Data Interoperability Pilot:

Analyzing the usability of generalized GSAA data in context of crop classification

under EXPERT CONTRACT NUMBER - CT-EX2018D319529-105

Final report

Created by Bernadett Csonka – February - April 2021.

Content:

1 The purpose of the classification pilot

The main target of this pilot study is **to analyze the usability of Geospatial Aid Application (GSAA) data in the context of crop type classification** on pilot sites, and **to assess land cover / land use nomenclatures regarding their fitness for an interoperable exchange of georeferenced crop information**.

- The pilot study was implemented along the following principles: 2 pilot sites were implemented with the use of open available GSAA data of y2020
- semantic harmonization, thematic generalization and geometrical generalization of GSAA data
- testing the performance as reference data (training/test) to derive a seamless crop map:
  - **2 versions of grouping crops: detailed crop types / main crop groups**
  - **3 versions of geometry: polygon / random points / inner centroid**
  - **3 types of supervised classification method**
- comparing the result of different geometric representation of training data using the misclassification matrix and the accuracy values related to crop types
- methods were defined respecting possibilities of automatization,
- LPIS was also used for masking eligible area.

Since the seamless geometry and the crop type of agricultural parcels is available in the Integrated Administration and Control System (IACS) as the result of multi annual GSAAs, this information can be used for modeling, analyzing and planning land use. On the other hand, the high sensitivity of GSAA data prevents member states to share original geometries real time due to privacy or economic considerations. The legal aspects of data security level is still an on-going debate during the implementation of the pilot. This pilot would like to offer alternative solutions to derive geometrically generalized data from GSA, which can be suitable for training effectively image classification algorithms.

Several approaches exist to detect the annual crops based on semi-automatic or automatic classification of multi-temporal satellite images. To assess the potential applications of a crop map based on GSAA parcel data, a classification exercise has been run on two pilot study areas. Different semantic, thematic and geometric generalization approaches have been tested in combination with classification algorithms.

The case with IACS-GSAA data is, however, specific: high amounts of data (i.e. declared parcels) are available. In several Member States (MS) over 90% of the Arable Land (AL) is declared. However, the quality of these declaration is only partially assured by the On The-Spot Checks (OTSC), which amounts to a 5-8% of the total area. Therefore, the quality of the GSAA dataset and approaches based on ML coupled with large training data sets have to be explored to establish an operational methodology.

Another important aspect to investigate is the possibility to generate crop maps using GSAA parcels data. Currently, the EU MSs are developing detailed methodologies for detecting land management actions and the growth of crops by using satellite images for Checks by Monitoring (CbM). It is foreseen that the result of the “marker approach” will lead to a more in-depth validation of GSAA parcels against remote sensing data. The “marker approach” is a systematic monitoring of crop development and agricultural practices of a parcel, based on the combination of image signal time series and predefined values and limits describing the land phenomenon. Therefore, the quality of data to train models for crop mapping will highly increase. On the other hand, crop type maps, are not a direct output of the marker-based approach, they are outside the scope of the CbM. The methodology of CbM does not require generating a seamless land use thematic map for the entire country.

There are several directions of spatial data modelling, where a seamless cropmap brings fundamental values, such as modelling ecosystem services contributing to biodiversity, ensuring food security, analyzing land use change scenarios, handling competitive sectors via policy etc. Further to this, generating crop maps using GSAA parcels data would contribute to the new policy context that targets at creating comprehensive and comparable data throughout the EU for Performance-based Monitoring and Evaluation Framework (PMEF) in agri-environment-climate policy. Monitoring can be used in a variety of contexts. It is known that a land use map containing detailed categories of permanent grasslands, permanent crops and arable land featuring individual crop types could become the spatial standardized baseline for calculating several indicators tracking the change of biodiversity, climate change, land management trends etc., and running prediction models. Planning of interventions might even require sub-parcel categorization integrating several other aspects, leading to challenges of data interoperability.

By generating crop maps using GSAA parcels data this pilot study is planned to investigate along the following principles:

- **semantic harmonization** of open and available GSAA datasets of different countries;
- **thematic generalization** of crop types of GSAA data via aggregating individual crop types into group types following a hierarchical classification nomenclature;
- **geometric generalization** of spatial representations of the training samples derived from GSAA data, meaning that simplified geometries will be tested to represent the training data derived from the original GSAA geometries;
- use of well-established algorithm of supervised crop classification to automatize the parametrization of the model, to define a method sensitive for regional and seasonal differences, and to contribute to a **harmonized crop classification system**;
- comparison of the different **geometric representation of training data** using the misclassification matrix and the thematic accuracy values related to crop types.

The methodology developed in this study is based on the following assumptions, which is based on prior experiences:

1. Based on studying the free available GSAA datasets, and the willingness of member states to participate, 2 pilot sites had been selected. On the 1st pilot site of Germany – North Rhine Westphalia (NRW), a pixel-based classification method was preferred to object-based models, because a summarized decision for an object (i.e. crop parcel) might hide the differences of classification to evaluate the result visually. On the 2nd pilot site (Austria) the results of an object-based machine learning model were tested.
2. The difference between the classification of the training data based on the 3 geometric generalization (whole parcel's polygon, number of random points proportional to parcel's area, parcel's inner centroid) is assessed by using 2 types of popular image classification method: maximum likelihood supervised classification model, and the Random Forest (RF) supervised learning algorithm.
3. The overall quality of GSAA datasets used in this pilot study was not evaluated, but was accepted as a reliable input based on prior knowledge and on the OTSC sample. This was confirmed by a visual inspection comparing the parcels' geometry to the image time series. Correctness of the geometry topology (no overlaps, no invalid geometries) was tested.
4. In this study the image classification is based on optical Sentinel 2 images only. Better results can be obtained if the approach is extended with the combined use of radar and optical S1+S2 image for crop detection.
5. The pilot sites are small enough to be regarded as a single agricultural region where the crops have relatively the same vegetation profiles and can be evaluated with the use of the same image time series. Independently from this, the chosen method is targeting to be capable to deal with regional differences.
6. The minimum parcel size to participate in training/test data was set to 2000 m2.
7. The classification was run for agricultural area masking it with the arable land theme of the LPIS. This step excludes the misclassification effects sometimes caused by areas of forest, natural habitats and household gardening. This method might not work for all MSs, due to the limitations of LPIS data on non-declared but maintained land.

2 Selecting the area of the pilot

2.1 Method of pilot site selection

The two basic conditions of selecting a pilot site were 1) the availability of data and 2) the willingness of the MS administration to join the pilot. In the next step, test areas of approximately of 60Km X 60 Km were selected based on the following criteria:

Criterion	*Analysis and results*
Open availability of the GSAA vector data for year 2020 including not only the main land cover classes (AL/PG/PC/NAEA) but also the type of crop itself.	GSAA data of several member states available on INSPIRE portal was investigated. Only very few cases were found where not the LCC categorization, but the crop type was presented. Even the crop specific data of the selected North-Rhone-Westphalia region was delivered directly for this project by the PA.
Open availability of LPIS data	LPIS with the main LCC categories (AL/PG/PC/NAEA) and forest LCC data is usually available in INSPIRE.
Availability of S2 optical images for y2020	Due to the fact that time series of S2 L2A optical images had to be able to present the vegetation curve of each crop, areas covered by minimum 8 100% cloud-free images were selected. This is an ideal situation with seamless image data on the entire test area, what ensures better verification due to the homogeneous performance of the classification algorithm.
Representation of several AL crop types	The area is AL dominant, because the classification of arable crops had been in focus.
Representation of different parcel sizes	A visual analysis was made to select a zone where larger and also smaller parcels do appear.
Quality of GSAA vector data	This factor definitely had a high priority, as the current resources do not allow to develop methods for filtering the potentially non-proper parcels of the GSAA. In this study such plot area was chosen, where we know from previous experiences that the boundary of the declared parcels do quite exactly fits to the real location. Before running the classification, inhomogeneous crops were filtered by visual versification, and excluded from the training and test sample. The future implementation of CbM will also decrease the importance of this step, because the full monitoring workflow for BPS/SAPS/greening parcels is willing to verify the GSAA quality via the markers.

Two pilot sites have been chosen with similar size, one in **Germany, North Rhine-Westphalia (NRW)** and a second one in **Austria**. Both areas have relatively small parcels, with high diversity of crops and an outstanding general quality of GSAA data.

Size/grouping	160 000 ha GSAA parcels: 32 228 Group1: 27 crops / Group2: 16 crops	103 000 ha GSAA parcels: 29211 11 groups of crops
Training/test data:	polygon / random points / centroid + segmentation 50-50% stratified random	polygon / random points / centroid + segmentation 50-50% stratified random
Satellite images	Sentinel-2, L2A, 8 dates, cloud free, Principal Component Analysis: 9 bands	Sentinel-2, L2A, 11 dates, cloud free, Principal Component Analysis: NDVI + 10 bands
Classification methods:	Pixel based, 2 types: maximum likelihood, random forest	Object (parcel) based: random forest
	NRW = DEVELOPING THE METHOD	A = TESTING THE METHOD

2.2 Analyzing the GSAA / LPIS data available on the INSPIRE portal

As the 1st step of the pilot, data availability of several member states had been tested. The following links had been used for successful data download:

Country/region	Link
Netherlands	https://www.pdok.nl/downloads/-/article/basisregistratie-gewaspercelen-brp-
France	https://www.data.gouv.fr/fr/datasets/registre-parcellaire-graphique-rpg-contours-des-parcelles-et-ilots-culturaux-et-leur-groupe-de-cultures-majoritaire/#
Estonia	https://kls.pria.ee/geoserver/inspire_gsaa/wfs?service=WFS&request=GetCapabilities
Catalunya	https://analisi.transparenciacatalunya.cat/Medi-Rural-Pesca/Mapa-de-cultius-de-Catalunya-amb-origen-DUN/yh94-j2n9
Denmark	https://www.geodata-info.dk/srv/dan/catalog.search;jsessionid=8BE554223A7F4F9A1329DDE37DE83004#/search?resultType=details&any=gsaa&fast=index&_content_type=json&from=1&to=20&sortBy=relevance
Belgium WA	http://geoportail.wallonie.be/catalogue/44b10a46-4025-4020-a943-e8ffd5ccbd21.html
Belgium FL	https://geoservices.informatievlaanderen.be/overdrachtdiensten/Landbgebrperc/wfs?service=WFS&version=1.1.0&request=GetCapabilities
Slovenia	https://inspire-geoportal.ec.europa.eu/proxybrowser/#fq=resourceType%3Adataset&q=gsaa
Slovakia	https://www.nlcsk.org/arcgis/services/MPRV/LU_LPIS/MapServer/WFSServer?service=WFS&AcceptVersions=2.0.0&request=GetCapabilities https://docs.google.com/spreadsheets/d/1ZH0soNQ4uwuUTC9IGpCgbfD9Xy3jS0s6jGO-AoGv2d8/edit#gid=239495147 On Portal data.gov.sk – LPIS: https://data.gov.sk/dataset/system-identifikacie-polnohospodarskych-pozemkov-lpis – GSAA: https://data.gov.sk/dataset/hranice-uzivania LPIS description: https://portal.vupop.sk/arcgis/rest/services/LPIS/citlive/MapServer/layers GSAA app user guide: https://gsaa.mpsr.sk/2018/help/ziadatel/help.html#trueeditova-popis
Austria	https://geometadatensuche.inspire.gv.at/metadatensuche/srv/ger/catalog.search;jsessionid=852753D3EECB6DC16A207EF5D4B9BAA5#/metadata/c2412b1f-b77d-435a-b76a-d56f4609266e
Luxembourg	https://data.public.lu/fr/datasets/referentiel-des-parcelles-agricoles-flik/

<table>
<tr>
<td>Germany - NRW</td>
<td>
https://inspire-geoportal.ec.europa.eu/download_details.html?view=downloadDetails&resourceId=/INSPIRE-4fed3eb0-06fa-11ea-8480-525400695e9c_20201221-173502/services/1/PullResults/126401-126420/datasets/20&expandedSection=metadata
https://www.opengeodata.nrw.de/produkte/umwelt_klima/bodennutzung/landwirtschaft/
GSAA download can be found here – it do not contains the individual crops, just the crop group and a bool field if the parcel is under the crop div. requirement or not.
https://www.opengeodata.nrw.de/produkte/umwelt_klima/bodennutzung/landwirtschaft/LWK-TSCHLAG_EPSG25832_Shape.zip
<table>
<tr><th>Data</th><th>Declared and verified eligible parcels</th></tr>
<tr><td>Attribute</td><td>technical description</td></tr>
<tr><td>ID</td><td>distinct key</td></tr>
<tr><td>INSPIRE_ID</td><td>distinct INSPIRE-compliant ID-Key</td></tr>
<tr><td>FLIK</td><td>identifier for physical blocks</td></tr>
<tr><td>WJ</td><td>financial year</td></tr>
<tr><td>AREA_HA</td><td>area in ha</td></tr>
<tr><td>USE_CODE</td><td>crop group code</td></tr>
<tr><td>USE_TXT</td><td>AF = Agricultural fodder, AH = Other industrial crops, PA = Land withdrawn from production, GL = Permanent grassland, DA = Permanent culture, EW = Protein crops, EP = Energy crops, GM = Vegetables, GT = Cereals, GR = Greening / Landscape features, HF = Root crops, HP = Kitchen herbs / medicinal and aromatic plants, OE = Oilseeds, SF = Other land, SL = Decommissioning, ZP = Decorative plants</td></tr>
<tr><td>D_PG</td><td>permanent grassland yes / no</td></tr>
<tr><td>CROPDIV</td><td>crop diversification yes / no</td></tr>
<tr><td>EFA</td><td>ecological focus area yes / no</td></tr>
<tr><td>ELER</td><td>eligible for second pillar yes / no</td></tr>
<tr><td>DAT_BEARB</td><td>date of validity</td></tr>
</table>
LPIS download can be found here:
https://www.opengeodata.nrw.de/produkte/umwelt_klima/bodennutzung/landwirtschaft/LFK-AKTI_EPSG25832_Shape.zip
and the matching description you can find here:
<table>
<tr><th>Data</th><th>Physical blocks in NRW</th></tr>
<tr><td>Attribute</td><td>technical description</td></tr>
<tr><td>ID</td><td>distinct key</td></tr>
<tr><td>INSPIRE_ID</td><td>distinct INSPIRE-compliant ID-Key</td></tr>
<tr><td>FLIK</td><td>identifier for physical blocks</td></tr>
<tr><td>NUTZ_CODE</td><td>Land use (A, G, S, F, K)</td></tr>
<tr><td>NUTZ_TXT</td><td>A = Arable land, G = Permanent grassland, S = Other land, F = Eligible in the 2nd pillar, K = Permanent Crops</td></tr>
<tr><td>GUELT_VON</td><td>date of validity</td></tr>
<tr><td>AREA_HA</td><td>area in ha</td></tr>
</table>
</td>
</tr>
</table>

The selection of the pilot sites did not only depend on the availability of GSAA and LPIS data itself, but also on the willingness of the local administration to participate in the pilot, on the prior information about the quality of the data and on the availability of suitable S2 image time series. North Rhone Westphalia and Austria fulfilled all requirements. Slovakia could be the next candidate of a pilot site, due to the possibility that they will publish a full GSAA data for y2020 soon.

Experiences found during analyzing input vector datasets:

- Searching the data is already a challenge, because there are no fixed INSPIRE key-words attached to the different IACS-related datasets. Also, the location of the data in the INSPIRE dictionary is not standardized. Metadata is available but the level of data description under "resource abstract" varies a lot, as it is a free text. Under the current rules of publishing the metadata a stricter requirement of how to describe LPIS and GSAA data would solve this problem. A guide of detailing the content of the free text could be useful. The current stage is, that regarding some member states the description do not even contain "LPIS" or "GSAA", only contains the national name of the system. Further to this, meaningful descriptive data of LPIS could be further thematized and that would make easier the use of the data. Correct data content description has a strong importance because LPIS and GSAA data varies, and the member states often upload generalized content, what could result to the fact that it is quite difficult to understand the exact content. An example for what data could be retrieved regarding an LPIS upload:

    - Date of validity from-to, in case of rotational update cycles parts with seamless update cycle should be uploaded as a single unit.
    - Type of reference parcel
    - Is the data spatially continuous or not?
    - Thematic content of the layer: name and values of attribute filed should be documented regarding the following data contents, as an example:
        - SAPS eligible area and non-eligible area,
        - main land cover categories,
        - detailed land cover categories – the list and description of categories would be also an essential information, that is why the way of correctly publishing these attributes in the registry should also be clarified and communicated towards the data uploader users,
        - Permanent grassland (PG) categorization if exists,
        - PG ELP categorization, if exists,
        - PG pro-rata categorization, if exists,
    - Main data source of defining the reference parcel,
    - Rule of minimum mapping unit applied for AL,
    - Rule of minimum mapping unit applied for PG,
- Rule of minimum mapping unit applied for PC. In several countries not the Paying Agency as direct data manager uploads the data to INSPIRE. That is why the descriptive information and the generalization of the classes might not be well presented.
- In general case distinguishing crops is not a requirement for direct payments. However, for certain categories of (big) farmers yes, as they have to comply with the crop diversification, or other greening requirements. As an experience it is detected, that most of the MSs asks all the farmers to declare a detailed crop type for their parcels – also for those, where the payment scheme does not require it. On the other hand, the detailed crop type is usually not published in the GSAA data under the INSPIRE. The reason can be, that it is not specifically required, and the level of detailing the GSAA might lead to additional questions, so to fill up the correct metadata to explain the content is quite an investment. In practice most member states choose a simplified solution: uploading the crop group categories. These crop group categories are the groups of area-based direct payments of the given MS. For image classification purposes grouping of crops from the perspective of similar vegetation curves is needed, and it is more likely to derive from the lowest level of category breakdown.
- Crops planted as secondary ones in a vegetation year, called catch crops should be distinguished from the same type of crop cultivated as a main vegetation of that year.
- It would be very important to collect the detailed crop category in GSAA also for those parcels that are declared as a collective category: like crops for wildlife, set aside where certain plants are allowed to be maintained etc.
- Although INSPIRE technical guidance (TG) requires that *"the title of a single dataset shall refer to the claim year, to the Member state and if applicable, to the region"* it is not fully respected by all client, and it would be very important to implement correctly the TG. Geometric and semantic consistency of LPIS and GSAA should be checked, because it is not documented what version of an annual data is uploaded to INSPIRE.

2.3 North Rhine Westphalia (NRW) pilot site

A pilot site was selected in Germany, North Rhine Westphalia, because the GSAA and LPIS data were available on the INSPIRE web page. In addition, the Paying Agency (Landwirtschaftskammer Nordrhein-Westfalen) had previously shown an interest in sharing IACS data.

A cropland dominated area of 42x39 km in North Rhine Westphalia -Germany was selected as 1st pilot site:

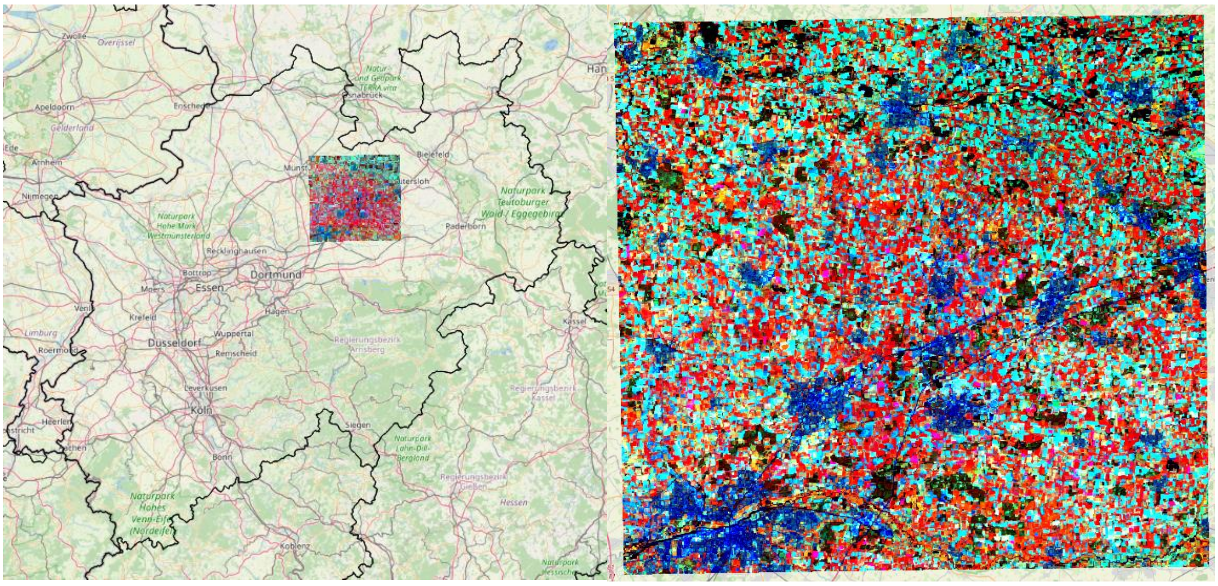


Image 1.: The NRW pilot site: Open Street map and false color composite of R=NIR, G=SWIR1, B=RED

The reference parcel type of the region is physical block (PB), with the categorization of

- A = Arable land,
- G = Permanent grassland,
- S = Other land,
- F = Eligible in the 2nd pillar,
- K = Permanent Crops

There are 24 016 PBs on the test site delineating 110 809 ha of eligible area, out of which in y2020 46 449 crop parcels were declared, meaning that in average there are 1,9 parcels declared in a PB. 97,57 % of the eligible area is declared. The area is quite a typical winter cereal/maize dominant crop rotation, with only 15% of PG and big varieties of crops with relatively small areas. The production of Acre grass varieties for energetic purposes is increasing (4 % in y2020) and the asparagus and strawberry production is also typical in the area.

There are 121 types of crops declared on the North Rhine Westphalian pilot site, out of which 13 crops cover 93 % of the declared area with more than 1% of area share compared to the total declared area. The other 108 crops are all under 600 ha/crop type and cover 7 % of the area. The average parcel size is quite small: 2,86 ha for the 12 types of main crops, covering 93% of the area, and 3,05 ha if all parcels are counted in the average.

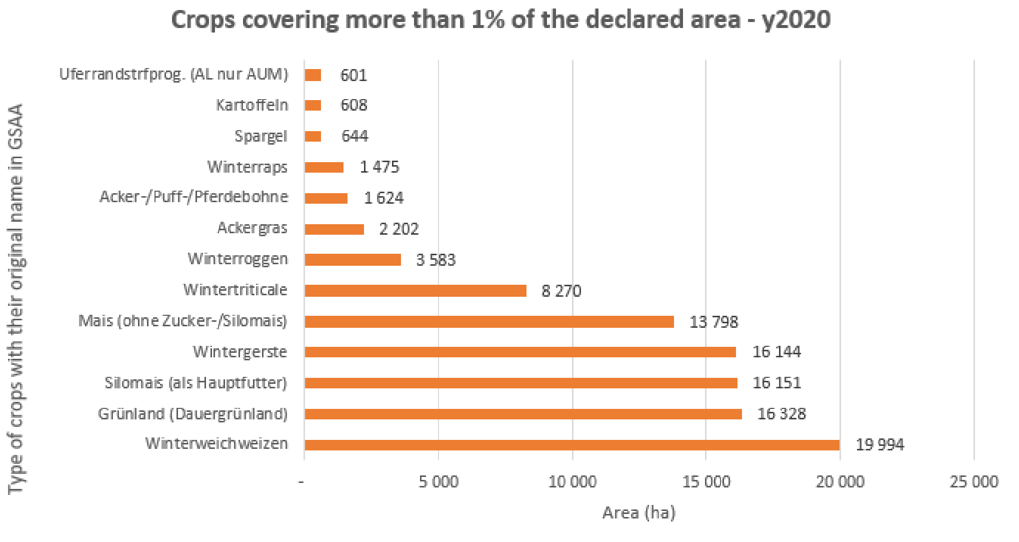


Image 2: Crop types evaluated on the North Rhine Westphalia (Germany) pilot site

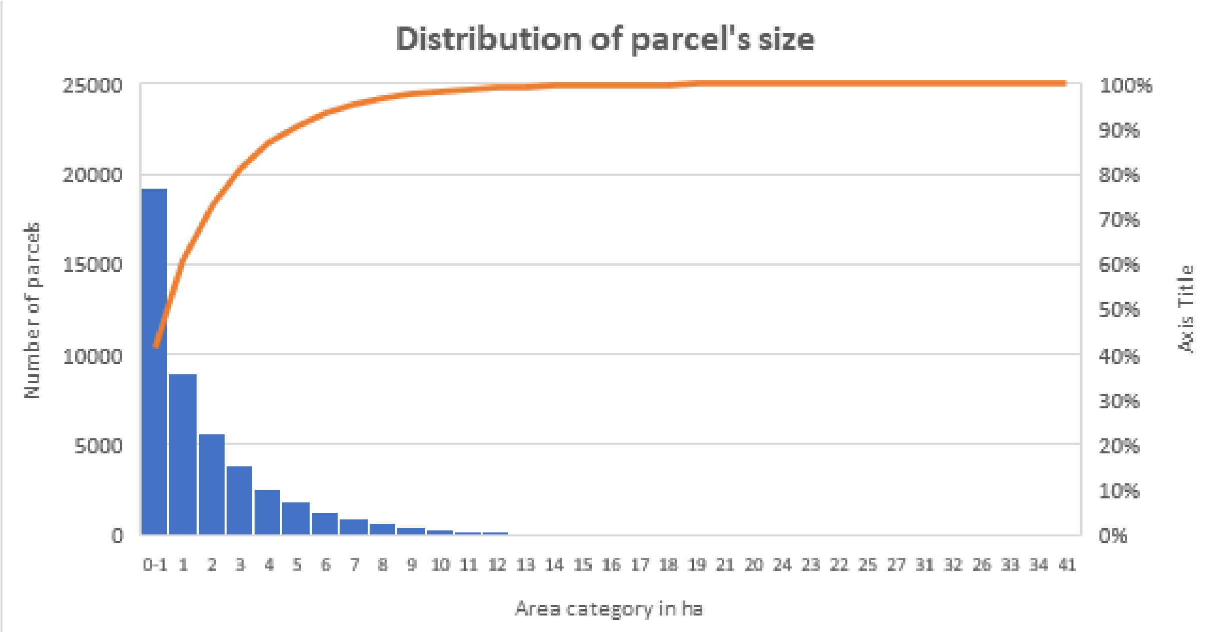


Image 3: Distribution of parcel's size on the North Rhine Westphalia (Germany) pilot site

Type of crops with their original name in GSAA	***Type of crops EN***	***NR of parcels***	***Area (ha)***	***average parcel size (ha)***	***% of sum declared area***
Winterweichweizen	Winter wheat	4912	19 994	4,070	18%
Grünland (Dauergrünland)	Permanenet grassland (PG)	11432	16 328	1,428	15%
Silomais (als Hauptfutter)	Silomaize	4871	16 151	3,316	15%
Wintergerste	Winter barley	4267	16 144	3,784	15%
Mais (ohne Zucker-/Silomais)	Maize (excluding sweetcorn and silo)	3831	13 798	3,602	13%
Wintertriticale	Winter tritcale	2302	8 270	3,593	8%
Winterroggen	Winter rye	1229	3 583	2,915	3%
Ackergras	Ackregrass	1349	2 202	1,632	2%
Acker-/Puff-/Pferdebohne	Ackre and other energetic grass	373	1 624	4,354	2%
Winterraps	Winter rape	310	1 475	4,757	1%
Spargel	Asparagus	180	644	3,579	1%
Kartoffeln	Potato	218	608	2,789	1%
SUM/AVE		***35274***	***100 822***	***2,858***	***93%***

Table 1: Main data of parcels evaluated on the North Rhine Westphalia (Germany) pilot site

2.4 The Austrian pilot site

As a 2nd pilot site Austria was chosen, because the GSAA data is fully available on open platforms, and the Paying Agency (Agrarmarkt Austria - AMA) together with its contractor the EOX IT Services GmbH have a proven experience of running ML models for predicting type of crops. The expert team presented an interest to investigate into testing how GSAA can support classification methods to derive crop specific thematic maps as a potential input for AMS.

The quality of the GSAA, including the localization of parcel boundaries is outstandingly high in Austria, which made the GSAA data suitable to test directly to train ML models, without an extensive a priori investigation of GSAA. An area of 30x31 km^2 south of the city Linz was selected representing the main variety of crops mostly in small parcels:

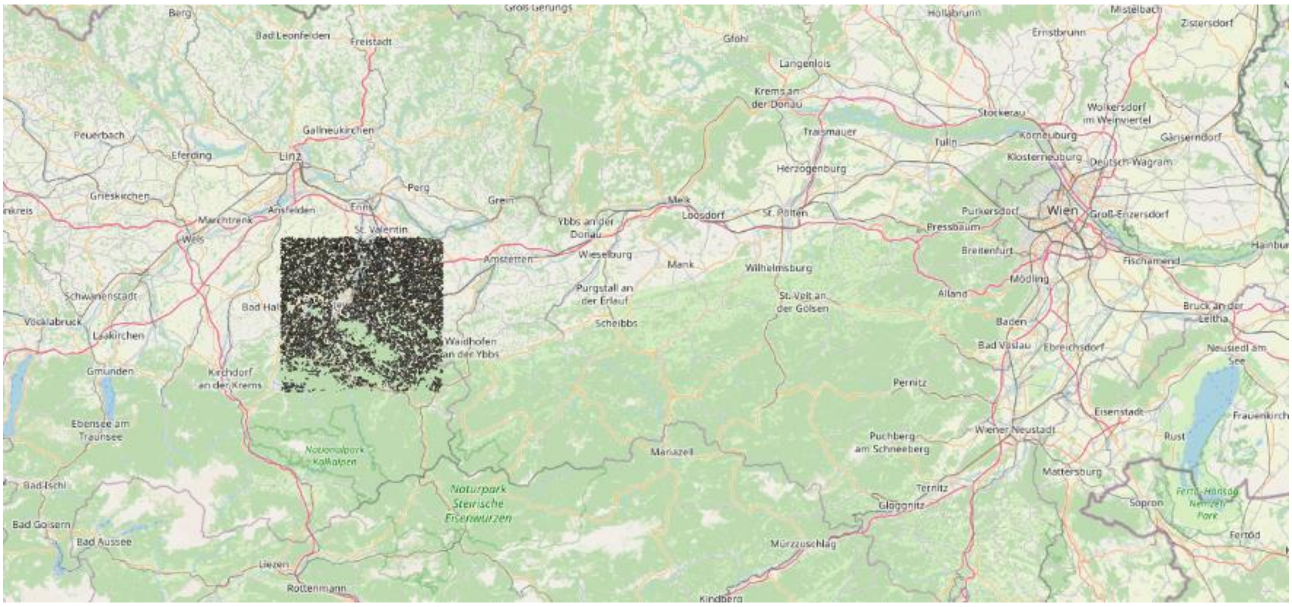


Image 4.: The Austrian pilot site: Open Street map and false color composite of R=NIR, G=SWIR1, B=RED

2.5 Selection of satellite images

2.5.1 NRW pilot site

Due to the extent of the study, a time series of 8 optical Sentinel2 images were selected, covering the whole crop growing period. Evaluation and selection of the images were done based on visual inspection. The entire site has a seamless image composite. This is an ideal situation to test how the different training data can be used to classify crops under the condition that input data and the classification algorithm are unchanged.

Acquisition date	***Type of satellite***	***Preprocessing status***	***Cloud free***
2020.02.07	s2a	BOA – L2A	100%
2020.03.23	s2b	BOA – L2A	100%
2020.04.17	s2a	BOA – L2A	100%
2020.05.07	s2a	BOA – L2A	100%
2020.06.01	s2b	BOA – L2A	100%
2020.06.26	s2a	BOA – L2A	100%
2020.08.05	s2a	BOA – L2A	100%
2020.09.19	s2b	BOA – L2A	100%

Table 2: Satellite images used on the North Rhine Westphalia (Germany) pilot site

The visual inspection of the image stack using one image per month showed that it still did not follow with enough details the development of crops with a shorter vegetation periods (like pea). Furthermore, it was not sensitive enough to classify crops with catch crop. The recommendation is to use a minimum bi-weekly coverage. The over 6 weeks of period in between 2 images were not always suitable to detect each multi crop combination on the test site. Data of S2 will ensure the detection of ploughing after harvesting the main crop, and this would exactly be enough to increase the accuracy of certain crop types.

The following bands were used in 20 meters resolution: blue/green/red/red edge1/ red edge 2/ red edge3/NIR/SWIR1/SWIR2.

Number of band used	***Sentinel-2 Bands in original order***	***Central Wavelength (µm)***	***Resolution (m)***
-	Band 1 - Coastal aerosol	0.443	60

1	Band 2 - Blue	0.490	10
2	Band 3 - Green	0.560	10
3	**Band 4 - Red**	**0.665**	**10**
4	Band 5 - Vegetation Red Edge1	0.705	20
5	Band 6 - Vegetation Red Edge2	0.740	20
6	Band 7 - Vegetation Red Edge3	0.783	20
7	**Band 8 - NIR**	**0.842**	**10**
-	Band 8A - Vegetation Red Edge	0.865	20
-	Band 9 - Water vapour	0.945	60
-	Band 10 - SWIR - Cirrus	1.375	60
8	**Band 11 – SWIR1**	**1.610**	**20**
9	Band 12 – SWIR2	2.190	20

Table 3: Bands of satellite images used on the North Rhine Westphalia (Germany) pilot site

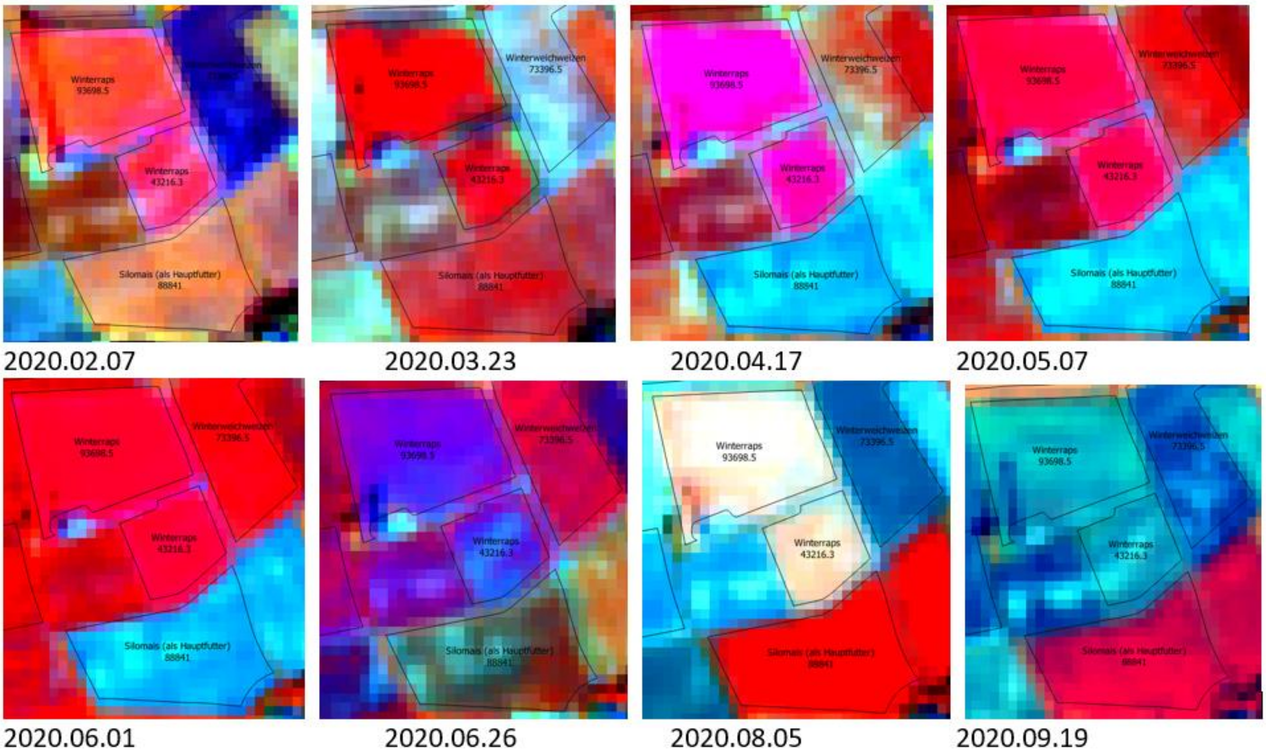


Image 5: An example of the 8 images using R=NIR, G=SWIR1, B=RED false color composite was used for visual verification of the processes - North Rhine Westphalia (Germany) pilot site

2.5.2 Austrian pilot site

A series of 11 optical only Sentinel2 images was selected. In the image selection, no focus was laid on establishing an equitemporal time series. However, it was of concern to select as many cloud-free images as possible within the vegetation period.

Acquisition date	***Type of satellite***	***Preprocessing status***
2020.01.09	*s2a*	*BOA – L2A*
2020.03.19	*s2a*	*BOA – L2A*

2020.04.05	*s2a*	*BOA – L2A*
2020.04.08	*s2a*	*BOA – L2A*
2020.05.18	*s2a*	*BOA – L2A*
2020.06.27	*s2a*	*BOA – L2A*
2020.08.13	*s2a*	*BOA – L2A*
2020.09.05	*s2a*	*BOA – L2A*
2020.09.15	*s2a*	*BOA – L2A*
2020.09.22	*s2a*	*BOA – L2A*

Table 4: Satellite images used on the Austrian pilot site

As in the NRW test site, the atmospheric bands (1, 9, 10) were excluded in the Austrian test site. In contrast to the NRW site, also the Band 8A, "Vegetation Red Edge", was included.

2.6 Analyzing the appearance of different crops on the NRW test site

Due to the lack of a priori knowledge about the crop management of the given region, quite an effort was spent to analyze the appearance of the crops. The reason of this was to understand the similarity and the difference among the vegetation curves of each crop type, because this is the fundament of implementing the grouping and regrouping of crops.

To understand and to compare the development of crops an NDVI stack was created, with the following parameters:

NDVI stack - false color composite	***Order of NDVI values***	***Acquisition date***	***Type of satellite***	***Band used for NDVI***	***GSD used for NDVI***
	1	2020.02.07	s2a	B8+B4	10m
BLUE	2	2020.03.23	s2b	B8+B4	10m
	3	2020.04.17	s2a	B8+B4	10m
	4	2020.05.07	s2a	B8+B4	10m
RED	5	2020.06.01	s2b	B8+B4	10m
	6	2020.06.26	s2a	B8+B4	10m
	7	2020.08.05	s2a	B8+B4	10m
GREEN	8	2020.09.19	s2b	B8+B4	10m

Table 6. Bands of satellite images used to derive NDVI on the North Rhine Westphalia (Germany) pilot site

Image 6. False color composite in y2020 to visualize the NDVI stack where R = NDVI 2020.05.07, G= NDVI 2020.09.19 B = NDVI 2020.03.23

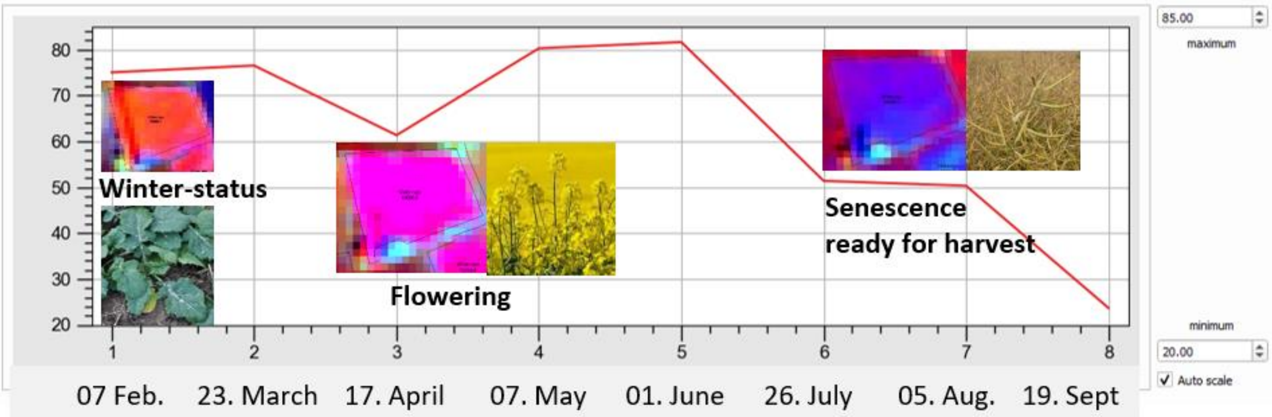


Image 7. The multi temporal NDVI profile (on scale 0-100) of a rapeseed parcel

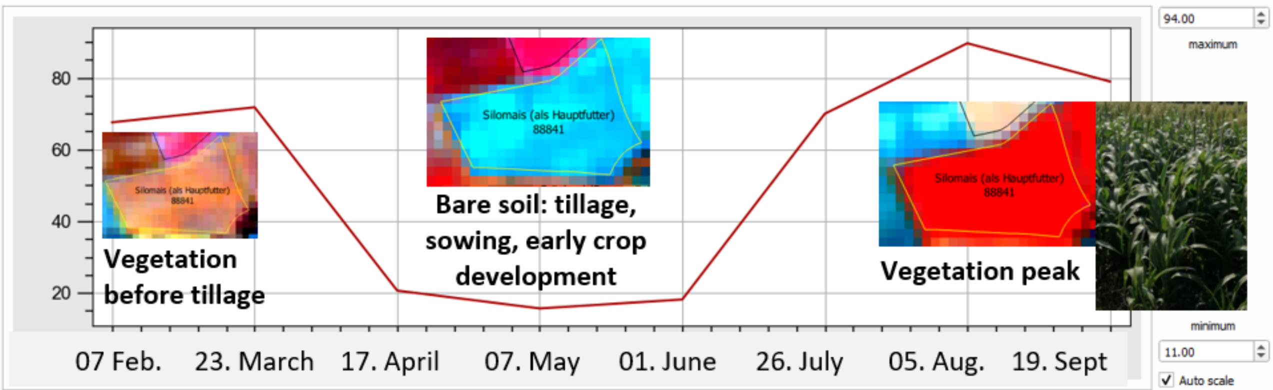


Image 8. The multi temporal NDVI profile (on scale 0-100) of a silo-maize parcel

There are 3 crops being specific for the NRW region, that is why special attention was taken for analyzing their development and appearance on the image time series.

Horse-bean, Vicia faba

(Acker-/Puff-/Pferdebohne)
In this example there are 2 parcels with 4 weeks of time shift in the crop development. It seems that classification had some weakness, it performed on 83 %, regarding the successful performance of the training on the cluster image.

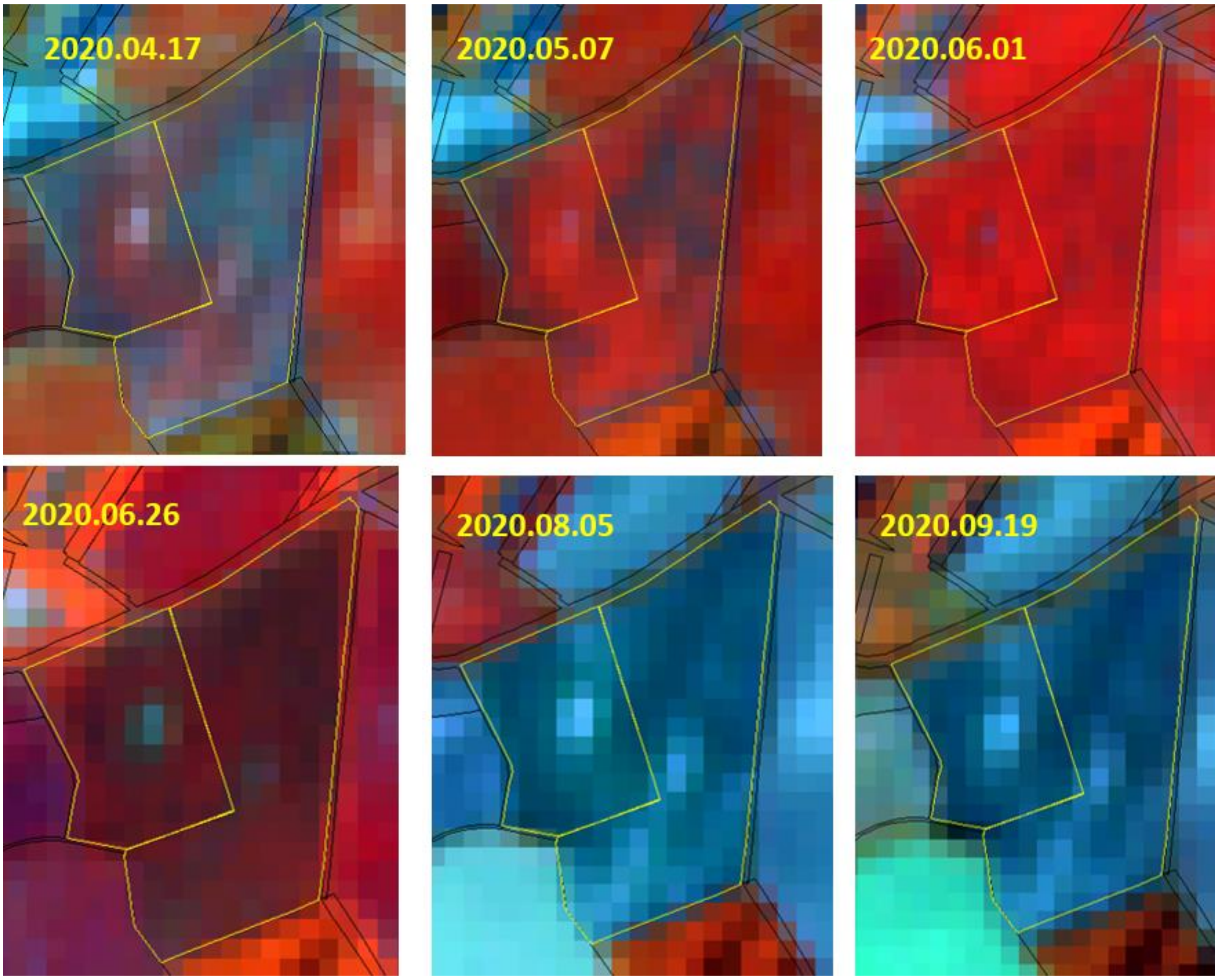


Image 9. Vicia faba - example1 False color comosite of R=NIR, G=SWIR1, B=RED

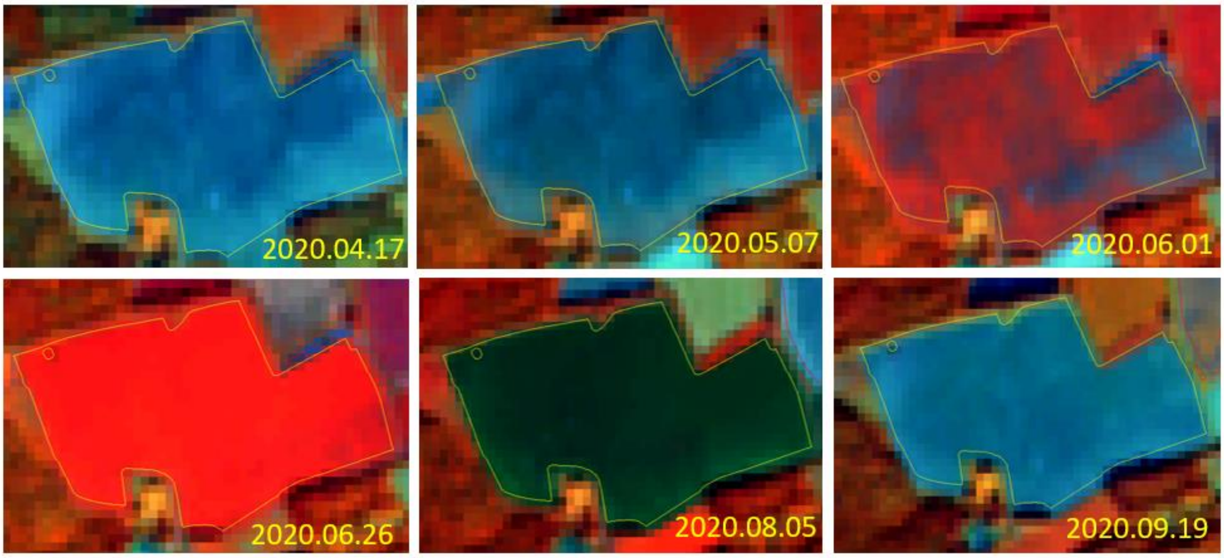


Image 10. Vicia faba – example2 False color comosite of R=NIR, G=SWIR1, B=RED

Asparagus / Spargel
is a multi annual crop, where the white or black plastic cover also determines the reflectance.

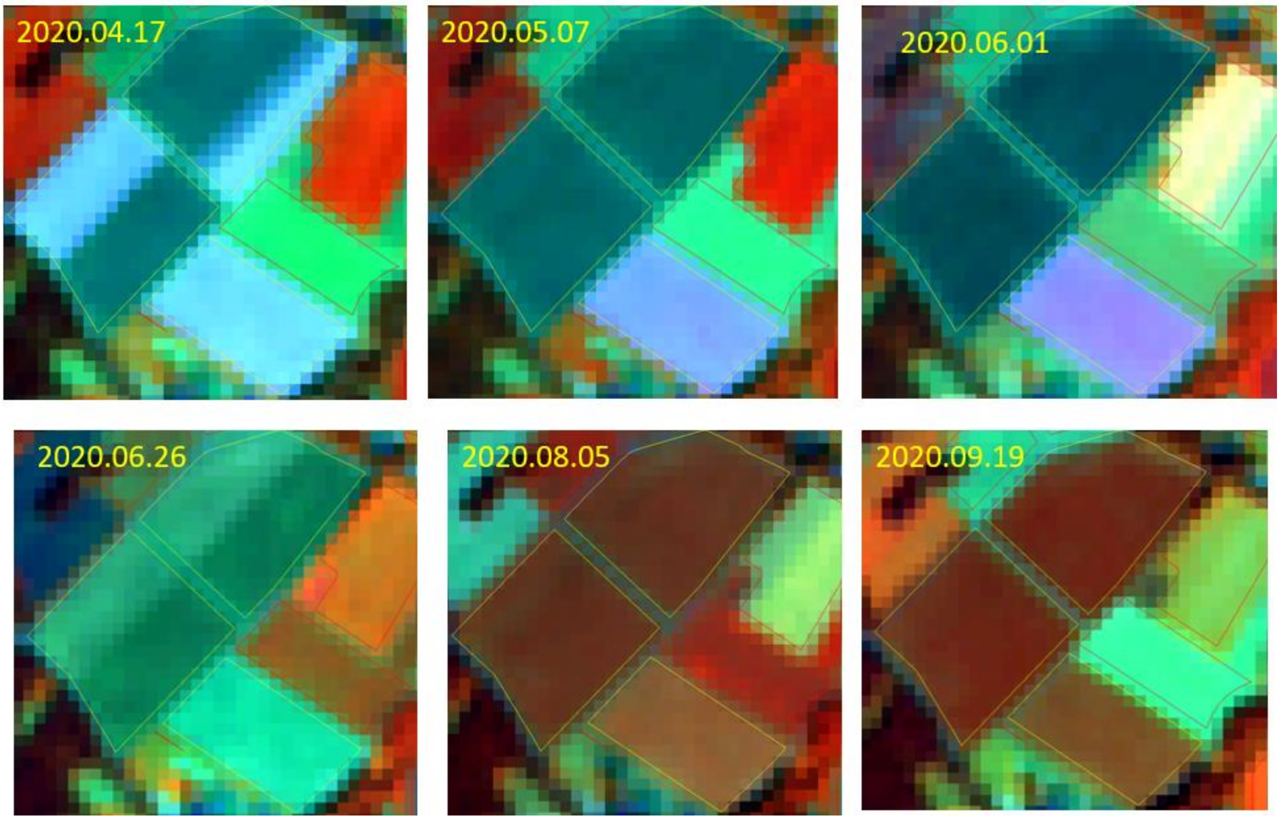


Image 11. Asparagus - false color comosite of R=NIR, G=SWIR1, B=RED

Acre grass / Ackergras is a multi annual crop

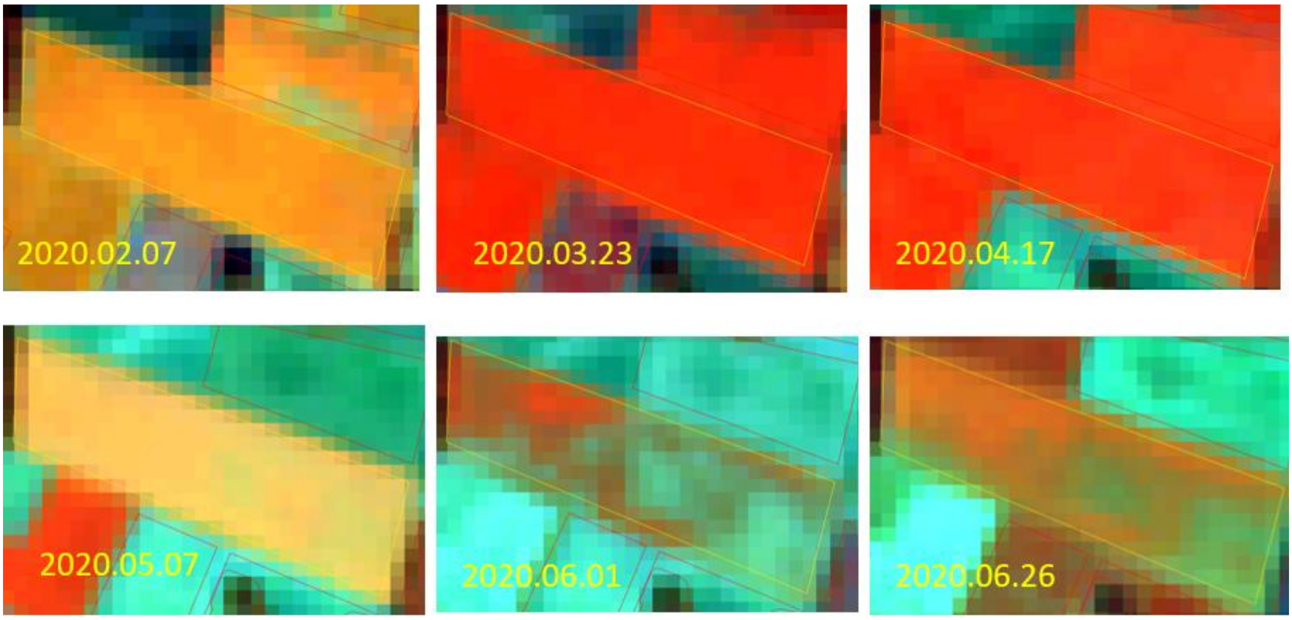


Image 12. Acre grass - false color comosite of R=NIR, G=SWIR1, B=RED

2.7 Selection of bands considered to be the most relevant

2.7.1 NRW pilot site

There are 72 bands of all 8 the images together, as the following 9 of the S2 bands are stacked: blue/green/red/re1/re2/re3/nir/swir1/swir2. A few among the 72 bands bringing significantly important information to identify different crops, but there are also bands correlated highly in such a way that they contain redundant information. To filter out the most relevant bands Principal Component Analysis (PCA)[i] was implemented.

The image masked with LPIS eligible area was the input of the PCA. The result is the following:

Sum of eigen values counted for all 72 bands:	23 893 121
Sum of eigen values counted for the 18 bands selected as most important:	23 544512
Ration of information content of the selected 18 bands:	98,5 %

The 18-band image what is the result of the PCA was the input of the classification. The following false color composite represents the most relevant data combination of the transformed bands.

Principal components	*Eigen value*	*% of information*	*False color composite*
1	9 636 429	40,3	B = meaning that where the blue color is dominant the 1st component of signal combinations plays a significant role
2	4 012 766	16,8	G = meaning that where the green color is dominant the 2nd component of signal combinations plays a significant role, and

			it is a larger range, such it consists of majority of the spectral information
3	2 240 261	9,4	R = meaning that where the red color is dominant the 3rd component of signal combinations plays a significant role

The following representation of the result was used to decide the band selection:

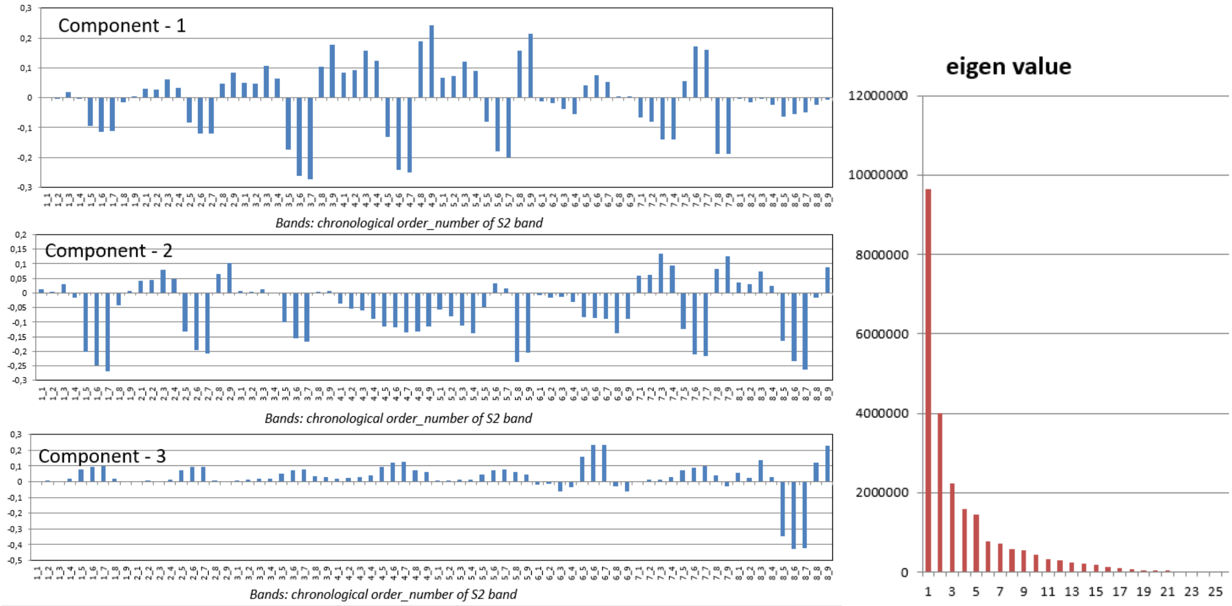


Image 13. Bands used to derive 1st, 2nd and 3rd principal component

The first 3 component can be visualized in a false color composite, to check visually how the different type of crops are separated.

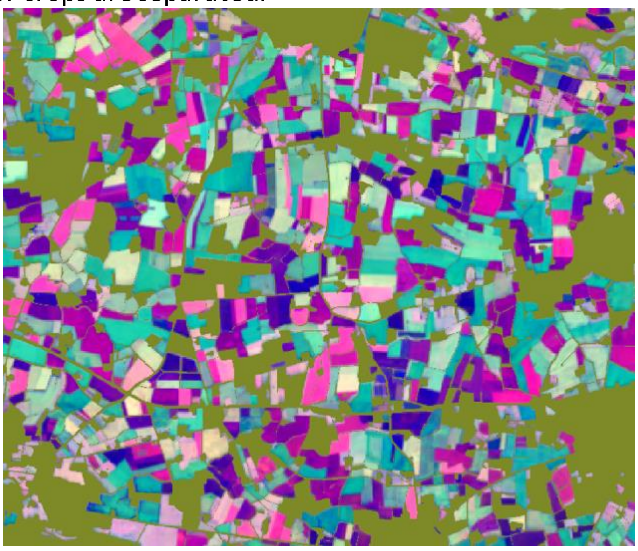
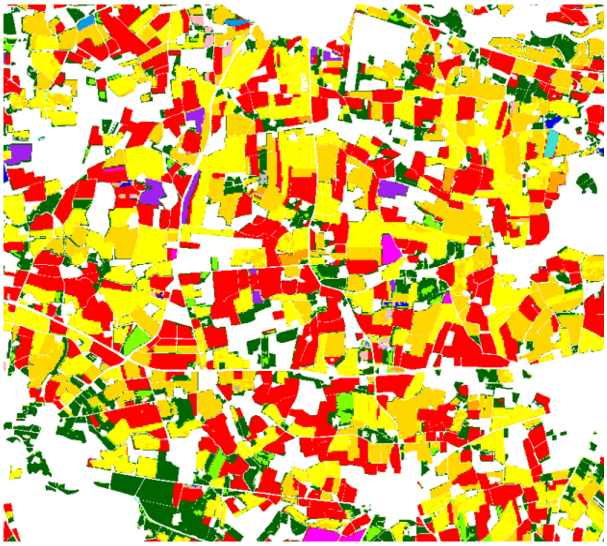


Image 14. R: PC3, G: PC2, B: PC1



Main crop classes identified (legend is in table 8.)

2.7.2 Austrian pilot site

The feature space of 110 features was reduced by applying a PCA to components explaining 95% of the Variance. For the polygon input data, this resulted in 15 used principal components, in case of the centroid and random point input data, 16 principal components were used.

3 Preparation of reference data

3.1 Geometrical preprocessing and separating training and test samples

The reference data is derived from the IACS -GSAA data of year 2020. The following initial steps of preparation were used before applying any kind of generalization. Further generalization of the reference data was derived from the output of this geometrical preprocessing. Separation of training and test was implemented at this preprocessing phase to ensure that the same training and test parcels are used at each version of the process tested, and to ensure that the result is comparable on a way, that it is independent from the test and training separation.

Steps applied:

1. Selecting GSAA polygons fully overlapping with the image of the test area.
2. 50% of training and 50% of test data were separated, with stratified random sampling. Crops were used as subsets, assigning a percentage to define the total number of randomly selected features in the subset. The percentage value is not applied to the whole layer, but to each crop category. In this case the share of crops both in the training and in the test sample represents the same share as it is in the entire population. Not all classification algorithms are sensitive to the fact if the training data properly represent the same share of a crop as it is in the entire population.
3. Implementing a negative buffer of the GSAA polygons to exclude from the classification the mixed pixels by the boundary of the parcels. As the NRW region is not hilly, buffer do not have to handle the possible DTM-derived boundary discrepancies. When defining the size of the buffer the parameters of vector to raster transition also should be taken into account. In this study the most common solution was used: a valid pixel of the raster is created if >50% of the pixel's area is overlapped by the input vector.
   As a result of buffering parcels all along narrower than 2 times the inner buffer will disappear, and parcels where only part of their extension is narrower than 2 X inner buffer will be split. To handle these a multipart to single part is implemented, and only those polygons are kept that have a minimum width along the entire extent over 60 meters. This will be a condition later at the step of generating simplified representation geometries as an inner point as reference data.
   (3A) – In practice **for POLYGON topology of training and test data 10 and 20 m of negative buffers were used**. 10m was used in case NDVI or stacks of S2 bands having 10 meters of Ground Sampling Distance (GSD) (for B3,4,8, NDVI), while 20 m was used if bands with GSD=20 m participates in the classification. Suitability of handling boundary-elongated clean pixels and the quality of crop development by the side of parcel margins was validated visually. Drop of NDVI by the side of a parcel is strongly determined by the success of the crop management, and the representation of less-developed crop by parcel margins might lead to the separation to an independent class of parcel margins rather than classifying the pixels to the given crop.

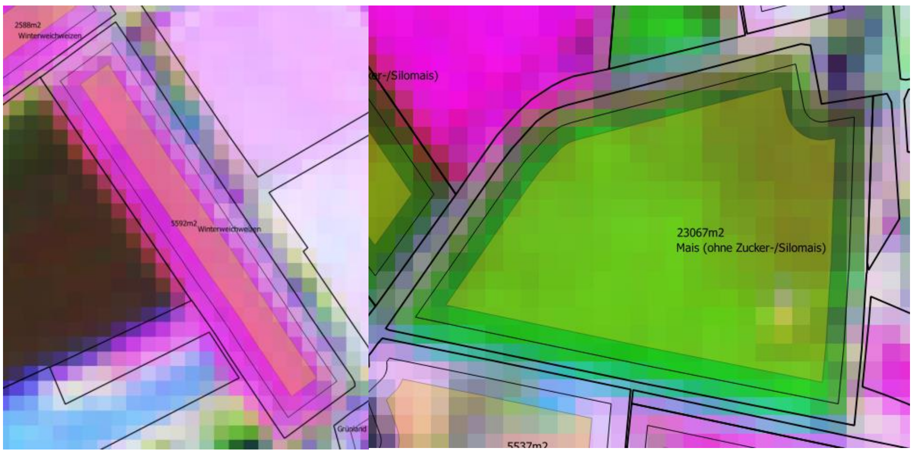


Image 16. NDVI_stack _8dates_10m 10 m of buffer is appropriate with 2000 m2 of size limit R=Band5, G=Band8, B=Band4

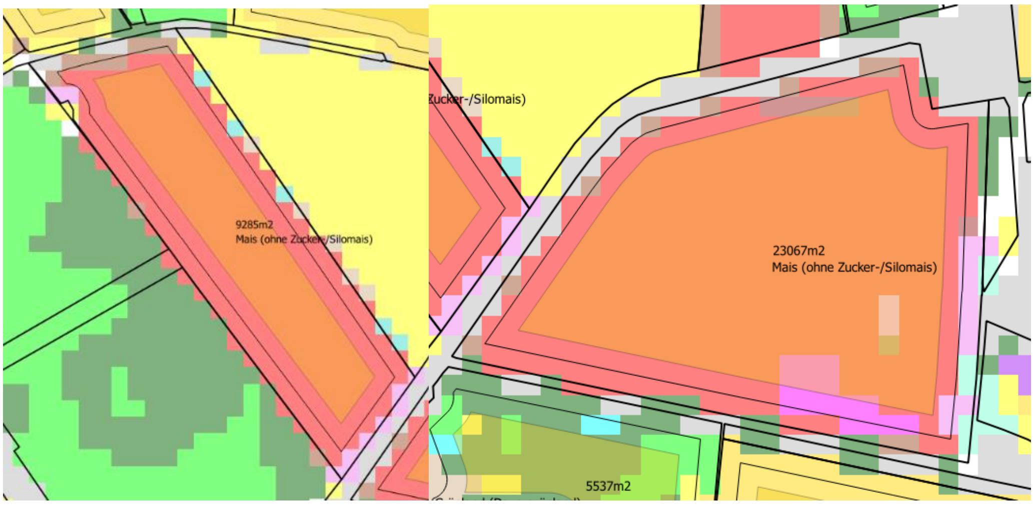


Image 17. RED_NIR_SWIR_8dates_20m – 72 ISODATA clusters 20 m of buffer with 2000 m2 of size limit

(3B) – **for polygon input used for deriving random POINTS inside the parcels** 30 meters (1,5*GSD) of negative buffer is calculated, thus random generator is only working inside this area.

4. (equal to steps 3A and 3B) Applying geometrical function of multipart to single part, to keep polygons over 200 m2 as single units what had been split because of the buffer.
5. (equal to steps 3A and 3B) Keeping the parcels what are over 2000 m2 regarding their original geometrical size.

3.2 Generalization in geometry the training and test datasets used in the pilot site

To clearly compare (reducing additional factors) how a different type of training and test data works regarding its geometrical representation 3 types had been tested:

1. using as training and test the entire polygon of the crop parcel,
2. generating points inside the polygon of the crop parcel, called multipoint solution,
3. generating a single inner center, to represent crop parcel, by creating a point at the "center" of an area and adjust the position of the point so that it always falls within the area.

While generating the three versions of reference data, the following solutions were used:

- all random points will fall into pixels participating in the rasterized training and test data,
- the parcels are represented both by their entire area and by points and multipoints;
- 30 meter inner buffer of the GSAA polygons will ensure that all random points will always be located inside the rasterized GSAA polygons.
- in the classification each point is represented by a circle with 10 meter radius, resulting, that the spectral value of maximum 9 pixels will be used with no overlap, and the 60 meters used for minimum distance among the points ensures to avoid the overlap,
- training and test separation are stable, split before generating the two point type of geometrical representation.

Two versions of generalizing GSAA parcels with points were tested:

1. **Maximum 8 or less random multipoints** inside the polygon of the crop parcel, with minimum 60 meters of distance applying a proportional decrease by parcel size. To reach this, the random generator must be set to use only one try to locate the point. It means that in average for parcels smaller than 3240 m2 the number of points will be proportionally decreased. The function is run in QGIS under the menu Vector/Research tools/Random Points Inside Polygons

```
'Random_points_in_polygons'
'MAX_TRIES_PER_POINT' : 1,
'MIN_DISTANCE' : 60,
'POINTS_NUMBER' : 8,
```

```
    'SEED' : None }
```

2. **Single point, as inner centroid** of a parcel with a 10 meters buffer, meaning that maximum 9 pixels will be represented.

An example how training reference data is presented:

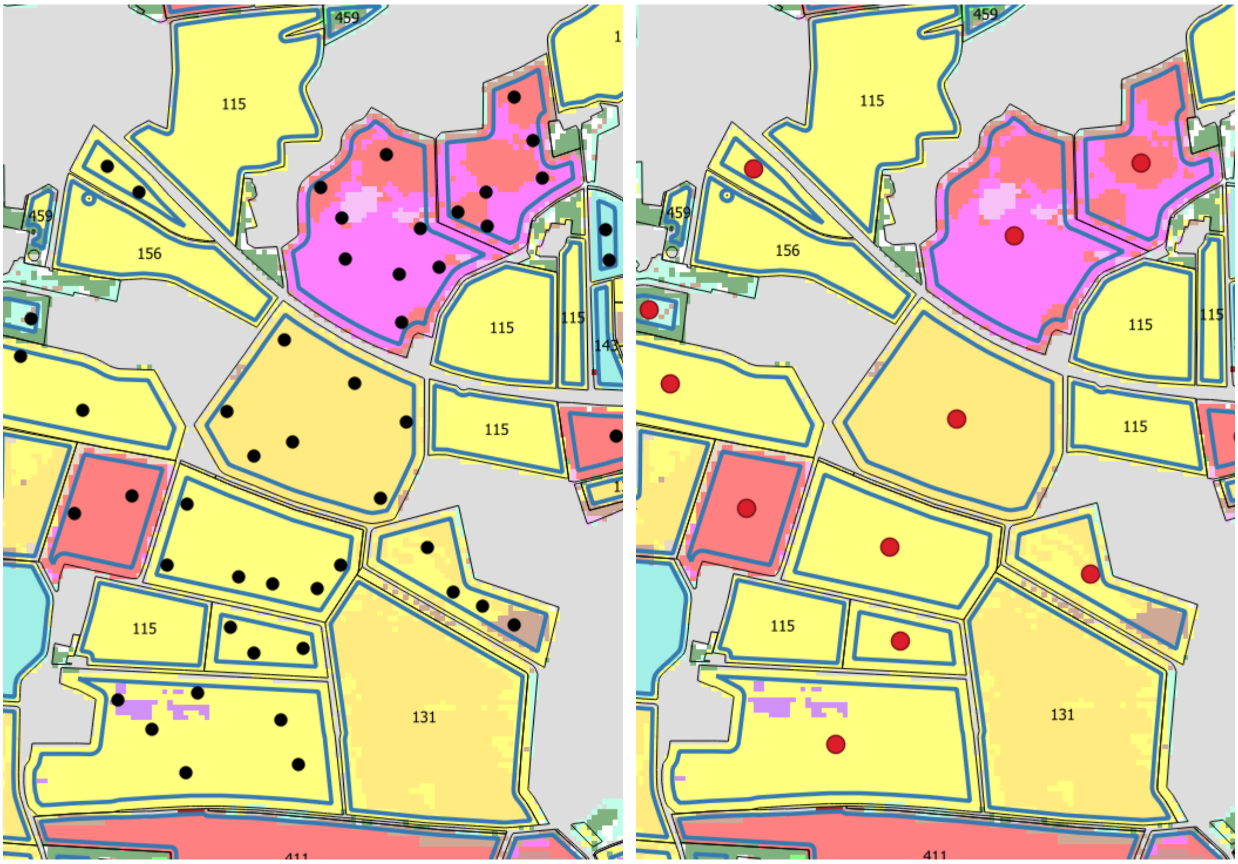


Image 18. Blue lines are representing the polygons of training data with the 20 m of buffer, black points are the result of random point generation, while the red points are the inner centroids of the polygons. The background image is an example of a supervised pixel based classification of the 8-multi-temporal stack of RED+NIR+SWIR bands using maximum likelihood and 60 clusters

To find the best combination of input data and image processing methods a few test runs were implemented. If the inner centroid was used only with the 10 meters buffer around the point, the classification was not very satisfactory. To improve the classification result a segmentation was implemented around the single points, and that leaded to a surprisingly good solution.

To test the performance of training and test data in image classification, the following number of features had been used on the NRW pilot site:

Number of features used	*GSAA polygon with Buffer 20 m, over 2000 m^2*	*Random maximum 8 points*	*Inner centroid of the parcel*
Training – strat. random 50%	17 030	36 436	17 030
Test – strat. random 50%	17 073	38 183	17 073
Derived feature used from classification	Each individual pixel of the parcel	Each individual pixel overlapping with a 10 meters buffer around the point	Each individual pixel of the segmented unit

Table 7: Result of deriving the different representation of training and test data on the NRW test site

3.3 Semantic grouping of training and test data on the NRW pilot site

Based on analyzing the performance of the crop types on the pilot area there were 2 main sets of reference data prepared using 2 levels of semantic grouping:

1. Crop level preparation (CODE-1 in Table 8.): keeping each unique crop individually, only applying such groupings, and filtering of minimum amounts what leads to clear distinguishable crop categories Only crops with exactly the same phenological presence and identical management technology are merged.
2. Crop group level preparation (CODE-2 in Table 8.): Summarizing the main crop types to larger categories commonly used in crop rotation planning and in land management evaluation. This 2nd categorization is implemented after the 1st, so the categories defined in step-1 are summarized further on.

From classification perspective the 1st category, to separate 27 crops is targeting a some critical issues, and the 2nd level of generalization can always be derived from the 1st output and summarizing the crops with highest rate of mixing.

The classes of the 2 level of generalization:

Aspects	*CODE-1 Preparation of reference data of separate Crops*	*CODE-2 Preparation of reference data on Crop Group level*
The required breakdown of output classes	Individual Crop level	*Crop Group level*
Potential use in AEM	Monitoring and evaluating crop diversification on the entire area (not only for greening-participants). Evaluating trends of changing crop share and of regional crop pattern.	Detecting widely used traditional rotations on group level, to define additional requirements for AEM. Monitoring the change of AL/PG/PC/NAEA/Forest share, Developing for land use planning scenarios.
Method of grouping GSAA crop categories	Assign to a single category crops what practically contains the same crop with the same management practice and will lead to the fact that a separate training data will never form an independent class during classification. Examples: fallow land and fallow land for greening with the same maintenance rules are a group; different garden vegetables flowers and herbs are merged into kitchen garden categories. Assign to a single category crops with the representation of very similar development (vegetation curve) on the given set of input satellite images. Examples: different tree-plantations, apple, peach, plum are merged as permanent crops (PC).	Creating crop groups what belongs to the same category of land management from the point of view of the target level of distinction. For example, for monitoring crop rotations groups of (1) winter cereals, (2) spring cereals (3) maize types (4) clovers and forage crops, (5) mixed row crops are necessary. Summarizing crops of the same categories makes the crop map more accurate, but crops with unique spectral profiles and large enough area to be well trained (like rape seed or Ackergras on NRW site) should not be merged, because it could lead to unnecessary misclassifications.

Semantical categories remained, and color legend used in the crop maps opn NRW site Number of crop categories remained	KOD1 Winter wheat / Winterweichweizen Sumer barley /Sommer gerste Winter Barley /Winter gerste Winter rye / Winterroggen Summer oat / Sommerhafer Winter triticale/ Wintertriticale Mays / Mais Mays for silo / Silomais Sugarbeet /Zuckerrüben Peas / Gemüseerbse Potato/Kartoffeln Winter-rapeseed/Winterraps Permanent crops (PC) Fallow land (FL)/ Bracheflache Permanent grassland (PG) Forest (F) / Forstflachen, Aufforstung Soyabean /Sojabohnen Winter oat/Winterhafer Mixed vegetables / Gemenge Leguminosen Asparagus/Spargel Strawberry /Erdbeeren Game-fodder/Wildacker auf lw. Flache Clover, clover-mix /Kleemischung Cup-plant/Silphium (Durchwachs., Becher) Broad bean / Acker-/Puff-/Pferdebohne Ackergrass/Ackergras Seed of grass/Grassamenvermehrung After preparing the training and test data on site NRW for CODE-1 27 crops out of 59 crops remained.	KOD2 Winterwheat/Winterweichweizen Summer cereal cereal Winter cereal Mays/Mais Row crop mix Root-beat row crops Peas Winter rapeseed/Winterraps Fallow land (FL))/ Bracheflache Permanent grassland (PG) Forest (F) / Forstflachen, Aufforstung ALG - garden Asparagus/Spargel Strawberry /Erdbeeren Acker grass /Acker-/Puff-/Pferdebohne Multi Annual AL -TG After preparing the training and test data on site NRW for CODE-2 16 out of 27 crops remained

Table 8. Two level of grouping GSAA crops according to the image classification

The main driving factors of grouping the crops from the perspective of the best classification result is the following:

1. Always merge crops with exactly the same phenological presence and identical management technology. These are crops represented on the image time series on a very similar way, with quite an identical vegetation curve, that is why their clusters will by definition not be separated. Example: crops where seed and grain production is separate in the GSAA, or maize for corn and popcorn etc.
2. Merge crops with similar phenological presence and having relatively small area to represent (under 10% of the similar larger area). If these crops with small area are fully excluded, they will be classified into the larger class being phonologically the closest.
3. Annalise the available time series of satellite images, and merge those crops, what have short vegetation period, and it likely cannot be covered by the available images. Merging of crops cannot be always done by an automatic schema, it is also depending on the actual set of images. Example: vegetables with sweet peas, if there is no chance to catch a 3-4 weeks period.

3.4 The preparation of training and test data on the Austrian site

The preparation of training and test data on the Austrian site was implemented along the same step as for the NRW pilot-1 site. Like pilot-1 training the following sets of test data was created:

- a polygon of declared GSAA parcels,
- a number of random points proportional to the area of the polygon, with a maximum of 8 points per parcel
- an inner centroid representing the GSAA polygon.

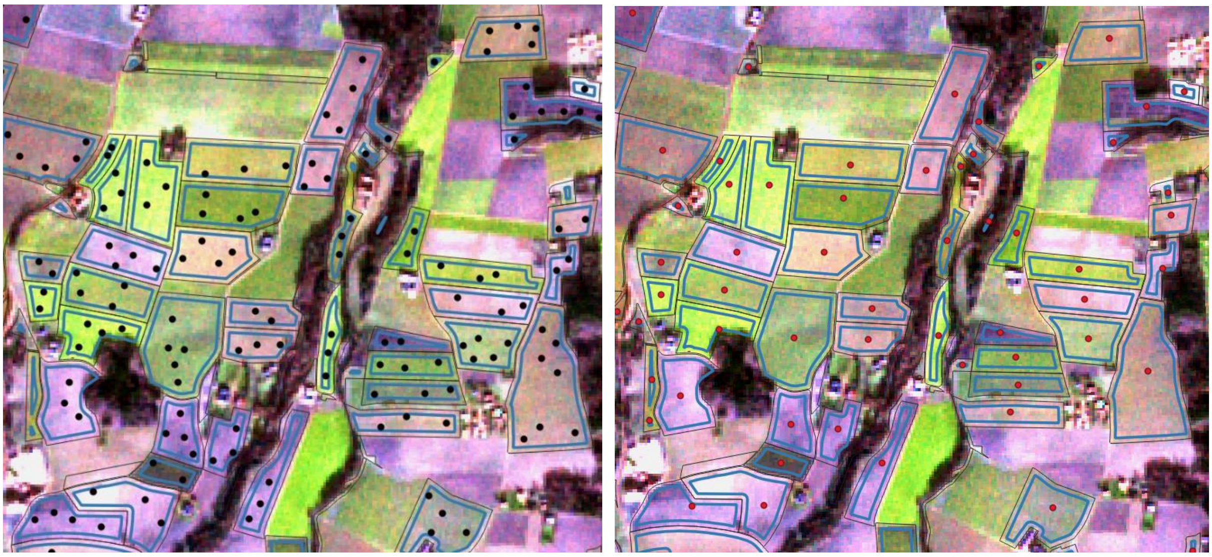


Image 19: Blue lines are representing the polygons of training data with the 20 m of buffer, black points are the result of random point generation, while the red points are the inner centroids of the polygons.
The background image is S2 R=1, B=2, G=3 – true color composite on 09-Jan-2020.

As the result of the geometrical processing to test the performance of training and test data in image classification, the following number of features was used on the NRW pilot site:

Number of features (parcels) used	GSAA polygon with Buffer 20 m, over 2000 m2	Random maximum 8 points	Inner centroid of the parcel
Training – strat. random 50%	7955	19 237 points	7955
Test – strat. random 50%	7938	19 510 points	7938

Table 9. Result of deriving the different representation of training and test data on the Austrian test site

The method of distribution of random points is the same as with the NRW test site. Since the parcels in the Austrian test site are smaller in general, the average number of points within the parcels is considerably lower.

The semantic grouping of training and test data was provided by AMA. With an applied minimum of 100 training parcels per crop, 14 different crops grouped into 11 groups were considered:

Crop Name	Crop Group	Crop Group (DE)	n features in training set
mowing meadow as permanent pasture three and more uses	mowing meadow	Mähwiese	2415
winter wheat	winter wheat	Winterweizen	1124
corn maize	corn maize	Körnermais	898
permanent pasture	permanent pasture	Dauerweide	782
winter barley	Winter barley	Wintergerste	576
silage maize	silage maize	Silomais	504
mowing meadow as permt pasture two uses	mowing meadow	Mähwiese	476
soy beans	soy beans	Sojabohnen	295
clover grass	fodder grass	Futtergräser	197
sugar-beet	Beet	Rüben	192
changing meadow	fodder grass	Futtergräser	132

green fallow land	green fallow land	Grünbrache	130
winter rapeseed	winter rapeseed	Winterraps	121
maize corn-crop Mix (CCM)	corn maize	Körnermais	113

Table 10. Type of crops and groups used on the Austrian test site

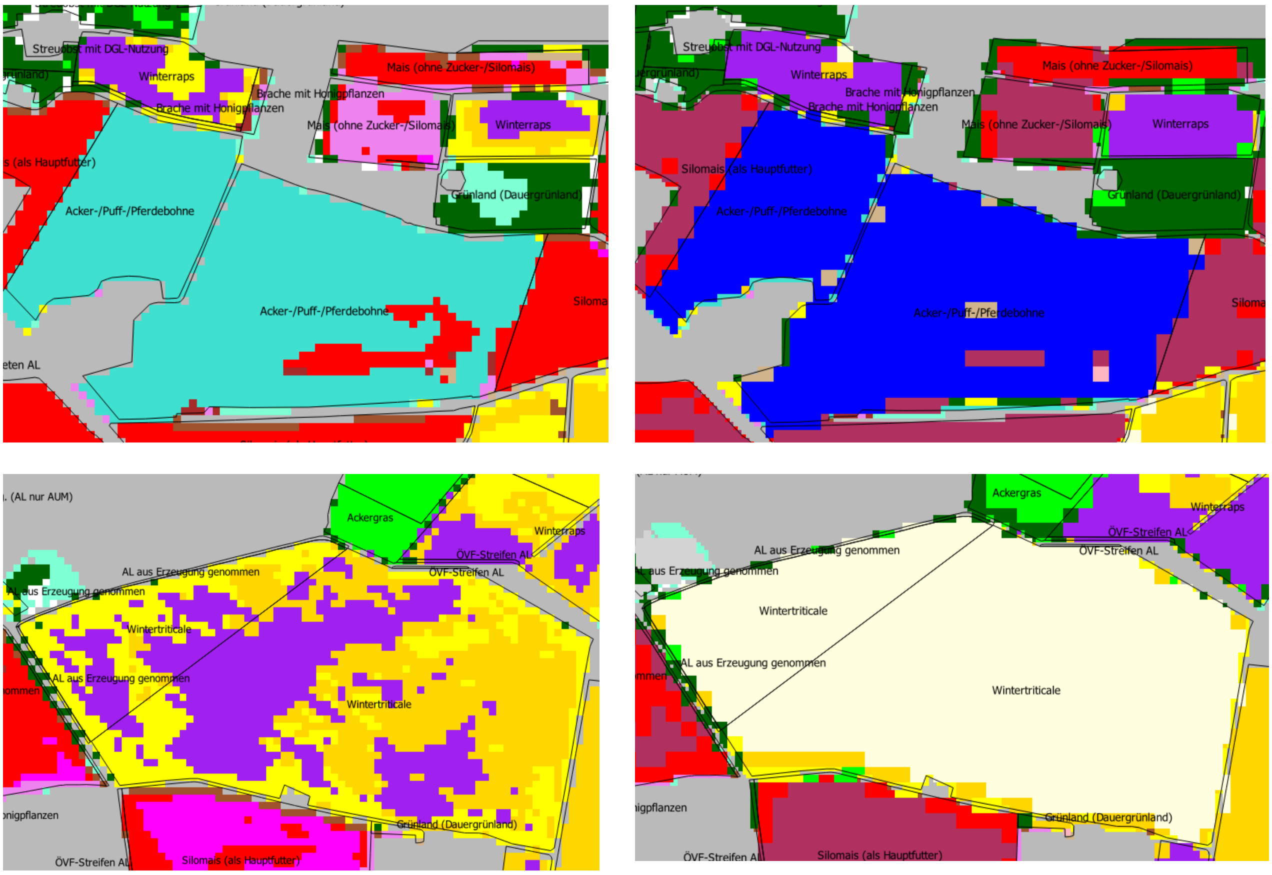
4 Performing supervised crop classification using a well-established algorithm with the use of 3 different type of generalized GSAA dataset

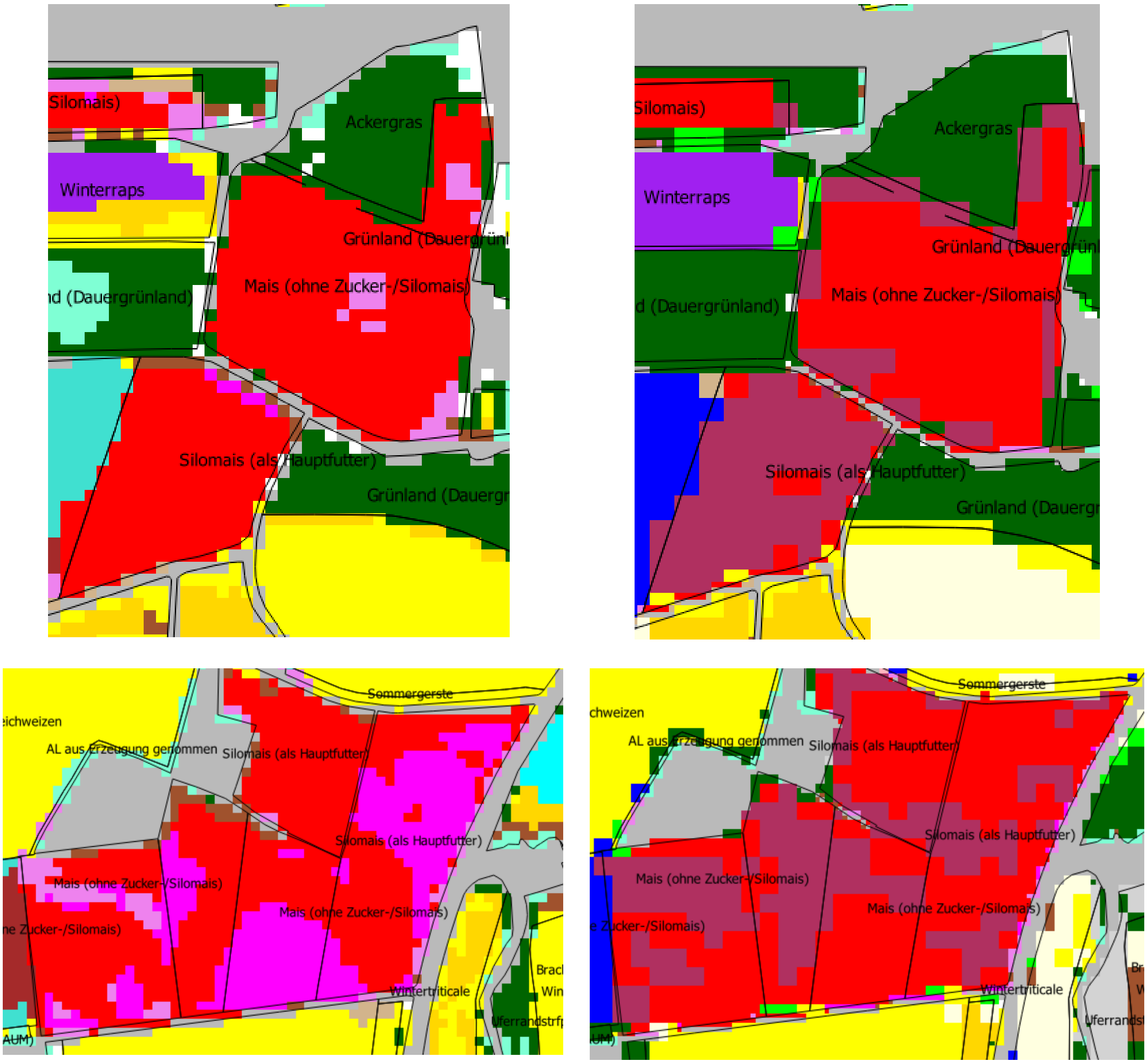
4.1 Classification methods for comparing the performance of the 3 types of generalized reference data on the NRW site

4.1.1 Testing different data preparation methods on the NRW pilot site

To clarify the final method for testing the performance of the 3 different type of training data, the following experiences had been gained:

1. **Principal component analysis**: Running a supervised pixel-based classification of the 8-multi-temporal stack of RED+NIR+SWIR bands with the use of the polygonal training data of the detailed crops. The 1st test was implemented using a maximum likelihood supervised classification and LPIS eligible area mask. The results were promising, but the need of integrating more bands was found, to detect further sophisticated differences among some crops. This led to implementing the PCA and selecting the most important 18 principal components. Due to the fact there was no cloud and haze effect on the site the use of NDVI stack was not in focus. The example reflect clearly the advantage of using PCA to derive an input spectral signature set for classification.





RED-NIR-SWIR spectral bands(image-1) *Principal component analysis run on all 8 bands (image-2)*

Image 20. Differences of classifying the same time stack of images, classifying with the same method different combination of preprocessed signal values, using the CID1, 27 crops. Legend is presented in Table 8.

2. **Segmentation:** Classification with the single centroid training data did not show satisfactory result. Despite a high number of data, the separation on cluster level of some crops was still not well represented. Several clusters describing real crop area (was proven by visual inspection) had no training representation by the reference data. The solution was to run a segmentation algorithm, which could detect quite efficiently the area around a point still belonging to that homogenic land unit, potentially forming an agricultural crop parcel. The result of the segmentation was used as training and test data of classifying the entire area. The following examples shows how the segmentation performed the detection of the crops considered the area as a homogenous unit:

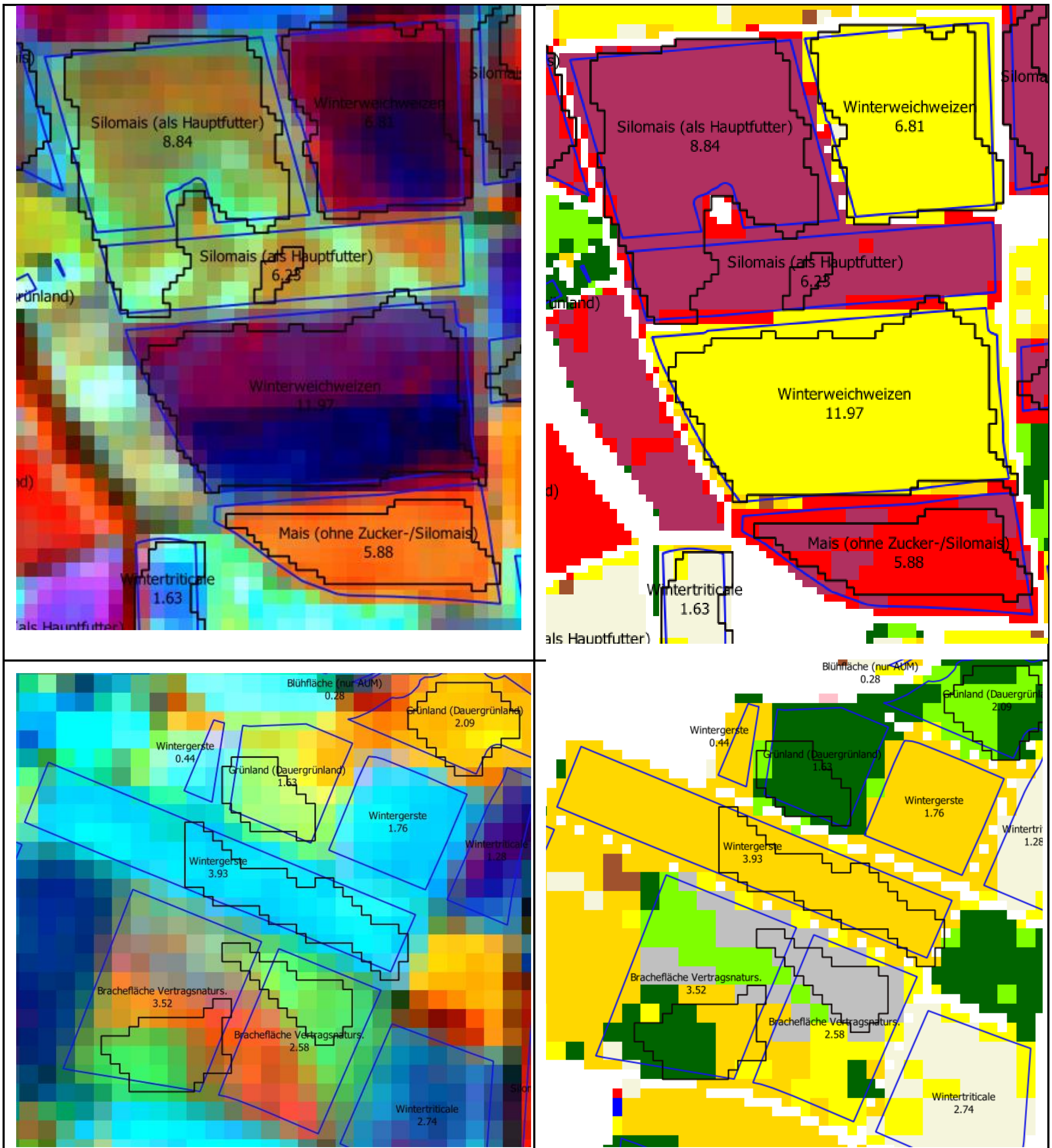


Image 21.
Blue boundary = GSAA parcel
Black boundary = segmented units
Background image: S2 - 26 June 2020 R=NIR, G=SWIR1, B=RED *false color composite*

Blue boundary = GSAA parcel
Black boundary = segmented units
Background image: cluster map of maximum likelihood on 18 principal components derived from 8 time series of S2:
winter barley = winter gerste
maize (corn for grain) = Mais (ohne Zucker- and Silomais)
silo-maize = Silomais (als Hauptfutter)
permanent grassland = grünland
fallow land = Bracheflache Vertragsnaturs
winter wheat = winterweichweizen

Image 21. Result of the segmentation around of the inner centroids of the GSAA crop parcels

On the NRW site 2 types of classifications had been implemented, with different supervised methods chosen – maximum likelihood (MXL) and random forest (RF) - to test the performance of the 3 types of reference data. In case of both classification methods 2 training sets, the detailed classes (CODE-1) and the summarized groups of the crops (CODE-2) were implemented. The result for CODE-2 was derived via summarizing the classes of the results generated for the detailed classes (based on CODE-1). The input spectral content of both classifications was the 18 principal components of S2 8 time series.

4.1.2 Pixel based maximum likelihood supervised classification

The following steps had been implemented identically, with the change of the input training data:

Order	Method	Maximum likelihood (MXL) based solution in NRW pilot
1	Selecting input satellite images	Visually selected a time series of 8 S2 images.
2	Masking area of interest	LPIS eligible area (AL/PG/PC) was used as a mask before image classification with the aim to reduce misclassification with categories out of interest
3	Selecting the most relevant spectral information	Selecting the most important 18 Principal Component among the 8 datesX9 = 72 bands
4	Defining classes with unsupervised classification	Pixel based ISODATA
5	Creating training signatures for the supervised classification	Intersecting the training data and the ISODATA cluster map
6	Supervised classification	Pixel based maximum likelihood
7	Reclassifying the final classes and creating the misclassification (confusion) matrix	Intersecting the training data and assigning the classes to the relevant crop types. Generating the final recoded class-map.
8	Accuracy assessment	50% of the reference data population is used as test data, what had been processed exactly on the same way as the 50% training. Calculated measures are integrated values (Overall accuracy, Kappa) and crop separate accuracy measures (user accuracy, producer accuracy, Kappa(i) Short (i) Hellden(i)).

Table 11: Steps of Pixel based maximum likelihood supervised classification

The result is the continuous final recoded cluster map, where the pattern of the permanent pastures along the riverside are well visible, built-in areas and forests are excluded.

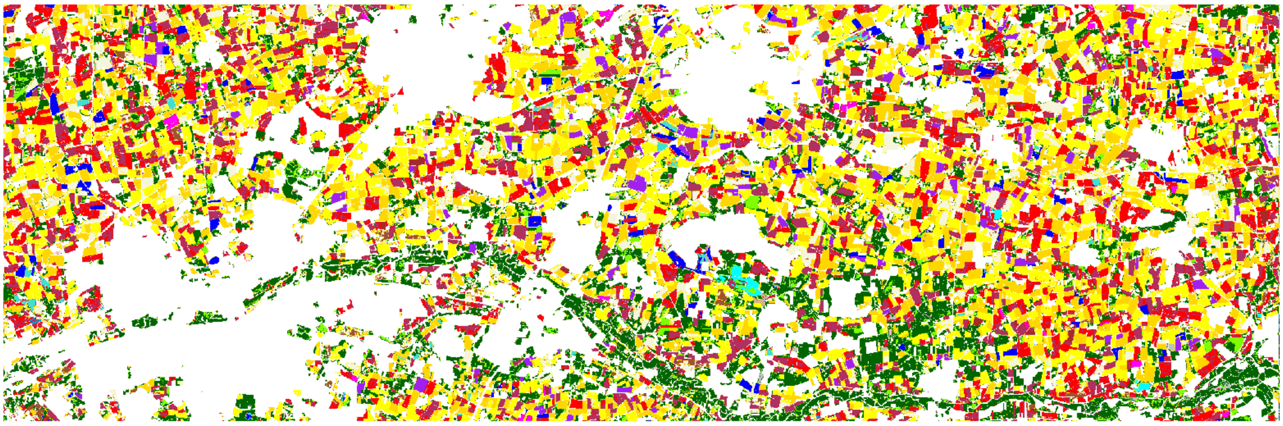


Image 22. Final recoded class-map on the south part of the NRW site, using polygon reference data and the maximum likelihood method

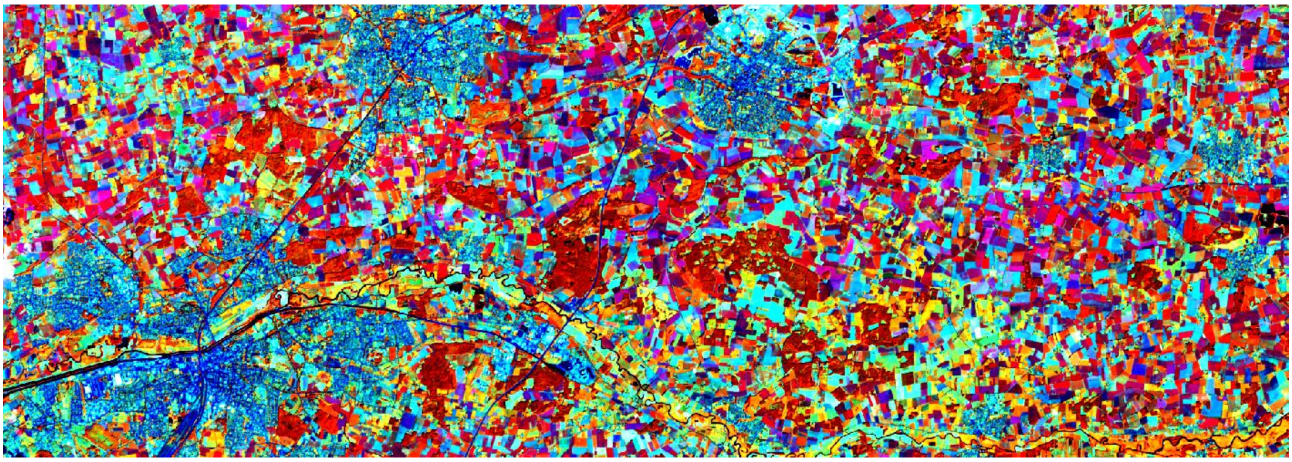


Image 23. As to compare: S2 image of 26 July 2020, using R=NIR, G=SWIR1, B=RED *false color composite*

From point 5 the steps are separately done with each type of reference data, using the same settings regarding each step, to be able to compare the result. In this chapter the examples and figures are all from the classification trained and coded by the detailed classes (see CODE-1) The following examples represent the difference of the results and the ratio of misclassifications visualized in the confusion matrix with the use of the test data.

The following examples are presenting the most common misclassifications inside a parcel in the 2 main crop categories. The annotation of the 1st polygon-based example (blue polygon) is the declared crop used as training data, while the cluster map presents the observed category on pixel level. In the test on the NRW site, the result is evaluated on pixel level, and no parcel level decision is derived. This could also be done, but only for the polygon-based version of reference data, meaning the respect of the GSAA boundaries on a higher state. The confusion matrices calculated for the test data are presented on the figure underneath. These were used to analyze the result.

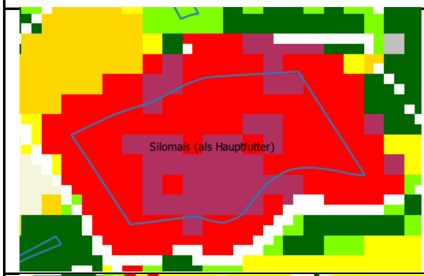
Confusion of **maize and silo maize** is somehow an “acceptable” case, due to the fact the 2 types of crop might look exactly the same in several conditions. In some regions, the maize species used for silo differs and also the technologies differ, but in several cases the only element makes it different is, if the time series detects the early-harvest of the silo-maize.

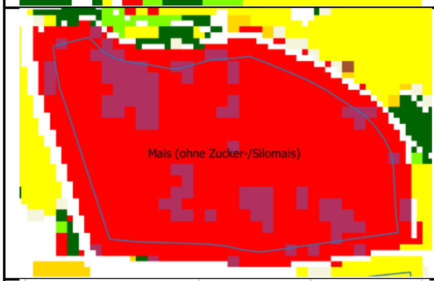
As crop separate accuracy measures shows, row crops as a group are distinguishable. Here are the crop group values of MXL summarized by CODE2 with the use of polygon training and test data:

	user	producer	Hellden(i)	Short(i)
Mais	99,2	97,7	98,48	97,0
ROWMIX	76,8	85,4	80,89	67,9
ROOT-types	97,4	99,4	98,40	96,9

The examples of parcels show the misclassifications of each crop type, according to CODE1:

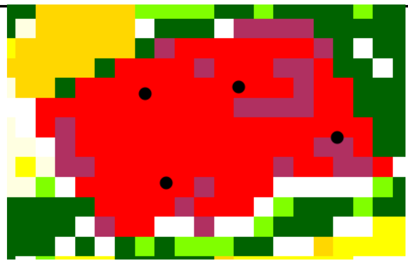
TRAIN = POLY

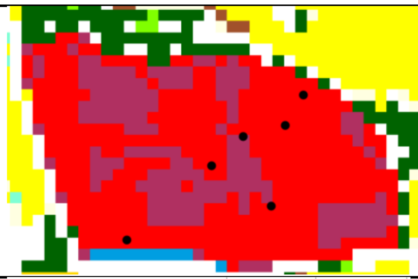




	% Maize	% Silo-maize
Winter wheat	0,25%	0,05%
winter barley	0,05%	0,12%
rye	0,05%	0,03%
Triticale	0,07%	0,05%
Maize	62,91%	35,65%
Silo-maize	36,38%	77,87%
Soybeen	0,07%	0,10%
Strawberry	0,02%	
Trifolii		0,02%
Acre grass	0,02%	0,17%

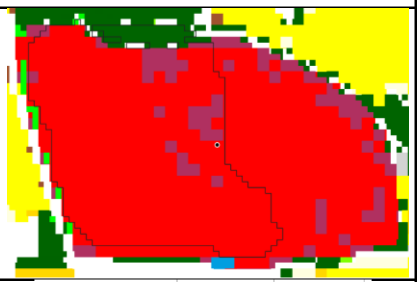
TRAIN = MAX. 8 POINTS





	% Maize	% Silo-maize
Winter wheat	0,10%	0,02%
winter barley	0,13%	0,11%
rye	0,05%	0,02%
Triticale	0,05%	0,07%
Maize	59,24%	28,38%
Silo-maize	39,93%	70,62%
rape seed		0,02%
PP	0,16%	0,15%
Forest		0,02%
Soybeen	0,24%	0,24%
Asparagus/Spargel	0,05%	0,11%
Acker-/Pferdebohne	0,03%	0,02%
Acre grass	0,03%	0,22%

TRAIN = CENTROID +SEGMENTS



	% Maize	% Silo-maize
Winter wheat		0,05%
winter barly	0,07%	0,11%
rye	0,07%	
triticale	0,14%	0,22%
Maize	**60,90%**	**28,14%**
Siolomaize	**38,54%**	**70,98%**
potato		0,16%
rape seed		0,05%
PP	0,14%	0,05%
strawberry	0,07%	
acre grass	0,07%	0,22%
SUM:	100,00%	100,00%

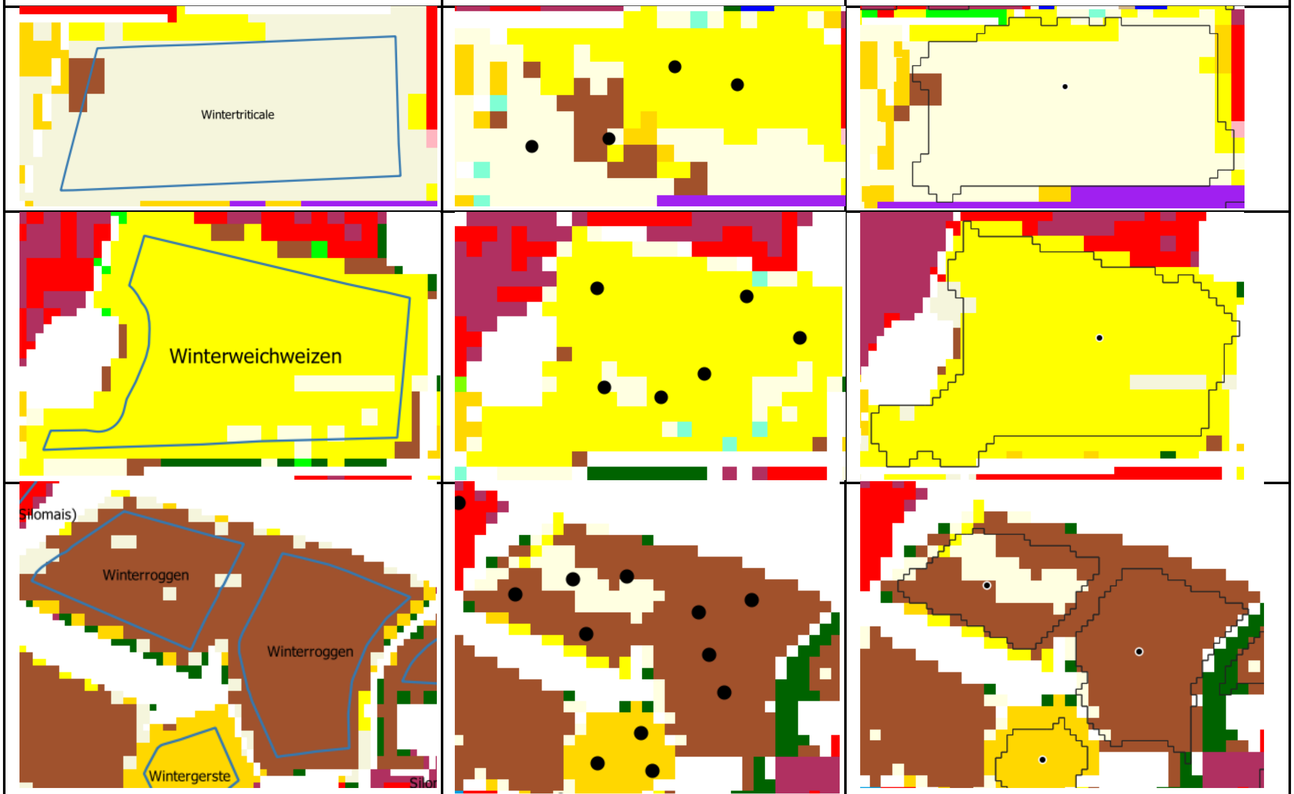
The mismatch of the **winter cereals** is also a common issue, but the separation in some context is clearly possible. These are the details depending fully on the combination of spectral content, which cannot be really followed by visual inspection. In the current example winter wheat triticale and rye are mismatched.

As crop separate accuracy measures shows, winter cereals are the best distinguishable. Here are the crop group values of MXL summarized by CODE2 with the use of polygon training and test data:

	user	producer	Hellden(i)	Short(i)
WINTER_WHEAT	92,3	94,4	93,30	87,4
SPRING_CEREALS	51,3	86,8	64,47	47,6
WINTER_CEREALS	95,8	93,0	94,36	89,3

The examples of parcels show the misclassifications of each crop type, according to CODE1:

TRAIN = POLY	TRAIN = MAX. 8 POINTS	TRAIN = CENTROID +SEGMENTS



TRAIN = POLY:

	winter wheat	summer wheat	summer barley	winter barley	rye	oat	triticale
winter wheat	92,34%	77,08%	10,82%	0,64%	4,93%	6,58%	7,65%
summer barley			30,86%	0,00%	0,00%	6,43%	0,00%
winter barley	0,82%		18,24%	95,33%	1,21%	3,34%	2,24%
rye	1,14%		2,54%	0,46%	70,18%	2,43%	5,44%
oat		1,62%	27,25%	0,02%	0,01%	60,90%	0,01%
triticale	5,38%	8,33%	3,32%	2,81%	20,61%	3,16%	84,25%

TRAIN = MAX. 8 POINTS:

	winter wheat	summer wheat	summer barley	winter barley	rye	oat	triticale
winter wheat	90,89%	68,42%	13,27%	0,93%	6,86%	11,02%	10,04%
summer barley	0,00%		15,31%	0,00%	0,00%	0,79%	0,00%
winter barley	0,61%		18,37%	94,69%	1,86%	1,57%	2,32%
rye	0,59%		3,06%	0,32%	65,88%	1,57%	5,75%
oat	0,00%		26,53%	0,00%	0,00%	62,20%	0,00%
triticale	5,53%	5,26%	2,04%	3,08%	21,47%	3,15%	80,87%

TRAIN = CENTROID +SEGMENTS:

	winter wheat	summer wheat	summer barley	winter barley	rye	oat	triticale
winter wheat	90,76%	50,00%	23,08%	1,09%	6,76%	8,20%	10,53%
summer barley	0,00%		9,62%	0,00%	0,00%	0,00%	0,00%
winter barley	1,08%		21,15%	94,57%	2,25%	4,92%	4,03%
rye	1,03%		0,00%	0,60%	65,37%	3,28%	6,72%
oat	0,00%		17,31%	0,06%	0,00%	50,82%	0,00%
triticale	6,66%	12,50%	5,77%	2,90%	23,16%	6,56%	78,05%

A typical crop of the region is **Acre grass**. So typical, that it was classified quite accurately. That is why it was a separate class even in the CODE2 level of grouping. Meanwhile other **graminoid dominant crops** in multi-annual cultivation are usually mixed up with permanent grasslands (PG) and with fallow land as well. Another aspect is, that when Acre grass is cut, it looks like a cereal stubble, but the grass regenerates quite fast.

As crop separate accuracy measures shows, winter cereals are the best distinguishable. Here are the crop group values of MXL summarized by CODE2 with the use of polygon training and test data:

	user	producer	Hellden(i)	Short(i)
Fallow land	14,5	44,2	21,79	12,2
PG	92,6	84,7	88,50	79,4
Forest	88,5	99,5	93,68	88,1
MassAAL-grass inc. Acregr	46,6	47,1	46,9	30,6

The examples of parcels show the misclassifications of each crop type, according to CODE1:

TRAIN = POLY	TRAIN = MAX. 8 POINTS	TRAIN = CENTROID +SEGMENTS

TRAIN = POLY

	FL	PG	Acreg.
winter wheat	4,55%	0,03%	0,38%
winter barley	1,04%	0,01%	0,06%
rye	0,89%	0,01%	0,04%
triticale	7,52%	0,01%	0,02%
maize	2,28%	0,04%	0,58%
silomaize	16,09%	0,06%	0,40%
potato	5,64%	0,01%	0,06%
FL	15,30%	0,19%	0,02%
PG	35,54%	92,41%	46,18%
strawberry	1,24%	0,02%	0,13%
clover	1,49%	0,12%	1,09%
Acreg.	3,81%	7,09%	50,89%

TRAIN = MAX. 8 POINTS

	FL	PG	Acreg.
winter wheat	2,22%	0,00%	0,16%
winter barley	3,70%	0,00%	0,00%
rye	2,96%	0,02%	0,00%
triticale	5,93%	0,02%	0,00%
maize	4,44%	0,11%	0,63%
silomaize	5,93%	0,09%	0,63%
potato	7,41%	0,06%	0,00%
PG	57,78%	94,50%	57,50%
strawberry	3,70%	0,06%	1,56%
broad bean	4,44%	0,00%	0,00%
Acreg.	1,48%	5,07%	39,38%

TRAIN = CENTROID +SEGMENTS

	FL	PG	Acreg.
winter wheat	11,72%	0,33%	1,25%
winter barley	12,50%	0,72%	1,25%
rye	0,00%	0,11%	0,25%
triticale	4,69%	0,11%	0,25%
maize	9,38%	0,52%	2,99%
silomaize	6,25%	0,25%	1,00%
potato	4,69%	0,03%	0,25%
PC	0,00%	0,33%	0,00%
FL	5,47%	0,11%	0,00%
PG	37,50%	88,06%	54,11%
clover	0,00%	0,03%	0,25%
broad bean	3,13%	0,00%	0,00%
Acreg.	3,91%	9,41%	38,40%

Table 13. Examples of classifications per crops showing the difference among the different training data

The misclassification matrixes above only present the % of the mixed pixels related to the test data of the given crop = user accuracy (presented by columns) but do not fully present how the classified crops are mixing with other crops of the test data = producer's accuracy (presented by rows, but only for 3 crop types).

As an example, 2 misclassification matrixes are visualizing 2 class-related accuracy values for the polygon based maximum likelihood classification regarding all the crops, summarizing into a single group the types not have been visualized among the parcels in this chapter. This "all other crop" category contains all the crops of CODE1 categorization what are not nominated in the matrix.

Threshold metrics are to quantify the classification prediction errors, designed to summarize the fraction, ratio, or rate of when a predicted class does not match the expected class in a holdout dataset.

Producer Accuracy	winter wheat	summer wheat	summer barley	winter barley	rye	oat	triticale	maize	silo-maize	fallow land	PG	acre-grass	all other crop types	Prod. Acc.
winter wheat	94,26%	0,43%	0,14%	0,49%	0,76%	0,12%	3,01%	0,16%	0,04%	0,06%	0,01%	0,03%	0,48%	94,26%
summer barley	0,46%	0,23%	72,34%	0,34%	0,00%	20,23%	0,00%	0,00%	0,00%	0,34%	0,11%	0,00%	6,97%	72,34%
winter barley	1,10%	0,00%	0,32%	96,47%	0,25%	0,08%	1,16%	0,04%	0,10%	0,02%	0,01%	0,01%	0,49%	96,47%
rye	7,88%	0,00%	0,23%	2,37%	73,56%	0,30%	14,51%	0,23%	0,15%	0,08%	0,03%	0,02%	0,84%	73,56%
oat	0,34%	0,53%	21,30%	0,76%	0,11%	63,85%	0,19%	0,00%	0,00%	0,15%	0,00%	0,00%	14,10%	63,85%
triticale	12,22%	0,10%	0,10%	4,80%	7,09%	0,13%	73,69%	0,10%	0,08%	0,22%	0,01%	0,00%	1,42%	73,69%
maize	0,24%	0,01%	0,00%	0,00%	0,12%	0,02%	0,01%	62,37%	35,35%	0,04%	0,03%	0,07%	1,76%	62,37%
silo-maize	0,15%	0,01%	0,04%	0,01%	0,02%	0,00%	0,04%	31,01%	66,37%	0,27%	0,04%	0,04%	1,97%	66,37%
fallow land	0,15%	0,00%	0,00%	0,00%	0,15%	0,00%	0,00%	0,44%	0,00%	45,78%	21,78%	0,30%	32,00%	45,78%
PG	0,03%	0,00%	0,06%	0,40%	0,57%	0,09%	0,17%	0,07%	0,20%	0,85%	84,73%	6,97%	5,93%	84,73%
acre-grass	0,03%	0,00%	0,00%	0,30%	0,25%	0,46%	0,18%	0,10%	1,17%	0,52%	36,86%	43,58%	16,45%	43,58%
all other crop types	0,13%	0,21%	0,12%	1,25%	0,12%	0,77%	0,06%	0,57%	0,78%	0,65%	0,32%	0,47%	95,01%	95,01%

User Accuracy	winter wheat	summer wheat	summer barley	winter barley	rye	oat	triticale	maize	silo-maize	fallow land	PG	acre-grass	all other crop types
winter wheat	92,34%	77,08%	10,82%	0,64%	4,93%	6,58%	7,65%	0,25%	0,05%	4,55%	0,03%	0,38%	1,43%
summer barley	0,00%	0,23%	30,86%	0,00%	0,00%	6,43%	0,00%	0,00%	0,00%	0,15%	0,00%	0,00%	0,12%
winter barley	0,82%	0,00%	18,24%	95,33%	1,21%	3,34%	2,24%	0,05%	0,10%	1,04%	0,01%	0,06%	1,10%
rye	1,14%	0,00%	2,54%	0,46%	70,18%	2,43%	5,44%	0,05%	0,03%	0,89%	0,01%	0,04%	0,36%
oat	0,01%	1,62%	27,25%	0,02%	0,01%	60,90%	0,01%	0,00%	0,00%	0,20%	0,00%	0,00%	0,71%
triticale	5,38%	8,33%	3,32%	2,81%	20,61%	3,16%	84,25%	0,07%	0,05%	7,52%	0,01%	0,02%	1,89%
maize	0,16%	1,16%	0,15%	0,00%	0,54%	0,84%	0,02%	62,91%	31,16%	2,28%	0,04%	0,58%	3,47%
silo-maize	0,11%	1,85%	2,10%	0,01%	0,12%	0,18%	0,08%	36,38%	68,05%	16,09%	0,06%	0,40%	4,53%
fallow land	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	15,30%	0,19%	0,02%	0,42%
PG	0,01%	0,12%	2,39%	0,29%	2,03%	2,69%	0,23%	0,06%	0,15%	35,54%	92,41%	46,18%	9,69%
acre-grass	0,00%	0,00%	0,00%	0,04%	0,16%	2,47%	0,04%	0,01%	0,15%	3,81%	7,09%	50,89%	4,74%
all other crop types	0,03%	9,61%	2,34%	0,41%	0,20%	10,97%	0,04%	0,22%	0,26%	12,62%	0,16%	1,43%	71,51%
User ACC.	92,34%	0,00%	30,86%	95,33%	70,18%	60,90%	84,25%	62,91%	68,05%	15,30%	92,41%	50,89%	71,51%

Table 14. User and producer accuracy values for the polygon based maximum likelihood classification

LPIS eligible area was used as a mask before image classification, that is why parcels not being part of the y2020 declaration had also been successfully classified:

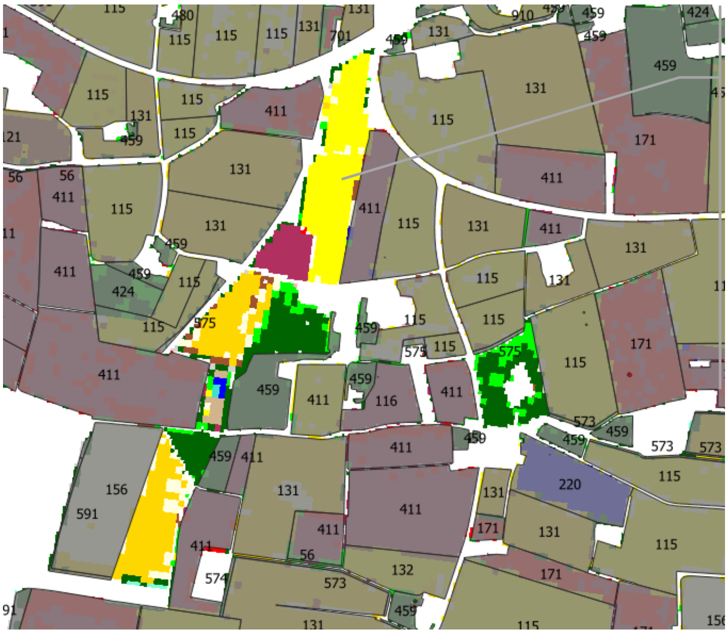


Image 24. Difference in extent of the derived cropmap and of the declared parcels in IACS - GSAA

The transparent gray vector layer shows the incontinuity of the declared parcels, overlapped with the crop classification cluster map result.

The yellow parcels are winter cereals, the green represents the permanent pastures, and the maroon is silo-maize.

4.1.3 Pixel-based Random Forest supervised classification

The second version of the test on the NRW site was a classification using pixel-based Random Forest (RF). In course of classification RF is sampling the input training data with points. It means, that both the GSAA polygons and the segmented units derived around the inner centroid of the crop parcels had been resampled by the RF algorithm such as the centroid points. The user cannot parametrize the resampling on a way, that balanced representation of a crop could be reached. .

"Choosing an appropriate metric is challenging generally in applied machine learning, but is particularly difficult for imbalanced classification problems. Firstly, because most of the standard metrics that are widely used assume a balanced class distribution, and because typically not all classes, and therefore, not all prediction errors, are equal for imbalanced classification.

Decision trees are a simple and powerful predictive modeling technique, but they suffer from high-variance. This means that trees can get very different results given different training data. A technique to make decision trees more robust and to achieve better performance is called bootstrap aggregation or bagging for short. Bagging is an ensemble algorithm that fits multiple models on different subsets of a training dataset, then combines the predictions from all models. [Random forest] is an extension of bagging that also randomly selects subsets of features used in each data sample.

Random forest involves selecting bootstrap samples from the training dataset and fitting a decision tree on each. The main difference is that all features (variables or columns) are not used; instead, a small, randomly selected subset of features (columns) is chosen for each bootstrap sample. This has the effect of de-correlating the decision trees (making them more independent), and in turn, improving the ensemble prediction. (J. Brownlee, 2020 – source: https://machinelearningmastery.com/bagging-and-random-forest-for-imbalanced-classification/)"

The following example represents how strongly imbalanced is the sampling made by the standard setting of the RF classifier.

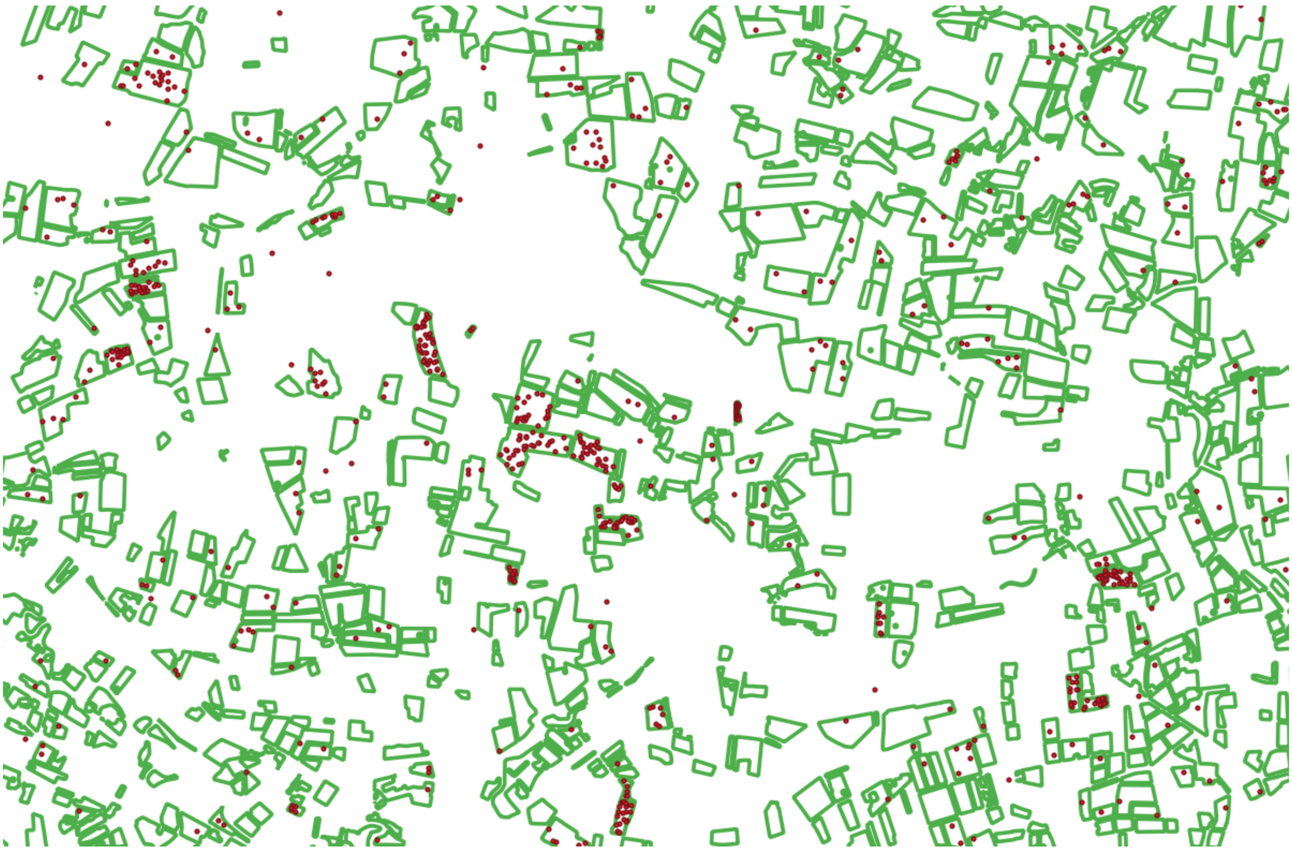


Image 25: Imbalanced sampling of the random forest model: red points are generated by the RF model as training signals, inside the entire training data (green crop boundaries).

A simple technique for modifying a decision tree for imbalanced classification is to change the weight that each class has when calculating the "*impurity*" score of a chosen split point. An easy way to overcome class imbalance problem when facing the resampling is to take the classes of the instances into account when training data are randomly selected from the original dataset. Another approach could be to use random independent resampling of the majority classes (crops) to create multiple datasets per crop categories with a balanced class distribution.

In the current study 3 types of reference data were classified with the same settings of the basic RF model, that is why the result can be compared. RF was run for the 3 types of reference data, using the same 18 principal component image and identical settings regarding each step, to be able to compare the result. The steps of implementation were the following:

Order	Method	Random Forest (RF) based solution in NRW pilot
1	Selecting input satellite images	Visually selected 8 time series of S2.
2	Masking area of interest	LPIS eligible area (AL/PG/PC) was used as a mask before image classification with the aim to reduce misclassification with categories out of interest.
3	Selecting the most relevant spectral information	Selecting the most important 18 Principal Component among the 8 datesX9 = 72 bands
4	Supervised classification	Random forest, number of samples was set to 500 and number of trees = 1000. The number of trees does not depend on the size of the area, but more on the diversity, what means partially the number of classes, but also the inner variability of a class, meaning the crop in our case. With higher tree number the oversampling regarding the large clusters will create less negative effect.
5	Accuracy assessment	50% of the reference data population is used as test data, which has been processed exactly on the same way as the 50% training. Calculated measures are integrated values (Overall accuracy, Kappa) and crop separate accuracy measures (user accuracy, producer accuracy, Kappa(i) Short (i) Hellden(i)).

Table 15: Steps of Pixel based random forest supervised classification

The RF result was compared to the result of the MXL classification. At the current stage of the analysis the advantage of the MXL method is that settings to increase the accuracy specified for a given class are available. In case of RF further investigation is necessary. The following overview examples represent the differences of the results:

Here is an example how the crops are distinguishable with different clustering methods. Both image classification process had used the same polygon training data (blue boundaries) and the detailed crop classes (only CODE1 level of grouping).

Result of random forest (RF) *Result of maximum likelihood (MXL)*

Image 26: Comparing the result of the two image classification method on the same area, with the same input dataset

❶ Broad bean (Acker-/Puff-/Pferdebohne) performed on a very similar way.

❷ Peas (Gemüseerbse) are short period crops with smaller sum area, that is why at MXL the chance that pea pixels joins maize is higher. It proofs the theory that Rf is not so sensitive for the area share of the reference, while MXL is. The result of the RF proofs the fact that most probably PCA was suitable to select the spectral information to detect the pea parcels.

❸ winter wheat and triticale is not an easy task to separate. In case of ❸/1 the MXL found the triticale as the observed crop but on case of ❸/2 example the RF detected the winter wheat as it is in the test.

The following example is the strangest finding of the comparison: mixture of Permanent crops (PC) and of fallow land was systematically classified by RF on locations where, in the reality, there are permanent grasslands along the river. The MXL classification detected well the PG areas.

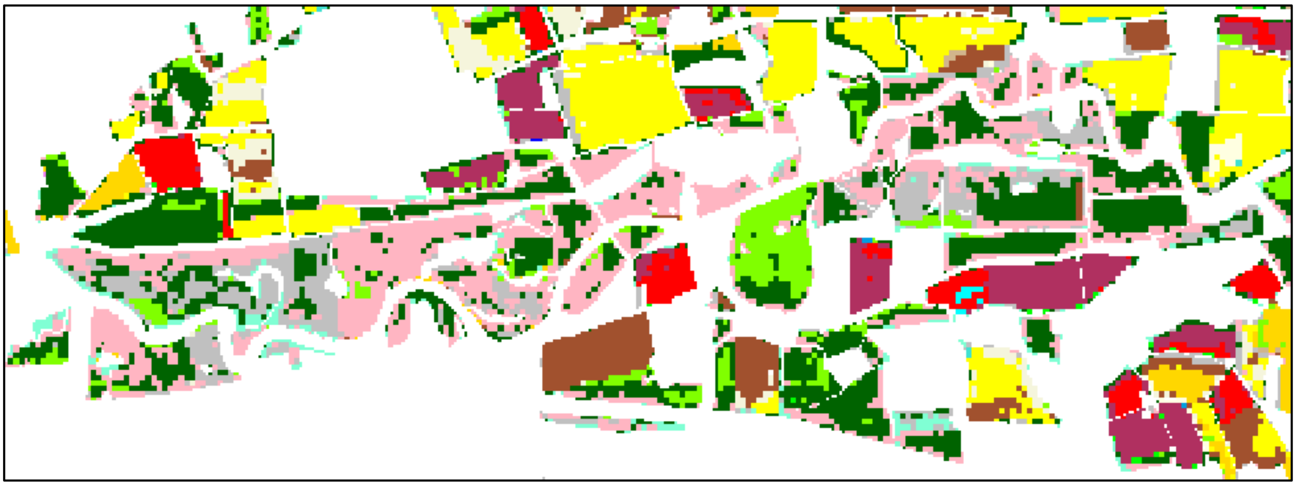


Image 27: RF – polygon reference data is used

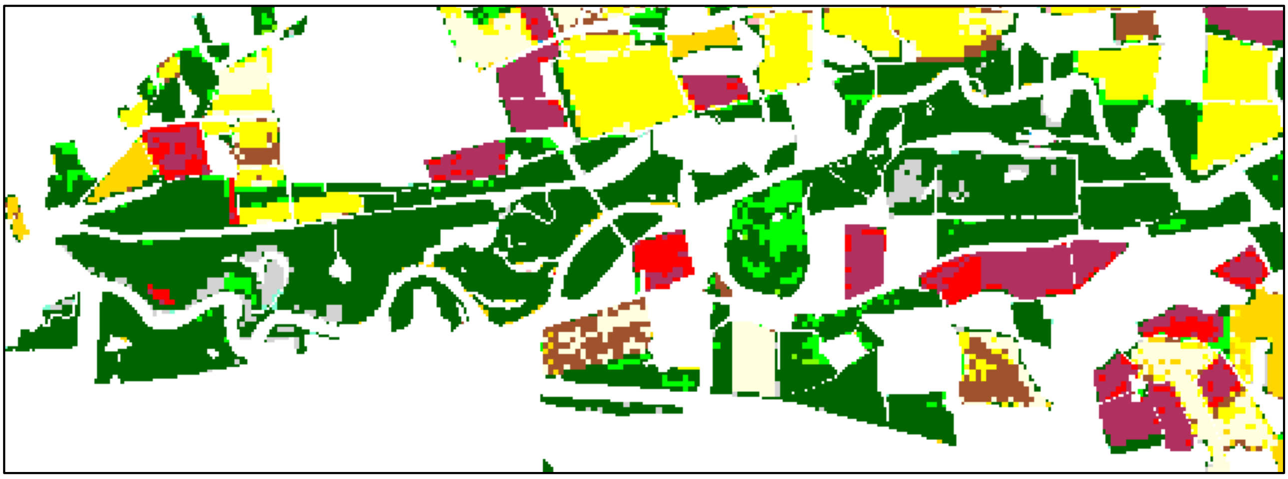


Image 28: MXL – polygon reference data is used

At the current stage the “black box effect” of RF makes it difficult how to improve the settings or the input data according to measurable parameters. The current result was reached via not automatized empirical experiments. High number of trees (1000) were chosen believing that it will be more sensitive to the large variety of crops.

The detailed CODE1 was used for classification, and the result for groups of COD2 classes were created via summarizing the classes of the 1st cluster map afterwards. CODE 2 level of grouping crops is built upon CODE 1 level. This ensures, that summarizing the predicted CODE1 classes along the rule as CODE 2 level is derived will lead to the predicted CODE2 classes. Due to the fact summarizing of CODE2 usually groups similar crops, the summed categories will lead to a clearer crop prediction map, what will be more accurate, as contains less mixing of crops. This method ensures us to study the importance and effectiveness of crop grouping as a single issue. If a new classification run would be implemented for CODE2 crop groups, the difference caused by a new random selection of training signatures may vary the result given the stochastic nature of the algorithm.

Further investigation would need to clarify how the own sample set of RF and the trees can be saved, what could lead to repeatable test solutions and achieving better classification accuracy. In this pilot, however, the target was just comparing how the different training samples perform.

4.2 Accuracy assessment and comparing the results of the classifications using the misclassification matrix on the NRW site

4.2.1 Integrated accuracy values on sample level

For accuracy testing the other 50% of the GSAA reference data had been used. In geometry the following training and test data combinations were used:

Type of reference data	***Training***	***Test***	***Value in the accuracy stats***
GSAA polygon (poly)	Poly	Poly	Area in ha, %
Random, maximum 8 points inside GSAA polygons (rpoint_max8)	10 meters of buffer around the Random points	10 meters of buffer around the random points	Number of pixels overlapping, %
Inner centroids of GSAA polygons (cpoint_segm)	Segmented units around inner centroid points	10 meters of buffer around the inner centroid points	Number of pixels overlapping, %

Table 16. Training and test data used by the three types of reference data

Segmentation of the test data was not implemented, to keep the original information source.

As the 1st step the confusion matrices for each 12 combinations were created. The combinations are the following:

- 2 types of classification method: RF and maximum likelihood (MXL)
- 3 types of test data derived from the GSAA: polygon and maximum 8 random point and inner centroid point and segmentation
- 2 versions of semantic grouping: CODE1 and CODE2.

Since different measures incorporate different information of the confusion matrix, accuracy for the entire sample is evaluated by the following integrated values[ii]:

- ***Overall accuracy:*** the sum of the correct pixels divided by the total number of pixels,
- ***Kappa***: used to be a widely used accuracy measure in remote sensing, can be understood as a measure of agreement between prediction and reality, determines if the values in a confusion matrix are significantly better than a randomly obtained one. Kappa is calculated for the NRW pilot site, to study the conclusions of Pontius-Millonesiii, stating that "*Kappa indices are useless, misleading and/or flawed for the practical applications in remote sensing. … abandon the use of Kappa indices for purposes of accuracy assessment and map comparison, and instead summarize the cross-tabulation matrix with two much simpler summary parameters: quantity disagreement and allocation disagreement.*"

As expected, the parcel-based reference data performed the best accuracy values of the classification. Further grouping of similar crops resulted in much higher -increase of more than 10% - accuracy values.

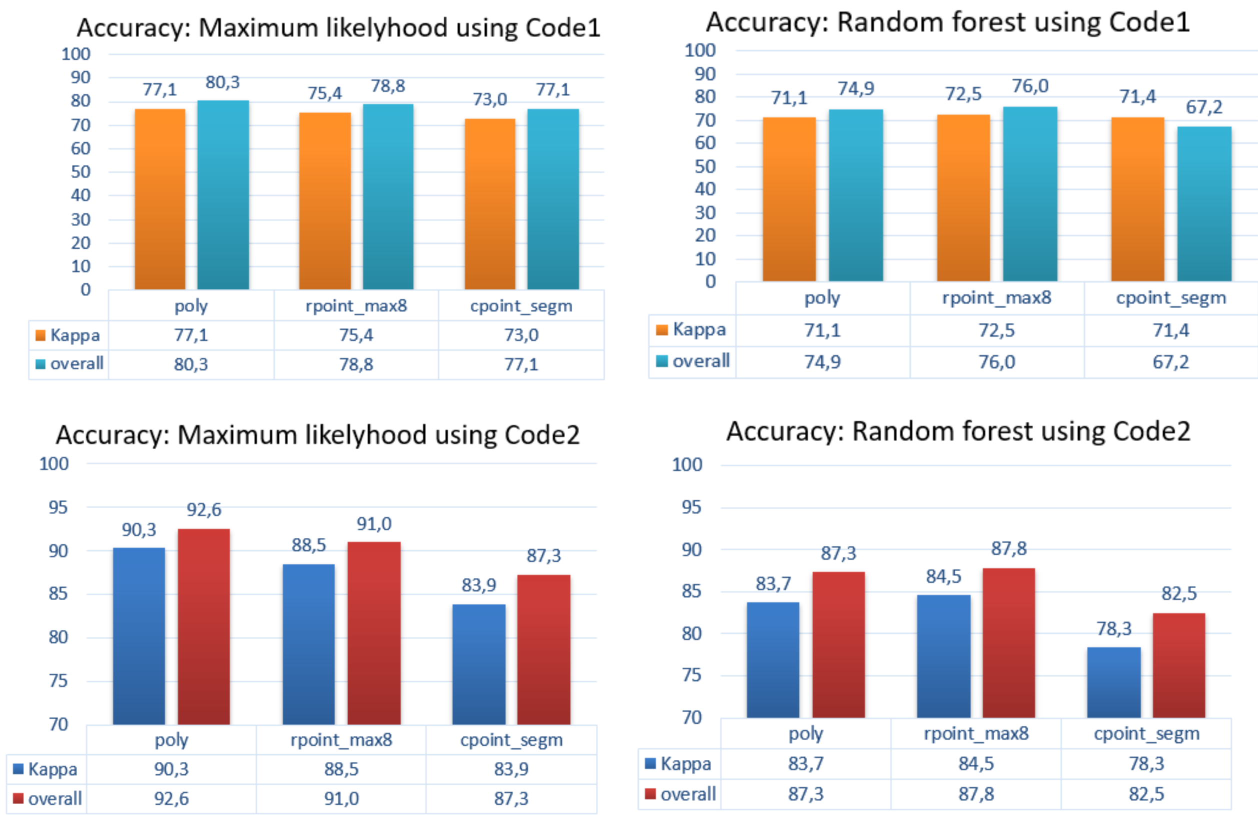


Image 29. Sample level summarized accuracy measures

4.2.2 Crop specific accuracy measures

To be able to compare the accuracy of classifications, where only one input data/parameter is changed, the calculation of crop specific accuracy measures is more suitable, than the average measures itself[iv]:

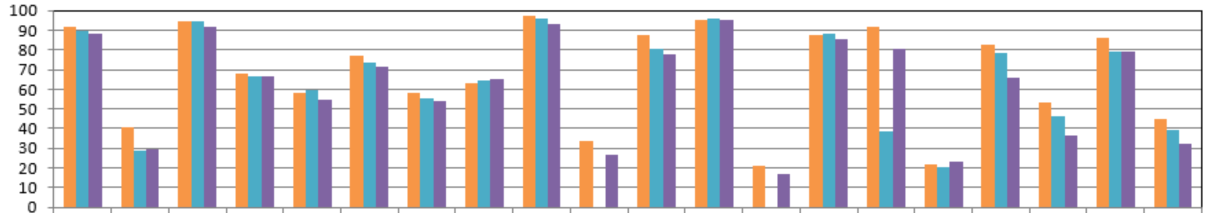
- ***User*** accuracy indicates the probability that a pixel on the class-map represents the same class on the test area.

- ***Producer accuracy:*** the sum of the correct pixels divided by the total number of pixels per each category classified separately, showing the probability of a pixel being correctly classified.
- ***Kappa(i):*** highlights only those instances that may have been correctly classified by chance. It is calculated using both the observed (total) accuracy and the random accuracy.
- ***Short (i):*** a class-specific symmetric measure defined as the ratio of the intersection of estimated and true classes to their union (in terms of set cardinality).
- ***Hellden(i):*** mean accuracy index; a measure of overlapping between the true and estimated classes (other instances, e.g. "true negative" (TN), are ignored).
- **Average accuracy:** means the simple mathematical average of the following crop specific accuracy measures: user accuracy, producer accuracy, Kappa(i) Short (i) Hellden(i).

The conclusion is, that probably the classification algorithm chosen determines more the difference in individual class accuracy than the type of reference data.

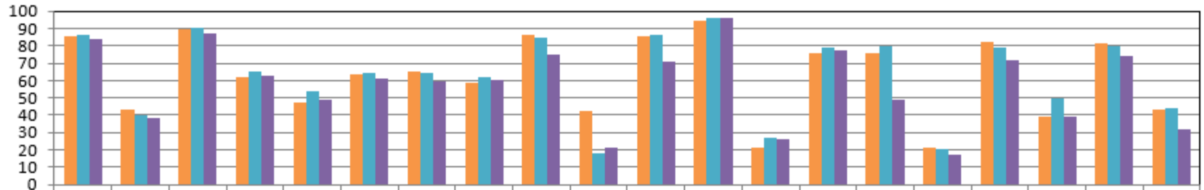
The following figures contain the "average accuracy":

Accuracy Code1 Maximum likelyhood



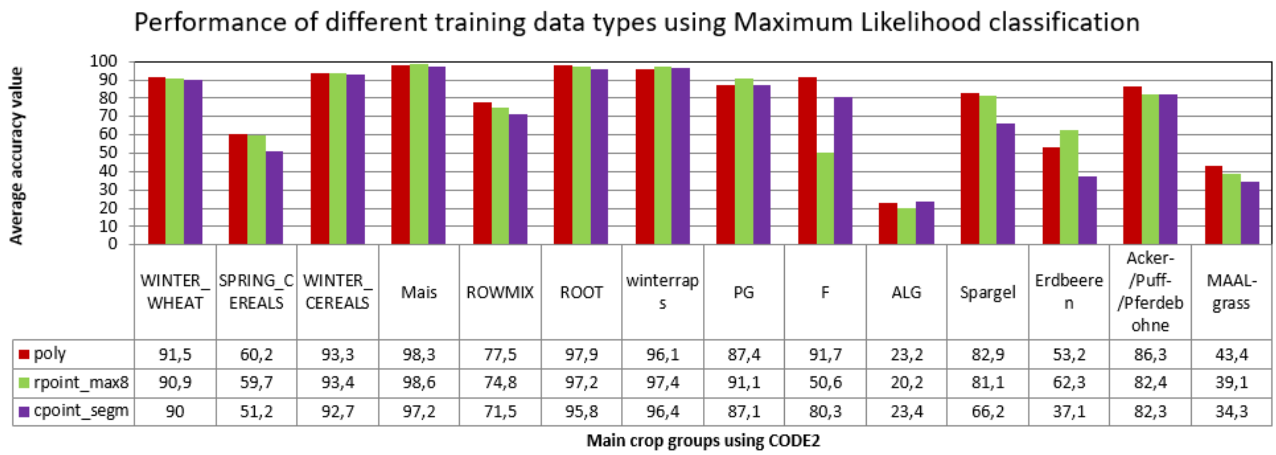
	Winter weich weizen	Sommer gerste	Winter gerste	Winter roggen	Sommerhafer	Winter triticale	Mais	Silomais	Zucker rüben	Peas	Kartoffeln	Winter raps	FL	PG	F	Gemenge Legumi nosen / Getreide	Spargel	Erdbeeren	Acker-/Puff-/Pferdebohne	Ackergras
poly	91,5	40,8	94,9	68,1	58,4	76,8	58,1	62,9	97,2	33,8	87,7	95,4	21,4	87,3	91,7	21,8	82,9	52,9	86,2	44,6
rpoint_max8	89,5	28,9	94,4	66,8	59,7	73,7	55,3	64,1	95,7	0,0	80,7	95,7	0,0	88,3	38,8	20,2	78,4	46,3	79,0	39,2
cpoint_segm	88,1	29,4	91,6	66,8	54,6	71,7	54,0	65,0	93,2	27,0	77,4	95,6	17,0	85,3	80,3	23,3	65,9	36,3	79,2	32,0

Accuracy Code1 Random forest

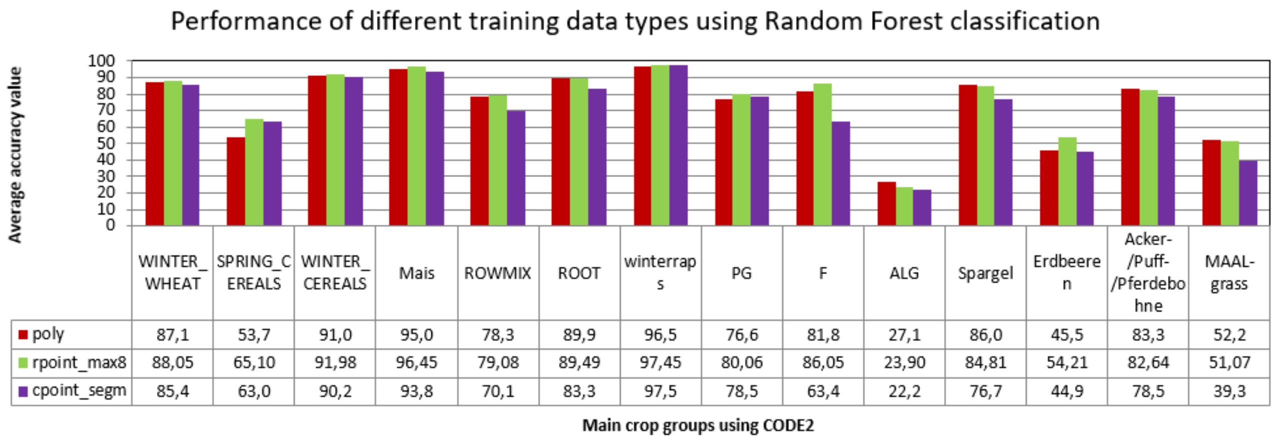


	Winter weich weizen	Sommer gerste	Winter gerste	Winter roggen	Sommerhafer	Winter triticale	Mais	Silomais	Zucker rüben	Peas	Kartoffeln	Winter raps	FL	PG	F	Gemenge Legum inosen / Getreide	Spargel	Erdbeeren	Acker-/Puff-/Pferdebohne	Ackergras
poly	85,8	43,1	89,6	62,2	47,1	63,8	64,9	58,7	86,2	42,2	85,2	94,3	21,5	75,6	75,4	21,1	82,5	39,5	81,8	43,1
rpoint_max8	86,7	40,0	90,2	64,9	53,7	64,6	64,8	62,3	84,7	18,3	86,3	95,9	27,2	79,2	80,1	20,6	79,1	49,4	79,6	43,9
cpoint_segm	84	39	88	62	49	61	59	60	75	21	71	96	26	77	49	17	72	39	74	32

Average accuracy Code2

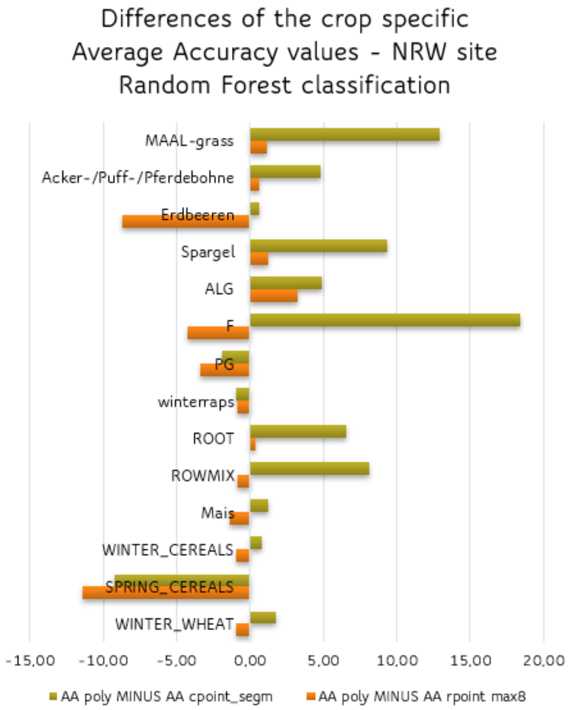


	WINTER_WHEAT	SPRING_CEREALS	WINTER_CEREALS	Mais	ROWMIX	ROOT	winterraps	PG	F	ALG	Spargel	Erdbeeren	Acker-/Puff-/Pferdebohne	MAAL-grass
poly	91,5	60,2	93,3	98,3	77,5	97,9	96,1	87,4	91,7	23,2	82,9	53,2	86,3	43,4
rpoint_max8	90,9	59,7	93,4	98,6	74,8	97,2	97,4	91,1	50,6	20,2	81,1	62,3	82,4	39,1
cpoint_segm	90	51,2	92,7	97,2	71,5	95,8	96,4	87,1	80,3	23,4	66,2	37,1	82,3	34,3



	WINTER_WHEAT	SPRING_CEREALS	WINTER_CEREALS	Mais	ROWMIX	ROOT	winterraps	PG	F	ALG	Spargel	Erdbeeren	Acker-/Puff-/Pferdebohne	MAAL-grass
poly	87,1	53,7	91,0	95,0	78,3	89,9	96,5	76,6	81,8	27,1	86,0	45,5	83,3	52,2
rpoint_max8	88,05	65,10	91,98	96,45	79,08	89,49	97,45	80,06	86,05	23,90	84,81	54,21	82,64	51,07
cpoint_segm	85,4	63,0	90,2	93,8	70,1	83,3	97,5	78,5	63,4	22,2	76,7	44,9	78,5	39,3

Image 30. Crop specific summarized accuracy measures



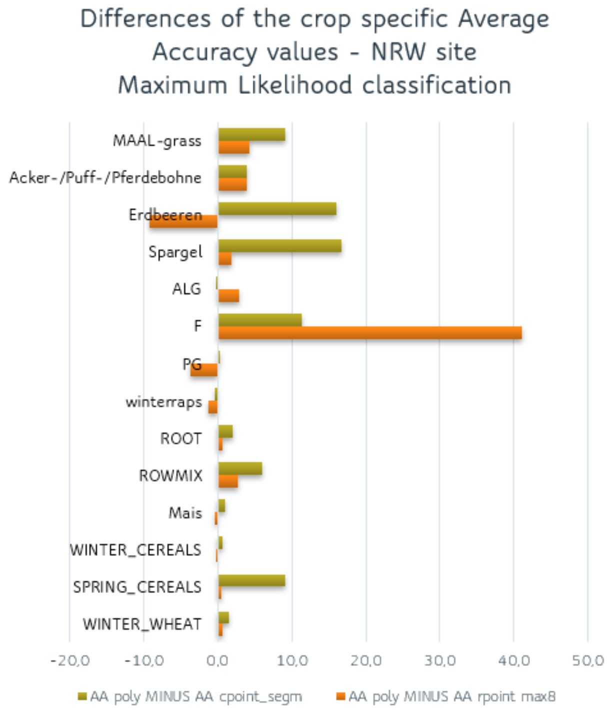


Image 31. Differences of the class specific average accuracy values

"Average accuracy" means the simple mathematical average of the following crop separate accuracy measures in %: user accuracy, producer accuracy, Short (i) Hellden(i).

4.3 An object based random forest classification implemented on the Austrian pilot site

In this study the performance of the training data was tested by an object Based Random Forest classification of Sentinel-2 image time series.

Order	Method	Random Forest based solution in Austrian pilot
1	Selecting input satellite images	Visually selected 11 time series of S2.
2	Rasterizing the declaration of the GSAA	Rasterization of the pixels which are overlapping the random/center points/buffered polygons.
3	Calculating NDVI	Calculating NDVI for all pixels intersecting the polygons/random points / internal centroids
4	Averaging bands/NDVI	S2 bands / NDVI are averaged for each parcel for the polygon/random point method.
5	Scaling	Scaling the features to values between 0 and 1 using a standard scaler.
6	Selecting the most relevant features	Selecting the most relevant features using am principal component analysis.
7	oversampling	Oversampling of underrepresented classes using a synthetic minority oversampling technique (SMOTE).
8	Supervised classification	Training of an object based (parcel based) Random Forest classifier.
9	Accuracy assessment	50% of the reference data population is used as test data, what was processed exactly on the same way as the 50% training. For accuracy assessment, the classified crop types are grouped into crop groups. Calculation of accuracy metrics.

Table 17. Steps of processing the Austrian pilot site

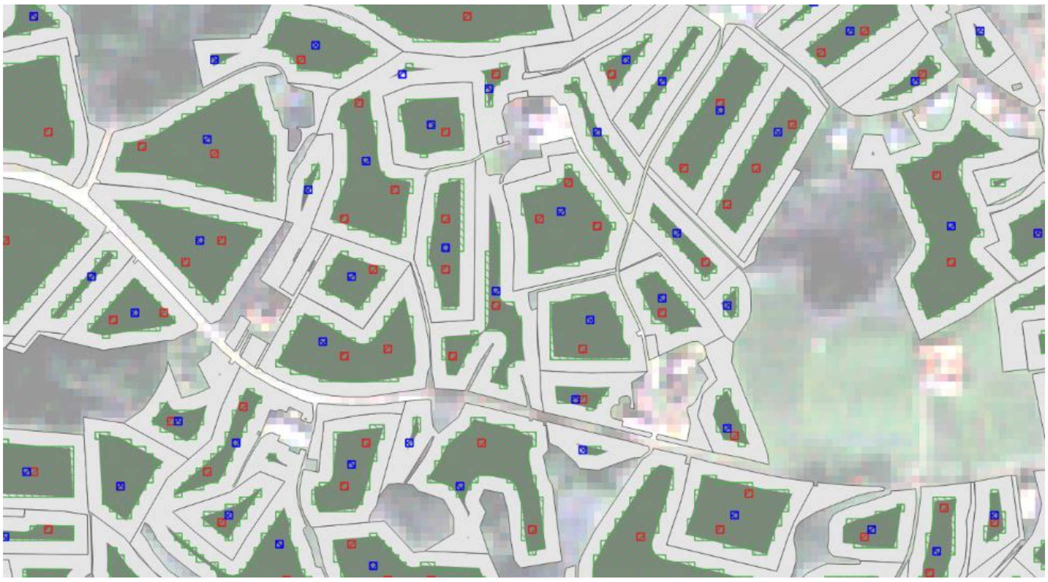


Image 31: 3 types of input data: Green: polygons; Red: random points; blue: centerpoints

For the GSAA polygon input data, pixel and NDVI values were averaged for all pixels from rasterizing the polygons after applying a negative 20m buffer. Accordingly, for the random point method, all pixels overlapping such a point were averaged for each individual parcel. For the center point method, the pixel overlapping the centerpoint was considered to be the feature for classification.

Only crops with at least 100 training parcels in the test area were considered for classification. As a result, 14 crops were considered which were grouped into 11 groups after classification:

The training input features, for each individual crop, were upsampled using a synthetic minority oversampling technique (SMOTE) to reduce the imbalance within the training data.

The random forest classifier was trained using 500 trees and a maximum tree depth of 9. Increasing the depth of the classifier resulted in overfitting of the model.

4.4 Comparing the results of the classifications using the misclassification matrix on the Austrian site

Using the test data, different metrics were used to determine the model performance.

Overall accuracy: percentage of parcels in the test set that were classified in the same crop group as the declaration.

Recall or producer accuracy: Percentage of correctly classified parcels for all parcels with a declaration in a particular crop group. `True Positive / (True Positive + False Negative)`

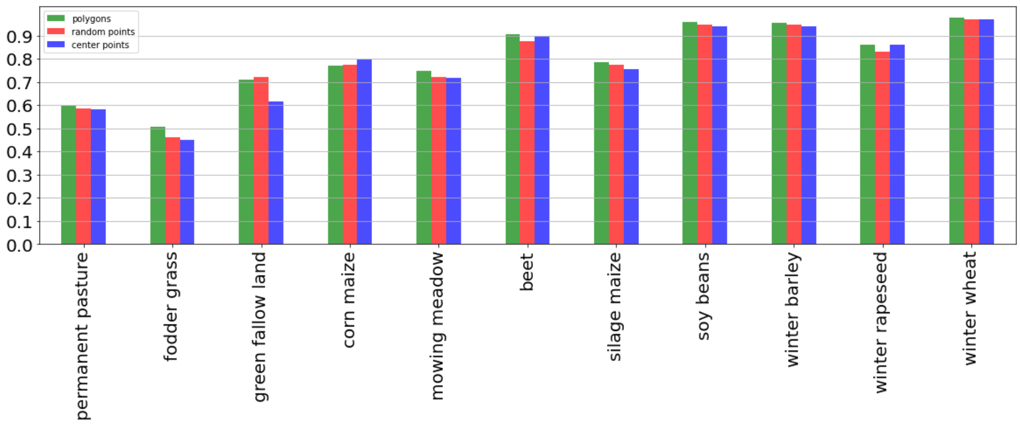
Precision or user accuracy: Percentage of parcels classified as a member of a particular crop group which are also declared to be in that crop group. `True Positive / (True Positive + False Positive)`

F1-score: Combined metric integrating precision and recall giving more weight to the lower number[v].

	Polygons	max.8 random points	centerpoints
Overall Accuracy	78.91 %	77.22 %	77.08 %
F1 - Score	80.09 %	78.59 %	78.53 %

4.4.1

4.4.2 Recall for all crop groups



4.4.3 Precision for all crop groups

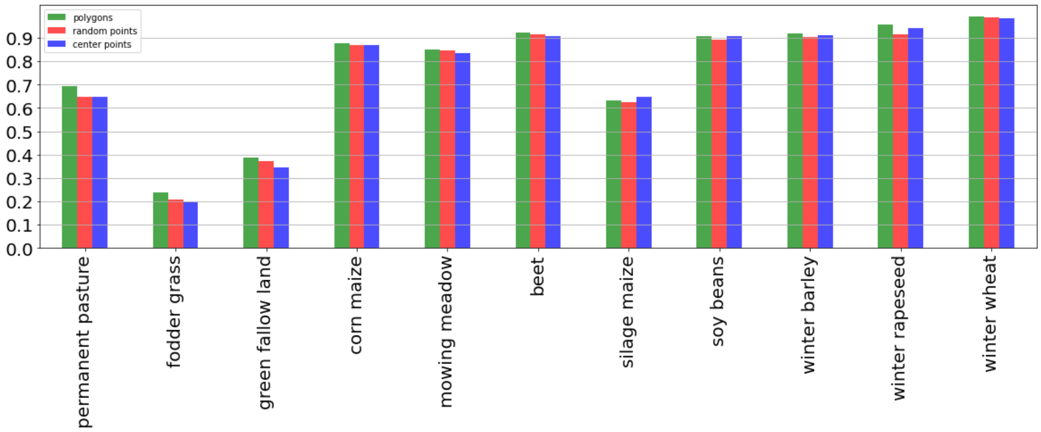
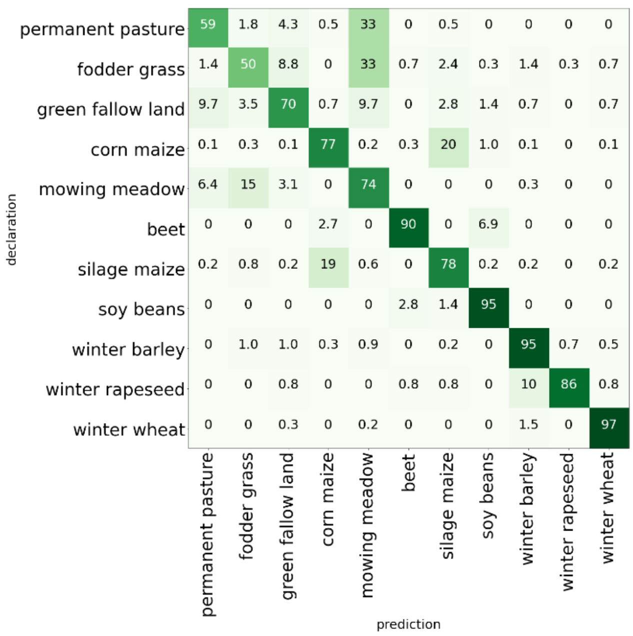
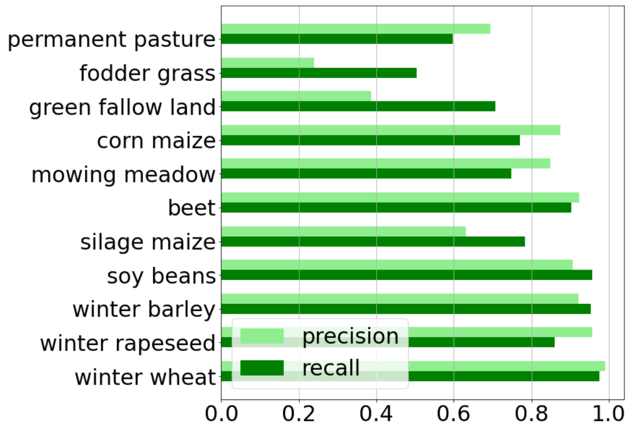


Image 32. Accuracy measures regarding the Austrian Pilot site

4.4.4 Metrics on parcel level - object based classification,

Method where GSAA Polygons were used as training and test data:





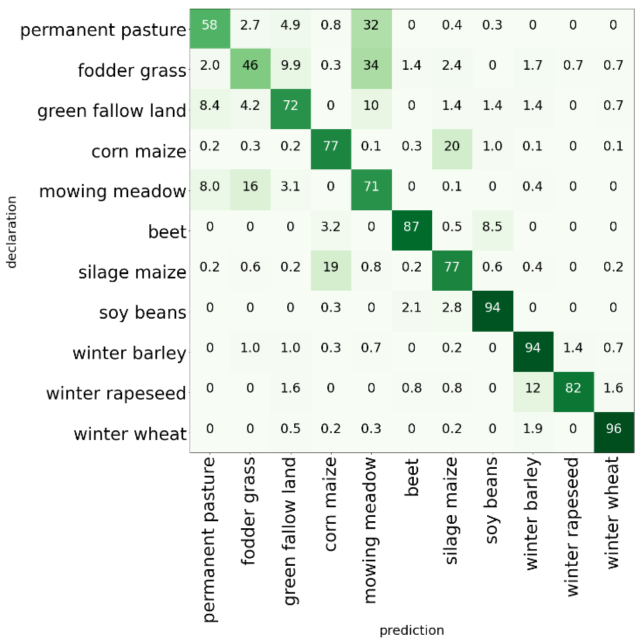
Relative accuracies for all crop groups (GSAA polygons)	precision/recall for individual crop groups

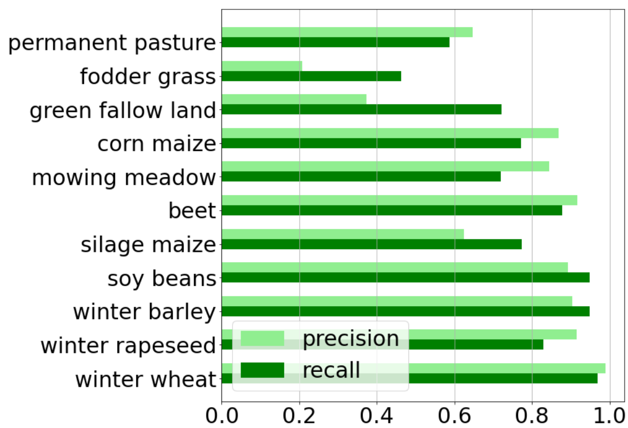
Image 32. Accuracy measures on crop parcel level regarding the Austrian Pilot site, where the entire GSAA polygon is used as training data

		prediction										
		permanent pasture	fodder grass	green fallow land	corn maize	mowing meadow	beet	silage maize	soy beans	winter barley	winter barley	winter wheat
declaration	permanent pasture	474	14	34	4	262	0	4	0	0	0	0
	fodder grass	4	149	26	0	99	2	7	1	4	1	2
	green fallow land	14	5	102	1	14	0	4	2	1	0	1
	corn maize	1	3	1	768	2	3	206	10	1	0	1
	mowing meadow	188	442	89	0	2184	0	1	0	10	0	1
	beet	0	0	0	5	0	170	0	13	0	0	0
	silage maize	1	4	1	96	3	0	391	1	1	0	1
	soy beans	0	0	0	0	0	8	4	275	0	0	0
	winter barley	0	6	6	2	5	0	1	0	561	4	3
	winter rapeseed	0	0	1	0	0	1	1	0	14	111	1
	winter wheat	1	1	3	1	2	0	0	1	17	0	1079

Table 18. Misclassification matrix containing absolute number of parcels (GSAA polygons)

Method where random points (max 8) were used as training/test data:





Relative accuracies for all crop groups (random points)	precision/recall for individual crop groups

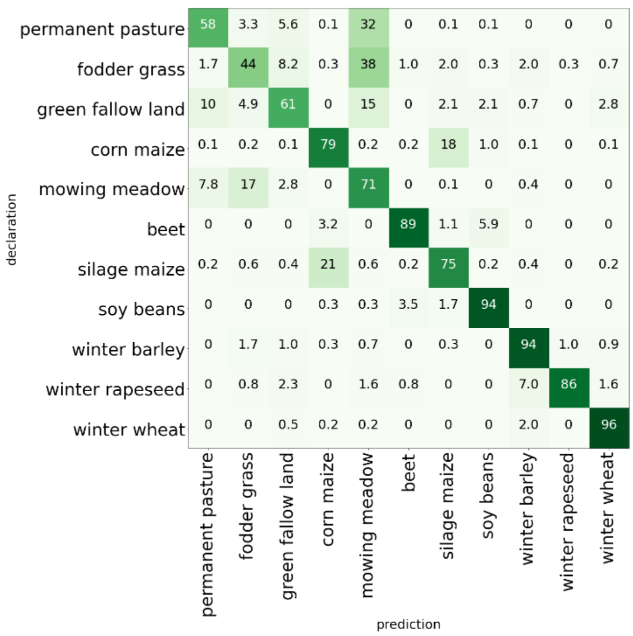
Image 33. Accuracy measures on crop parcel level regarding the Austrian Pilot site, where 8 random points are used as training data

Misclassification matrix containing absolute number of parcels (random points)

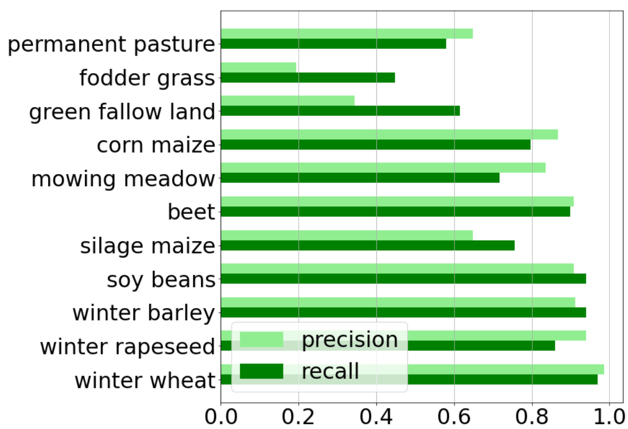
		prediction										
		permanent pasture	fodder grass	green fallow land	corn maize	mowing meadow	beet	silage maize	soy beans	winter barley	winter barley	winter wheat
declaration	permanent pasture	465	21	39	6	256	0	3	2	0	0	0
	fodder grass	6	136	29	1	102	4	7	0	5	2	2
	green fallow land	12	6	103	0	15	0	2	2	2	0	1
	corn maize	2	3	2	769	1	3	204	10	1	0	1
	mowing meadow	233	478	89	1	2097	0	4	0	12	0	1
	beet	0	0	0	6	0	165	1	16	0	0	0
	silage maize	1	3	1	97	4	1	386	3	2	0	1
	soy beans	0	0	0	1	0	6	8	272	0	0	0
	winter barley	0	6	6	2	4	0	1	0	557	8	4
	winter rapeseed	0	0	2	0	0	1	1	0	16	107	2
	winter wheat	0	1	5	2	3	0	2	0	21	0	1071

Table 19. Misclassification matrix containing absolute number of parcels (random points)

Method where center points were used as training/test data:



Relative accuracies for all crop groups (center points)



precision/recall for individual crop groups

Image 34. Accuracy measures on crop parcel level regarding the Austrian Pilot site, where inner center points are used as training data

		prediction										
		permanent pasture	fodder grass	green fallow land	corn maize	mowing meadow	beet	silage maize	soy beans	winter barley	winter barley	winter wheat
declaration	permanent pasture	460	26	44	1	259	0	1	1	0	0	0
	fodder grass	5	132	24	1	113	3	6	1	6	1	2
	green fallow land	15	7	88	0	22	0	3	3	1	0	4
	corn maize	1	2	1	794	2	2	182	10	1	0	1
	mowing meadow	228	498	82	0	2092	0	3	0	12	0	0
	beet	0	0	0	6	0	169	2	11	0	0	0
	silage maize	1	3	2	108	3	1	377	1	2	0	1
	soy beans	0	0	0	1	1	10	5	270	0	0	0
	winter barley	0	10	6	2	4	0	2	0	553	6	5
	winter rapeseed	0	1	3	0	2	1	0	0	9	111	2
	winter wheat	0	1	6	2	2	0	1	0	22	0	1071

Table 20. Misclassification matrix containing absolute number of parcels (center points)

When comparing the three methods, the polygon-based classification model reaches the highest overall accuracy. The models using random points or only the centerpoint reach slightly lower accuracies, but **overall, the results do not differ very much from each other**.

Overall, the grassland groups (permanent pasture and mowing meadow) reached the lowest real values, which were expected due to their spectral variability. Notable is also the confusion between winter barley and winter rapeseed, which accounted for 7-12% of all parcels declared as rapeseed. The most notable difference in classification recall occurred in grassland classes, with green fallow land being classified about 10% worse using the centerpoint method.

5 Conclusions and lessons learnt

5.1 The results of the pilot study

The following overall results are nominated:

- Methodology of processing GSAA data to become a reference of image classification is developed further
- Suitability of all 3 geometrical types of training data is proven, but determined by the classification methods

- Crop separate crop classification cluster maps are derived, and directions of further developments are drafted.

The outputs of this pilot can lead to:

- support the decision that in what format MSs should make available their GSAA data,
- defining standard rules and data structure of GSAA data,
- advising MSs how to define the thematic of GSAA data for best supporting satellite image analysis – useful for CbM as well,
- further testing of image classification solutions: integrating S1 (radar), testing other machine learning algorithms, automatizing how regional aspects can be taken into account
- demonstrate how a more detailed classification of habitats and land uses could serves the AMS, to monitor changes, or to delineate the target area of certain interventions;
- stepping further towards interoperability.

5.2 Proposing recommendations on the acceptable degree of geometrical generalization considering privacy and economic interests of the farmers and the MS

The pilot study has proven the successful use of geometric generalization of GSAA data for training models of crop type classification. In this pilot the performance of 3 training data was tested with 2 types of popular classification methods. The following 3 training data processing was defined during the pilot study:

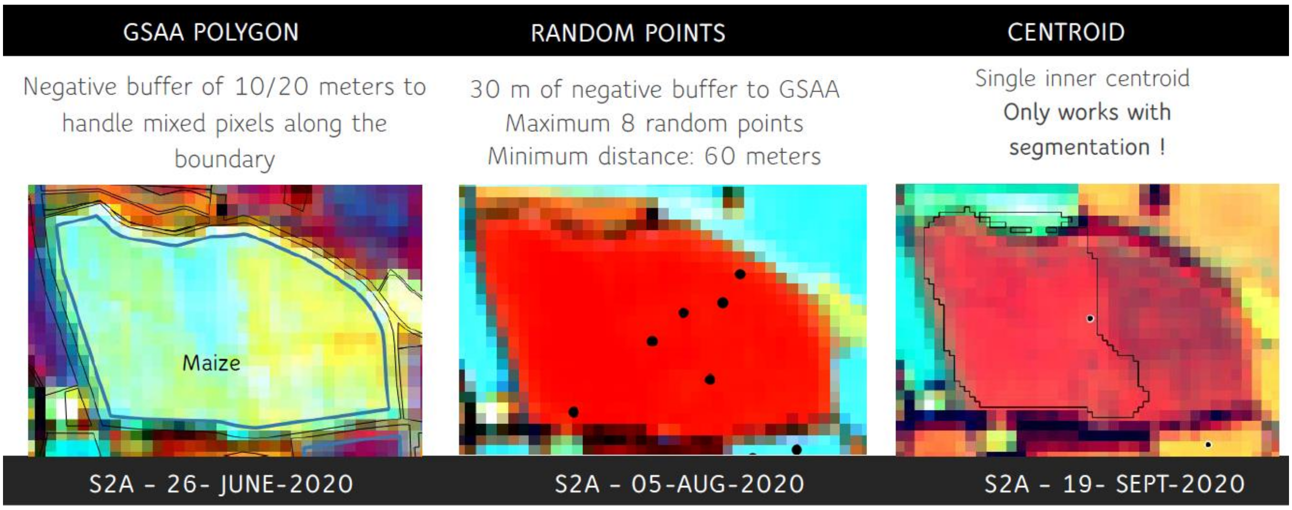


Image 35. Geometrical Generalization of GSAA data

The "maximum 8 point" training sample performed generally slightly higher misclassification rates compared to using the entire GSAA polygon when maximum likelihood classification was used. When random forest is used, the training representation with 8 points is the less suitable, this presented the highest difference in accuracy measures of recall and precision. The reason is, that the random forest algorithm includes the generation of training points, and it is designed for representing polygons, and not for further selecting points among a given set of points. Other machine learning algorithms could also be tested.

The single centroid point representation of GSAA data alone didn't produce accurate results. The use of inner centroid of a crop parcel only fits for the purpose when a segmentation algorithm is used to extend the sampling environment. Segmentation balances the loss of information as compared to the use of the entire GSAA polygon.

As comparing the different training input data, the difference in classification accuracy is under 2%. This interval of accuracy difference could also be compensated by the parametrization of the models or by the grouping of training data. As a general conclusion the use a single inner centroid of IACS crop parcels for crop classification is proven. Seamless crop map with similar accuracy values can be derived as for areas where the entire are of the declared parcels are available.

The GSAA data of 2 member state was used, but the data of other 3 member states (CZ, HU, RO) was also analyzed from classification perspective. The declarations are in all cases are on individual crop level, and thus the same similarity and complexity rules can be applied to define the groups suitable for classification. Although we know that there are strong regional differences regarding crop types and crop phenology behavior, a unified system of selecting training data among the GSAA parcels and to group the parcels for classification can be defined.

As the breakdown of classes for the crop map was defined to experiment the capability of crop distinction at the level of detailed crops (CODE1 in NRW site). Generalization and grouping of crops into wider categories would lead to better classification accuracy values, and will be suitable for AMS, for calculating performance indicators, and for strategic planning and evaluation of the CAP.

5.3 Use of GSAA and LPIS data for training and validation of supervised crop classification models

5.2.1. Generating training data for crop classification from a GSAA dataset

The breakdown of crop type to individual crop level would be fundamental to derive the same groups from different GSAA sources. This is one of the key point to standardize the process for European level. Crop types are however, not available in each context as the data is shared under INSPIRE, but always do exists in IACS managed by the Paying Agencies.

To avoid misleading models, aggregated data types such as "winter coverage" or "forage for wildlife" must be excluded from training. This ensures that these parcels will be classified according to the type of crop really present on them. The same effect is generated if the crop cultivated on a set aside is not available. Distinction between AL-set aside and AL-fallow land is necessary.

GSAA should contain the delimitation and the timing of catch crops and of winter cover crops, cultivated after or before the main crop. This has an impact on the categorization of training data and the selection of time series. Only those declared parcels can be used, which delineate single units of land management, i.e. which represent a continuous FOI.

Depending on its reference parcel type, LPIS data are fundamentally important. They allow masking out the natural vegetation, e.g. forest and wetland areas, to ensure that the classification runs on the eligible area only. Categorization of agricultural area in AL/PG/PC/NAEA give further input for classification. The geometric and semantic consistency of the GSAA and the LPIS data must be checked.

5.2.2. A recommended approach to derive training and test data from polygon representation of GSAA datasets:

1. Analyzing the crop and crop group types to filter out the heterogenous groups. Usually, it is needed when categories, which contain a summary of different crops, are aggregated in the GSAA based on other driving factors than the crop type.
2. Implement a negative buffer along the boundaries of the GSAA polygons to exclude the mixed pixels from the training data of the classification. The **10 and 20 m of negative buffers were** selected in function of the GSD of the input stack image. Setting up a minimum size of parcel corresponding to the GSD of the input image stack is also needed – minimum 6 pixels, leads to 2000 m2 if bands with GSD= 20 are used.
3. Define groups of crops through semantic grouping, which is determined by:
    - the target of the classification – the required level of detail of output classes;
    - the crop codes in the classification nomenclature. However, to keep separate the same crop with the same management practice in different classes (e.g. maize for corn and maize for

seed) will never form an independent class during the classification, even when separate training data is formed for them;
    - merging of at training input level with similar development (spectral signature) on the given set of input satellite images, is essential to get a usable result of cluster map, based on the prior knowledge, that these crops cannot be separated by the given classification algorithm. This can only be based on knowledge about the local crops (not always available), otherwise this step is difficult to automatize. It can be supported by additional examination of the vegetation curves using regression models, or by unsupervised classification.
4. Separate training and test samples – preferably with stratified random selection (proportional to the area or to the number of training parcels).
5. Delete those groups which do not reach the minimum amount of valid reference data. Crop types represented under the minimum amount of training information will also depend on the chosen classification algorithm.

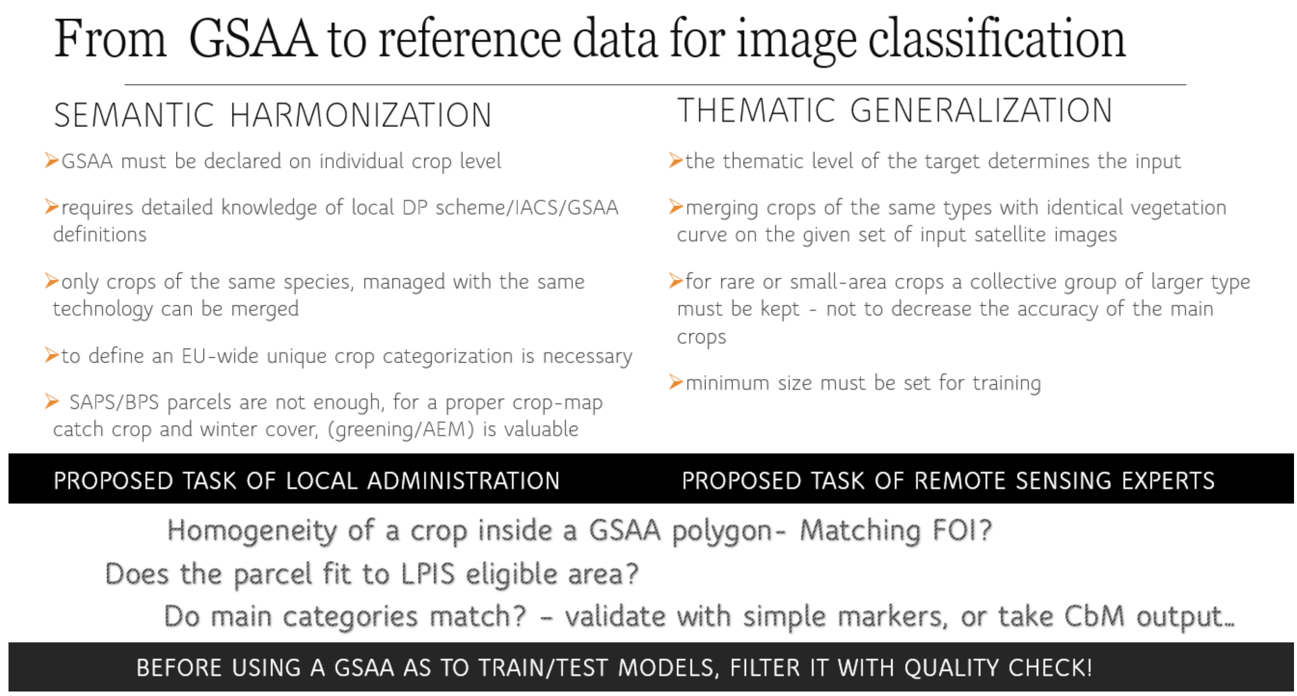


Image 36. Distribution of tasks of deriving reference data for image classification from a GSAA database

5.2.3. Deriving training and test data from point representation of GSAA datasets:

- Point 1, and 3 are valid also for point representation.
- Point 2 and 5 do not apply to point representation
- In case of point data detect the similar set of pixels around the point with a segmentation algorithm worked to be the most effective solution.
- Separate training and test samples (Point 4) can be parametrized proportionally to the number of points in the group, or to the area in case of option (2).

6 Further possibilities, next steps of the study

Based on the current activity, several directions for improving the data processing and classification methods are foreseen, which could eventually lead to the its standardization :

1. To extend crop maps generation to other countries the largest constraint is **the quality of GSAA** data. It has to be tested if it is suitable for training supervised classification models. Two main directions are planned: (1) to evaluate the validation level and quality of the parcels flagged as green in the CbM workflow. (2) filtering out of the training set the not trustable parcels by validating if a declared parcel correspond to a single feature of interest (FOI). This latter could be based on multi-annual S2 image time series and analyzing the vegetation curves of the declared crop in a region to determine local characteristics and variations.
2. This pilot was implemented on small test sites that can be considered as homogenous agricultural regions. To achieve similar accuracy of detailed crop-maps, a semi-automatic **methodology to delineate homogenous regions** could be developed. Another direction is to investigate methods that could handle the temporal shifts of similar vegetation curve of the same crop (an example is the dynamic time warping). The effect of the size and diversity of categorized area on the accuracy could be studied in depth.
3. The random forest algorithm resulted in worse classification accuracy than the maximum likelihood. That is why further testing on parametrizing the RF is needed. The next logical step would be to run it directly using the summarized crop groups (CODE-2) as training data, instead of merging the classes after the crop level classification. This would lead to understand how the training input determines the distribution of trees. Also, **the performance of other machine learning algorithms could be compared to the current results**.
4. We know from previous experiences (cit.), that the integrated use of radar (S1) and optical (S2) data leads to more accurate results for some crops. The main AL-crops had been surprisingly well separated with the S2 based processes, but for distinguishing multi annual graminid-dominant energetic crops (like Acre-grass) from the permanent grassland the data content of S1 time series seems to be necessary. A proposed test case could be to define methods on **how to integrate S1 decomposites into the classification input**.
5. Stepping further towards interoperability, pilot cases on **how a more detailed classification of dedicated habitats and land uses serves the AMS** could be implemented. For example, multi-annual classification of sensitive permanent grasslands, or PG-ELP categories could lead to the monitoring of changes, or to delineation of certain interventions on specific areas.

7 References and acknowledgements

7.1 Abbreviations

AEM – Agri-environment measures

AL/PG/PC/NAEA – arable land/permanent grassland/permanent crop/non-agriculture eligible area

AMA - Agrarmarkt Austria

AMS – Area Monitoring System

BPS -Basic Payment Scheme

BOA – Bottom of atmosphere

CAP – Common Agricultural Policy

CbM -Check by Monitoring

DTM – Digital Terrain Model

EU – European Union

FOI – feature of interest

FL – fallow land

FN – false negative

FP – false positive

GSAA – Geospatial Aid Application

GSD -ground sampling distance

IACS -Integrated Administration and Control System

LCC - Land Cover Classification

LPIS - Land Parcel Identification System

L2A – Level 2A

MS – member states

ML – machine learning

MXL - maximum likelihood classification

MS – member states of the EU

NDVI - Normalized Difference Vegetation Index

NRW – North-Rhine Westphalia

OTSC – on-the-spot control

PA – Paying Agency

PB – Physical block of LPIS

PCA – principal component analysis

PG-ELP – permanent grassland – established local practice

PMEF – Performance-based Monitoring and Evaluation Framework in agri-environment-climate policy

RF- random forest classification

RGB – red/green/blue of image color composite

SAPS – Single Area Payment Scheme

S1 – Sentinel-1 radar image

S2 - Sentinel-2 optical image

TG -temporary grassland

TN – true negative

TP – true positive

7.2 Technical resources used

Different combinations of open source SW shad ben used to complete the pilot study:

- QGIS version 3.16.1-Hannover for general GIS processes and for visualization of the results,
- SNAP - ESA Sentinel Application Platform 8.0.3 - to run classifications
- Semi-Automatic Classification Plugin (SCP) is a free open source plugin for QGIS that allows for the semi-automatic classification (also supervised classification) of remote sensing images, Congedo Luca (2020). Semi-Automatic Classification Plugin Documentation.
  DOI: http://dx.doi.org/10.13140/RG.2.2.25480.65286/1
- GDAL/OGR contributors (2021). GDAL/OGR Geospatial Data Abstraction software Library. Open Source Geospatial Foundation. URL https://gdal.org
- OGR, Numpy, SciPy, and Matplotlib under Jupyter Notebooks – a publishing format for reproducible computational workflows Authors: Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, Carol Willing, Jupyter Development Team
- Temporal/Spectral Profile Tool Plugin – License GNU GPL 3 © 2020 DHI-GRAS A/S to plot raster bands from the selected raster layers,
- Ms OfficeXcell to create matrixes and charts.
- Scikit-learn to train/test the classifier in the Austrian testsite and to calculate accuracy metrics. URL sklearn.org
- imblearn to apply oversampling. URL imbalanced-learn.org

1.1 Acknowledgements

I would like to thank to the national administrations for their willingness of participation and for supporting the availability and understanding of the data, namely Burkhard Ulonska and Berker Timo – from the side of Germany, the North-Rhine-Westphalian Paying Agency (Landwirtschaftskammer Nordrhein-Westfalen), - and Bernhard Eder – from Agrarmarkt Austria - AMA the Austrian Paying Agency. The co-operation of Gerhard Triebnig and Elias Wanko from EOX IT Services GmbH regarding the Austrian pilot was greatly appreciated. I would also like to express my special thanks of gratitude to Gizella Nádor for her support.

7.3 References:

[i] [Steven M. Holland, Univ. of Georgia]: Principal Components Analysis
[skymind.ai]: Eigenvectors, Eigenvalues, PCA, Covariance and Entropy
[Lindsay I. Smith] : A tutorial on Principal Component Analysis

[ii] Scofield, Graziela & Pantaleão, Eliana & Negri, Rogério. (2015). A Comparison of Accuracy Measures for Remote Sensing Image Classification: Case Study In An Amazonian Region Using Support Vector Machine. International Journal of Image Processing. 9. 2015.

[iii] Robert Gilmore Pontius Jr & Marco Millones (2011) Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment, International Journal of Remote Sensing, 32:15, 4407-4429, DOI: 10.1080/01431161.2011.552923 To link to this article: https://doi.org/10.1080/01431161.2011.552923

[iv] Congalton R. G., Oderwald R. G.: Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques, Photogrammetric Engineering and Remote Sensing, Vol. 49, No. 12, December 1983, pp. 1671-1678.
Maselli F., Conese C., Zipoli G., Pittau M.A.: Use of error probabilities to improve area estimates based on maximum likelihood classification Remote Sens. Environ., Vol. 31, pp. 155-160 (1990)
Rosenfield G. H., Fitzpatrick-Lins K.: A coefficient of agreement as a measure of thematic classification accuracy, Photogrammetric Engineering and Remote Sensing, Vol. 52, No. 2, February 1986, pp. 223-227.

[v] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html