 0 | 68 | 79.41 | 72.97 |
| **WB** | 4 | 7 | 67 | 8 | 0 | 0 | 0 | 86 | 77.91 | 81.71 |
| **WR** | 8 | 12 | 9 | 53 | 1 | 0 | 0 | 83 | 63.86 | 73.61 |
| **Maize** | 6 | 1 | 2 | 1 | 58 | 10 | 11 | 89 | 65.17 | 79.45 |
| **Potato** | 2 | 0 | 0 | 0 | 3 | 49 | 6 | 60 | 81.67 | 69.01 |
| **SB** | 0 | 0 | 0 | 0 | 8 | 11 | 58 | 77 | 75.32 | 77.33 |
| **Total** | 78 | 74 | 82 | 72 | 73 | 71 | 75 | 525 |  |  |
|  |  |  |  |  |  |  |  |  | **OA** | 75% |
|  |  |  |  |  |  |  |  |  | **Kappa** | 0.86 |
|  |  |  |  |  |  |  |  |  | **CV** | 0.75 |

**Table 5.** Confusion matrix with RF, all raw bands and without the rejection class "Other".

| | Winter wheat (WW) | Winter barley (WB) | Winter rapeseed (WR) | Corn | Potato | Sugar beet (SB) | Total | UA (%) | PA (%) |
|---|---|---|---|---|---|---|---|---|---|
| **WW** | 62 | 3 | 3 | 1 | 0 | 1 | 70 | 88.57 | 88.57 |
| **WB** | 7 | 63 | 4 | 1 | 0 | 0 | 75 | 84 | 92.65 |
| **WR** | 0 | 2 | 78 | 1 | 0 | 0 | 81 | 96.3 | 91.76 |
| **Corn** | 1 | 0 | 0 | 76 | 3 | 0 | 80 | 95 | 93.83 |
| **Potato** | 0 | 0 | 0 | 1 | 58 | 1 | 60 | 96.67 | 87.88 |
| **SB** | 0 | 0 | 0 | 1 | 5 | 78 | 84 | 92.86 | 97.5 |
| **Total** | 70 | 68 | 85 | 81 | 66 | 80 | 450 | | |
| | | | | | | | | **OA** | 92% |
| | | | | | | | | **Kappa** | 0.97 |
| | | | | | | | | **CV** | 0.91 |
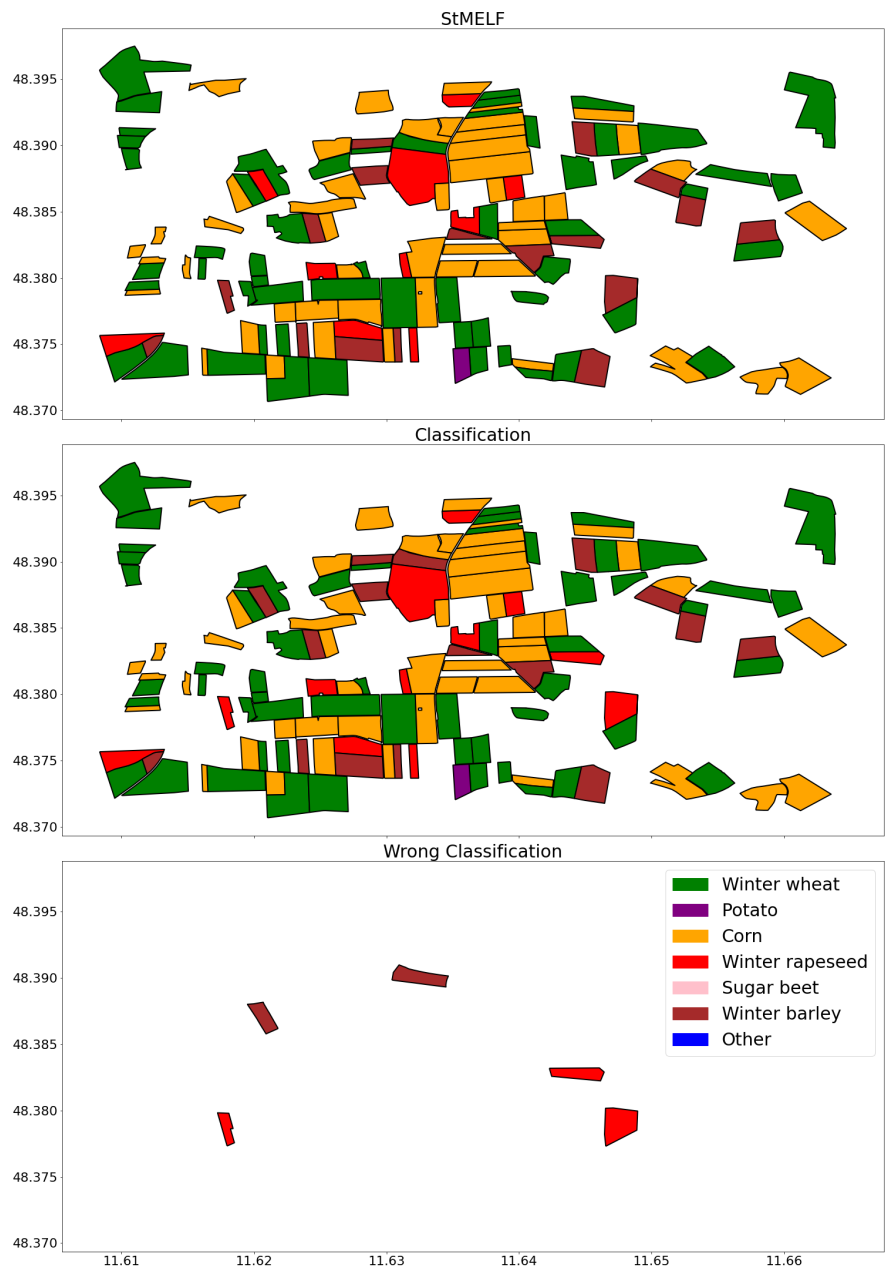
**Figure 6.** Crop classification based on RF, all bands and without the class 'Other'. The first image shows the recorded crop types (ground truth data) while the second image shows the classified crop types. The last image visualises incorrect classifications.

### 3.2. Prediction in an unknown year with anomalies

The prediction for an unknown year was analysed to develop a system for detecting crop types without labels for the year considered. Especially in times of climate change, this experiment is important to assess to what extent drought or unusual growing conditions affect the performance of classifiers. Figure 7 shows the raw reflectance values for band B6 and the crop types considered in our study. The reflectance values are influenced by weather conditions and the growth progress. Nevertheless, the trend remains comparable across years, and a separation between crops is given as well. We trained a model with data from 2016 and 2017 and predict crop types for 2018 using corresponding satellite observations. The OA was 79% with and 80% without the rejection class "Other" (see Table 6 and 7), which is very accurate considering that 2018 was not comparable to previous years. Applying the classifiers to the test area in Figure 8 reached an accuracy of 75% and 86% without the "Other" class, respectively. The results are even comparable to those of the models trained with data for 2016–2018. The temporal patterns are preserved even in an unusual year, which is why the classification performs well. To get a more detailed insight, the dependence of the accuracy on the classification time was investigated in Figure 9. Here, we see that the PA for winter rapeseed is already very high in spring, which is related to the clear light green color. At the end of July and in August, the accuracy for all other crop types converges. This is also visible in the overall accuracy where from the end of July the maximum is almost reached.
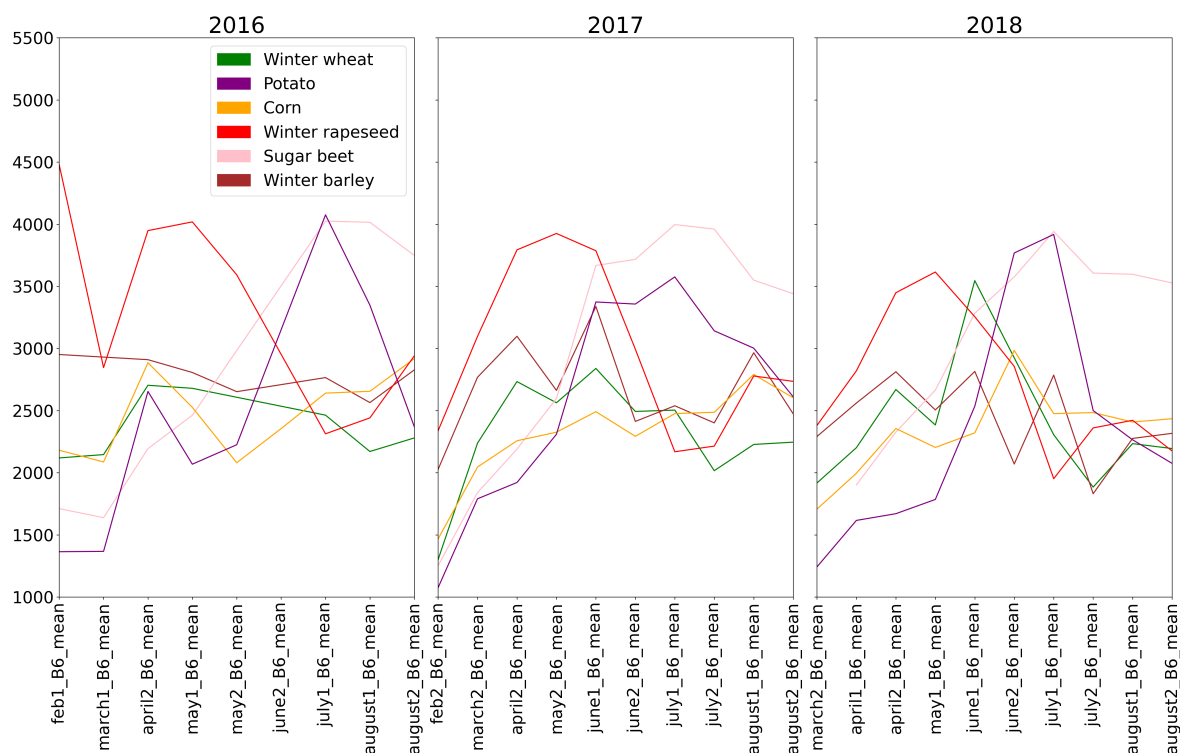
**Figure 7.** Comparison of B6 mean reflectance across years.

**Table 6.** Confusion matrix for RF with all bands and class "Other" based on a model trained with data for 2016 and 2017. The crop types were determined for 2018.

| | Other | Winter wheat (WW) | Winter barley (WB) | Winter rapeseed (WR) | Corn | Potato | Sugar beet (SB) | Total | UA (%) | PA (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| **Other** | 78 | 9 | 16 | 1 | 4 | 0 | 0 | 108 | 72.22 | 78 |
| **WW** | 8 | 83 | 11 | 3 | 0 | 0 | 0 | 105 | 79.05 | 83 |
| **WB** | 3 | 4 | 67 | 1 | 0 | 0 | 0 | 75 | 89.33 | 67 |
| **WR** | 1 | 3 | 5 | 95 | 0 | 0 | 0 | 104 | 91.35 | 95 |
| **Corn** | 3 | 0 | 1 | 0 | 67 | 1 | 0 | 72 | 93.06 | 67 |
| **Potato** | 7 | 1 | 0 | 0 | 28 | 99 | 34 | 169 | 58.58 | 99 |
| **SB** | 0 | 0 | 0 | 0 | 1 | 0 | 66 | 67 | 98.51 | 66 |
| **Total** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 700 | | |
| | | | | | | | | | **OA** | 79% |
| | | | | | | | | | **Kappa** | 0.91 |

**Table 7.** Confusion matrix for RF with all bands and without class "Other" based on a model trained with data for 2016 and 2017. The crop types were determined for 2018.

| | Winter wheat (WW) | Winter barley (WB) | Winter rapeseed (WR) | Corn | Potato | Sugar beet (SB) | Total | UA (%) | PA (%) |
|---|---|---|---|---|---|---|---|---|---|
| **WW** | 85 | 20 | 5 | 1 | 0 | 0 | 111 | 76.58 | 85 |
| **WB** | 10 | 71 | 2 | 1 | 0 | 0 | 84 | 84.52 | 71 |
| **WR** | 3 | 5 | 93 | 0 | 0 | 0 | 101 | 92.08 | 93 |
| **Corn** | 0 | 2 | 0 | 69 | 1 | 0 | 72 | 95.83 | 69 |
| **Potato** | 2 | 2 | 0 | 28 | 99 | 34 | 165 | 60 | 99 |
| **SB** | 0 | 0 | 0 | 1 | 0 | 66 | 67 | 98.51 | 66 |
| **Total** | 100 | 100 | 100 | 100 | 100 | 100 | 600 | | |
| | | | | | | | | **OA** | 80% |
| | | | | | | | | **Kappa** | 0.94 |

**Figure 8.** Prediction in 2018 using a model trained with data for 2016 and 2017 and without class 'Other'. The model achieved an OA of 86%.
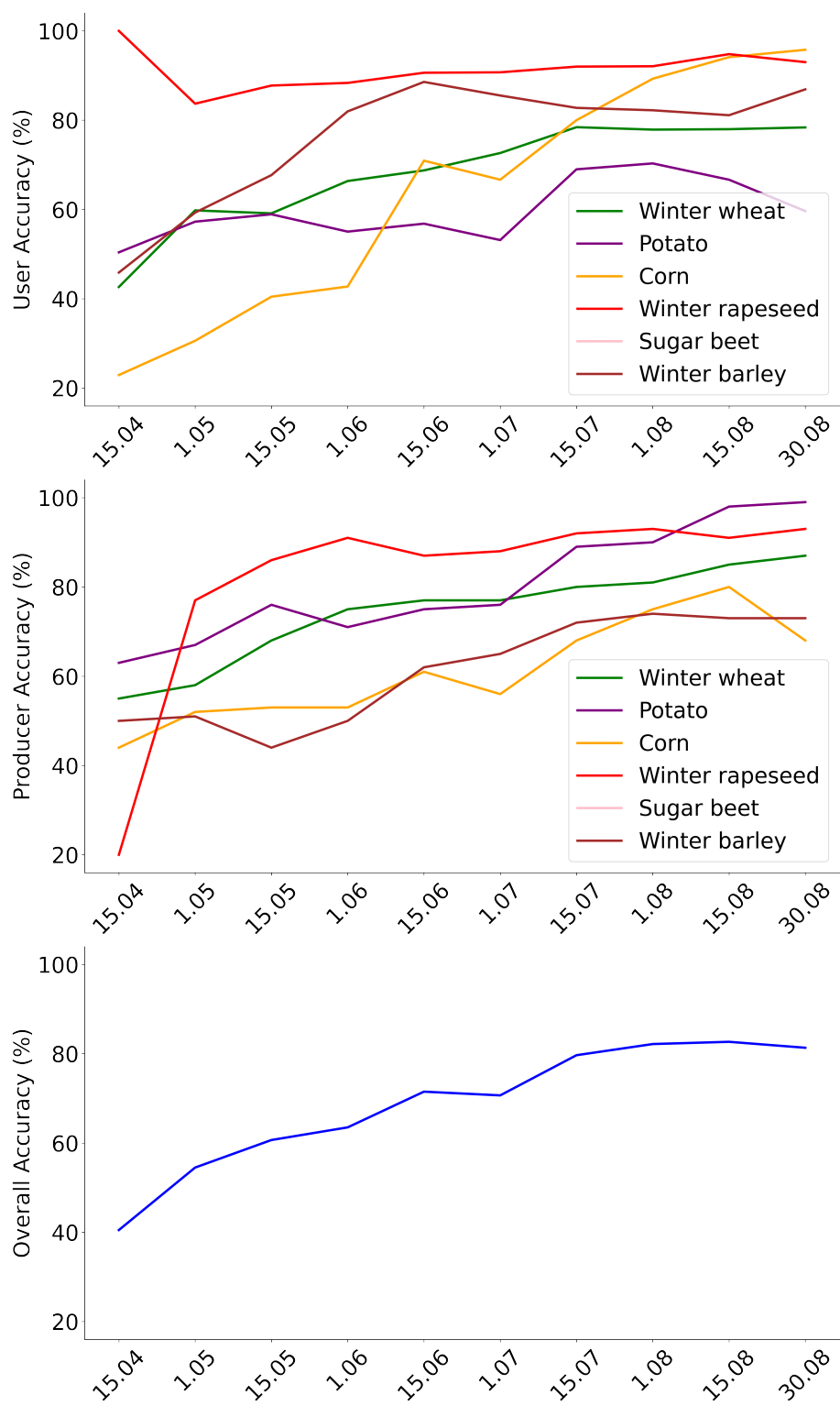
**Figure 9.** Prediction of crop types with model 2016/2017 as a function of time.

## 4. Discussion

An important factor is pre-processing, which includes temporal interpolation and data cleaning. While linear interpolation is the right choice in many applications, polynomial interpolation approximates a time series more accurately, but introduces errors, especially for data sets with gaps at the beginning of the growing season, which is why linear interpolation was used in this study. Two weeks temporal resolution proved to be appropriate as it adequately matched the phenology and was supported by sufficient data in the test region. Single gaps could be filled by linear interpolation. For future works, it would be interesting to extend the observation period, as, in this study, Sentinel-2 observations were used from February to the end of August. This extension would cover the sowing and harvesting periods for all crop types as indicated in Figure 1). In our investigations, a small amount of clouds was acceptable because Random Forest can handle a certain amount of noise [51,52]. Nevertheless, a completely cloud-free time series for each field could minimally improve the results, but at the cost of run time, since each field time series must be processed with respect to cloudy pixels or shadows.

Analysis of the raw bands and comparison with the NDVI showed that the raw bands exerted a significant influence on the accuracy of a classifier. In this work, we chose the NDVI because other indices such as the NDRE did not result in better separation of crops (unpublished results). A specific objective of this work was to compare the index-based classification for crop mapping with the raw band-based classification. The use of all bands exploited additional information that improved differentiation at specific phenological stages. In general, previous multi-temporal approaches point to the advantages of time for crop differentiation but at the same time use an index by which this advantage is partially lost. For example, winter rapeseed fields can be clearly distinguished in spring with RGB bands, which becomes more difficult with a biomass index. Analysis of predictability as a function of time showed that the spring season is important for winter rapeseed. This could avoid misclassification of winter barley as winter rapeseed. Such an advantage of RGB information was explicitly observed for winter rapeseed, but in general the information from the B5, B6, B7, B8, B11, and B12 bands was significant.

The rejection class "Other" for further plant types, such as grassland, reduced the classification accuracies obtained with SVM and RF. Results described in section 3.1 showed that training a classifier with as many crop types as possible and avoiding the use of a rejection class increased the accuracy. However, from a practical point of view, this result was important because sometimes a model does not know all the crop types in a target region. In this context, transfer learning offers a possibility to extend an existing model by further crop types.

This study assessed how well crop types can be classified in a year with widely differing weather conditions, since the weather in 2018 was hot and dry. Therefore, an experiment was conducted to classify crop types in 2018 using a model trained with 2016 and 2017 data. Despite reflectance changes, the trained model was not inordinately affected by this anomaly and predicted good results. The temporal patterns of all crop types in 2018 thus remained comparable and could be distinguished. The early detection of the crop types as outlined in section 3.2 showed that maximal classification accuracy was reached from July onward, which is also consistent with observations in [33]. This emphasises that a model based on training data from average years can also be used for future years without labels for the target year. This experiments were also important because many previous studies are based on a train/test split of all collected data, which accordingly results in high accuracies. We hypothesise that climatic change in our test region will not affect the crop-type classification strongly as long as good water availability can compensate the effects of higher temperatures.

Overall, the results provided very good classification accuracies. The models explained the variance in our data set very well. Referring to the best results in Table 5 where the model was trained with data from all years, it can be seen that sugar beet, winter rapeseed and corn all achieve UA's and PA's above 90%. Winter wheat, winter barley and winter rapeseed are occasionally interchanged, which degrades the accuracies to some degree. Similarly, potatoes are sometimes confused with sugar

beet and corn. We assume that due to the slight noise some time series become too similar although RF should deal with a small noise ratio. Special attention should be paid to winter wheat, winter barley and potatoes, as these have the highest potential for improvement in terms of accuracy. As a further extension of this work, it seems possible to apply a simplified approach to identify winter rapeseed based on spring color changes alone. As suggested in other research, radar-based information could also be used, which is also available from Sentinel-1. Retrieval of cloud-independent data may be beneficial in regions of high cloud cover, such as the tropics. However, Sentinel-2 data are preferable if they are available on a regular basis with at least one image every two weeks. For further quantification using other methods, our data set was partially made available to [2].

## 5. Conclusion

This study addressed several aspects of crop-type mapping. We used RF and SVM with their underlying configurations and achieved up to 92% overall accuracies. Using our approach, we classified seven crop classes with very high precision. Random Forest slightly outperformed SVMs with input features being critical for accuracy. Especially, the use of all raw bands uncovered additional information that could not be detected with a single index, such as NDVI, alone. In this context, we have shown that an index is not always useful. The use of a multi-temporal approach improved classification accuracies, but data gaps must be avoided, which was realised with a temporal resolution of two weeks and linear interpolation. Prediction of crop types for a year without labels and with extraordinary weather conditions showed that the spectral signature was still intact and crop types could be well determined at our study site. The accuracy of the prediction, depending on the time, increased steadily with the number of observations and reached a maximum from July onwards. An exception was winter rapeseed, which was already classified very well in spring. All other crops needed observations until July to reach the optimal classification level.

**Author Contributions:** The authors made the following contributions: **Michael Marszalek**: Conceptualisation, Methodology, Software, Validation, Writing - Original Draft, Review, Editing. **Maximilian Lösch**: Conceptualisation, Methodology, Software, Validation, Writing, Review, Editing. **Marco Körner**: Writing, Review, Editing, Supervision. **Urs Schmidhalter**: Writing, Review, Editing, Supervision, Project administration, Funding acquisition.

**Funding:** The research work was supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support program for the project 28-1-B3.030-16.

**Acknowledgments:** StMELF, the Bavarian State Ministry of Agriculture and Forestry supported this research with its crop type data set. The authors are thankful for this data, which was a key in the scope of this research. We also like to thank all colleagues, which contributed with valuable recommendations and improvements on this manuscript. Finally, we would like to acknowledge the work of ESA regarding its Sentinel mission, which will enable new applications.

**Conflicts of Interest:** The authors declare no conflict of interest. The funding organisation had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results. We requested a permission for the crop type data, which covers Bavaria. StMELF provides this data on request for selected research projects.

## References

1. Friedl, M.; Sulla-Menashe, D. MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC, 2019. [Online; accessed 31-August-2020].
2. Marszalek, M. CropTypes — Github repository, 2020. [Online; accessed 12-Mai-2020].
3. Wardlow, B.; Egbert, S. A comparison of MODIS 250-m EVI and NDVI data for crop mapping: a case study for southwest Kansas. *International Journal of Remote Sensing* **2010**, *31*, 805–830. doi:10.1080/01431160902897858.
4. Arvor, D.; Jonathan, M.; Simoes, M.; Dubreuil, V.; Durieux, L. Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. *International Journal of Remote Sensing* **2011**, *32*, 7847–7871. doi:10.1080/01431161.2010.531783.

5.    Tricht, K.; Gobin, A.; Gilliams, S.; Piccard, I.  Synergistic Use of Radar Sentinel-1 and Optical Sentinel-2 Imagery for Crop Mapping: A Case Study for Belgium.  *Remote Sensing* **2018**, *10*, 1642. doi:10.3390/rs10101642.

6.    Inglada, J.; vincent, a.; Arias, M.; Sicre, C.  Improved Early Crop Type Identification By Joint Use of High Temporal Resolution SAR And Optical Image Time Series.  *Remote Sensing* **2016**, *8*, 362. doi:10.3390/rs8050362.

7.    Saini, R.; Ghosh, S.  Crop Classification on Single Date Sentinel-2 Imagery using Random Forest and Support Vector Machine.  *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci* **2018**, pp. –5, 683–688. doi:10.5194/isprs-archives-XLII-5-683-2018.

8.    Immitzer, M.; Vuolo, F.; Atzberger, C.  First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sensing* **2016**, *8*, 166.  doi:10.3390/rs8030166.

9.    Natteshan, N.; Kumar, N. Effective SAR image segmentation and classification of crop areas using MRG and CDNN techniques. *European Journal of Remote Sensing* **2020**, pp. 1–15.  doi:10.1080/22797254.2020.1727777.

10.   Mazzia, V.; Khaliq, A.; Chiaberge, M.  Improvement in Land Cover and Crop Classification based on Temporal Features Learning from Sentinel-2 Data Using Recurrent-Convolutional Neural Network (R-CNN). *Applied Sciences* **2019**, *10*, 238.  doi:10.3390/app10010238.

11.   Chetan, M.; Dornik, A.; Urdea, P. Comparison of Object and Pixel-based Land Cover Classification through three Supervised Methods. *Zeitschrift für Geodäsie, Geoinformation und Landmanagement* **2017**, *5*, 265–270. doi:10.12902/zfv-0165-2017.

12.   Böhler, J.; Schaepman, M.; Kneubühler, M.  Crop Classification in a Heterogeneous Arable Landscape Using Uncalibrated UAV Data. *Remote Sensing* **2018**, *10*, 1282.  doi:10.3390/rs10081282.

13.   Wei, L.; Yu, M.; Liang, Y.; Yuan, Z.; Huang, C.; Li, R.; Yu, Y.  Precise Crop Classification Using Spectral-Spatial-Location Fusion Based on Conditional Random Fields for UAV-Borne Hyperspectral Remote Sensing Imagery. *Remote Sensing* **2019**, *11*.  doi:10.3390/rs11172011.

14.   Harfenmeister, K.; Spengler, D.; Itzerott, S.  A Progressive Crop-Type Classification Using Multitemporal Remote Sensing Data and Phenological Information. *PFG - Journal of Photogrammetry, Remote Sensing and Geoinformation Science* **2018**, *86*.  doi:10.1007/s41064-018-0050-7.

15.   Vuolo, F.; Neuwirth, M.; Immitzer, M.; Atzberger, C.; Ng, W.T. How much does multi-temporal Sentinel-2 data improve crop type classification? *International Journal of Applied Earth Observation and Geoinformation* **2018**, *72*, 122–130.  doi:10.1016/j.jag.2018.06.007.

16.   Marszalek, M.; Lösch, M.; Körner, M.; Schmidhalter, U.  Multi-temporal Crop Type and Field Boundary Classification with Google Earth Engine.  *Preprints* **2020**.  [PREPRINT], doi:10.20944/preprints202004.0316.v1.

17.   Belgiu, M.; Csillik, O.  Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis.  *Remote Sensing of Environment* **2018**, *204*, 509–523. doi:10.1016/j.rse.2017.10.005.

18.   Dimitrov, P.; Dong, Q.; Eerens, H.; Gikov, A.; Filchev, L.; Roumenina, E.; Jelev, G. Sub-Pixel Crop Type Classification Using PROBA-V 100 m NDVI Time Series and Reference Data from Sentinel-2 Classifications. *Remote Sensing* **2019**, *11*.  doi:https://doi.org/10.3390/rs11111370.

19.   Rußwurm, M.; Körner, M. Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders. *IJGI* **2018**, *7*, 129.  doi:10.3390/ijgi7040129.

20.   Rußwurm, M.; Lefévre, S.; Körner, M.  BreizhCrops: A Satellite Time Series Dataset for Crop Type Identification. *Remote Sensing and Spatial Information Sciences ISPRS* **2020**.

21.   Ma, L.; Li, M.; Ma, X.; Cheng, L.; Du, P.; Liu, Y.  A review of supervised object-based land-cover image classification.  *ISPRS Journal of Photogrammetry and Remote Sensing* **2017**, *130*, 277 − 293. doi:https://doi.org/10.1016/j.isprsjprs.2017.06.001.

22.   Pandya, M.; Baxi, A.; Potdar, M.; Kalubarme, M.; Agrawal, B.  Comparison of Various Classification Techniques for Satellite Data. *International Journal Of Scientific & Engineering Research* **2013**, *4*, 2013–1.

23.   Nitze, I.; Schulthess, U.; Asche, H. Comparison of Machine Learning Algorithms, Random Forest, Artificial Neural Network And Support Vector Machine To Maximum Likelihood For Supervised Crop Type Classification. GEOBIA; , 2012.

24.   Hao, P.; Wang, L.; Zhan, Y.; Niu, Z.  Using Moderate-Resolution Temporal NDVI Profiles for High-Resolution Crop Mapping in Years of Absent Ground Reference Data: A Case Study of Bole

and Manas Counties in Xinjiang, China. *ISPRS International Journal of Geo-Information* **2016**, *5*, 67. doi:10.3390/ijgi5050067.

25. Zheng, B.; Myint, S.; Thenkabail, P.; Aggarwal, R. A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. *International Journal of Applied Earth Observation and Geoinformation* **2015**, *34*, 103–112. doi:10.1016/j.jag.2014.07.002.

26. Sousa, V.; Salami, G.; Silva, M.; Silva, E.; Júnior, J.; Alba, E. Methodological evaluation of vegetation indexes in land use and land cover (LULC) classification. *Geology, Ecology, and Landscapes* **2019**. doi:10.1080/24749508.2019.1608409.

27. Blaschke, T. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* **2010**, *65*, 2–16. doi:10.1016/j.isprsjprs.2009.06.004.

28. Blaschke, T.; Hay, G.J.; Kelly, M.; Lang, S.; Hofmann, P.; Addink, E.; Queiroz Feitosa, R.; van der Meer, F.; van der Werff, H.; van Coillie, F.; Tiede, D. Geographic Object-Based Image Analysis – Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing* **2014**, *87*, 180 – 191. doi:https://doi.org/10.1016/j.isprsjprs.2013.09.014.

29. Petitjean, F.; Inglada, J.; Gancarski, P. Satellite Image Time Series Analysis Under Time Warping. *IEEE Transactions on Geoscience and Remote Sensing* **2012**, *50*, 3081–3095. doi:10.1109/TGRS.2011.2179050.

30. Maus, V.; Câmara, G.; Souza, R.; Ipia, A.; Ramos, F.; Queiroz, G. A Time-Weighted Dynamic Time Warping Method for Land-Use and Land-Cover Mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2016**, *9*, 3729–3739. doi:10.1109/JSTARS.2016.2517118.

31. yu HAO, P.; jun TANG, H.; xin CHEN, Z.; yan MENG, Q.; peng KANG, Y. Early-season crop type mapping using 30-m reference time series. *Journal of Integrative Agriculture* **2020**, *19*, 1897 – 1911. doi:https://doi.org/10.1016/S2095-3119(19)62812-1.

32. Mori, U.; Mendiburu, A.; Dasgupta, S.; Lozano, J.A. Early Classification of Time Series by Simultaneously Optimizing the Accuracy and Earliness. *IEEE Transactions on Neural Networks and Learning Systems* **2018**, *29*, 4569–4578.

33. Rußwurm, M.; Tavenard, R.; Lefévre, S.; Körner, M. Early Classification for Agricultural Monitoring from Satellite Time Series. International Conference on Machine Learning AI for Social Good Workshop, 2019.

34. Aneece, I.; Thenkabail, P. Accuracies Achieved in Classifying Five Leading World Crop Types and their Growth Stages Using Optimal Earth Observing-1 Hyperion Hyperspectral Narrowbands on Google Earth Engine. *Remote Sensing* **2018**, *10*. doi:10.3390/rs10122027.

35. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* **2017**, *202*, 18–27. doi:10.1016/j.rse.2017.06.031.

36. Mahdianpari, M.; Salehi, B.; Mohammadimanesh, F.; Homayouni, S.; Gill, E. The First Wetland Inventory Map of Newfoundland at a Spatial Resolution of 10 m Using Sentinel-1 and Sentinel-2 Data on the Google Earth Engine Cloud Computing Platform. *Remote Sensing* **2019**, *11, 43*. doi:10.3390/rs11010043.

37. LfStat. Statistische Berichte - A5113 201700: Flächenerhebung. In *Art d. tatsächlichen Nutzung in Bayern zum Stichtag 31*; Bayerisches Landesamt für Statistik (LfStat), 2017.

38. Allan, R.; Pereira, L.; Smith, M. *Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56*; Vol. 56, FAO, 1998.

39. FAO. Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56, 2020. [Online; accessed 23-July-2020].

40. GEE. A planetary-scale platform for Earth science data & analysis — Google Earth Engine, 2020. [Online; accessed 21-July-2020].

41. GEE. Sentinel-2 MSI: MultiSpectral Instrument, Level-1C — Earth Engine Data Catalog, 2012. [Online; accessed 21-July-2020].

42. Main-Knorn, M.; Pflug, B.; Louis, J.; Debaecker, V.; Müller-Wilm, U.; Gascon, F. Sen2Cor for Sentinel-2. Proc. SPIE 10427, Image and Signal Processing for Remote Sensing XXIII, 2017. doi:10.1117/12.2278218.

43. Sola, I.; Alvarez-Mozos, J.; Gonzalez-Audicana, M. Inter-Comparison of Atmospheric Correction Methods on Sentinel-2 Images Applied to Croplands. Proc. IEEE International Geoscience & Remote Sensing Symposium, 2018; , 2018. doi:10.1109/IGARSS.2018.8518890.

44. Doxani, G.; Vermote, E.; Roger, J.C.; Gascon, F.; Adriaensen, S.; Frantz, D.; Hagolle, O.; Hollstein, A.; Kirches, G.; Li, F. Atmospheric Correction Inter-Comparison Exercise. *Remote Sensing* **2018**, *10*, 352. doi:10.3390/rs10020352.

45. Lepot, M.; Aubin, J.B.; Clemens, F. Interpolation in Time Series: An Introductive Overview of Existing Methods, Their Performance Criteria and Uncertainty Assessment. *Water* **2017**, *9*, 796. doi:10.3390/w9100796.

46. Humbold State University. GSP 2016: Introduction to Remote Sensing, 2020. [Online; accessed 25-January-2020].

47. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E.; Louppe, G. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2012**, *12*.

48. Cortes, C.; Vapnik, V. Support vector machines. *Machine Learning* **1995**, *20*, 273–293.

49. Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*; Cambridge University Press, 2000. doi:https://doi.org/10.1017/CBO9780511801389.

50. Biau, G.; Scornet, E. A random forest guided tour. *TEST* **2016**, *25*, 197–227. doi:10.1007/s11749-016-0481-7.

51. Agjee, N.; Mutanga, O.; Peerbhay, K.; Ismail, R. The Impact of Simulated Spectral Noise on Random Forest and Oblique Random Forest Classification Performance. *Journal of Spectroscopy* **2018**, *2018*, 1–8. doi:10.1155/2018/8316918.

52. Pelletier, C.; Valero, S.; Inglada, J.; Champion, N.; Sicre, C.; Dedieu, G. Effect of Training Class Label Noise on Classification Performances for Land Cover Mapping with Satellite Image Time Series. *Remote Sensing* **2017**, *9*, 173. doi:10.3390/rs9020173.