Integration of remote sensing data for control in the system of direct agricultural subsidies (IACS)

Project duration: 2.11.2020–2.11.2022
Project manager: Prof. Dr hab. Eng. Beata Hejmanowska
Faculty of Geo-Data Science, Geodesy, and Environmental Engineering

AGH University of Krakow | RESEARCH UNIVERSITY EXCELLENCE INITIATIVE

RESEARCH UNIVERSITY
EXCELLENCE INITIATIVE

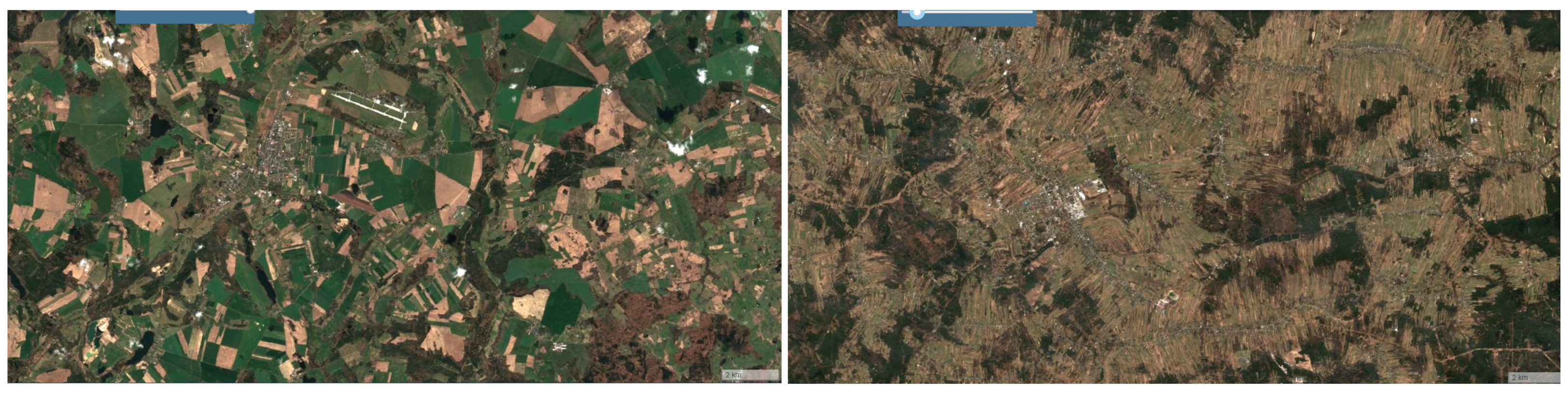
INTRODUCTION:

The aim of the project is to examine of the remote sensing data integration for control procedure in the system of direct payments in agriculture (IACS – Integrated Administration Control System).
In order to obtain subsidies, the farmer declares the **type of crop and its area**. The accuracy of this information is controlled in the IACS system. Until now, the inspection was carried out during a field visit or partly remotely, among others with the use of aerial photos processed into the so-called orthophotomaps, such as Google Maps.
Recently, the **EU replaceed the so-called on-site inspection** with remote sensing using the new products of the European Space Agency (ESA) of the Copernicus program, i.e. the **Sentinel-1 and Sentinel-2 images**. The method proposed by the JRC EU is based on analyzing dense time series of these images. In Northern and Central Europe, acquiring multiple images for the entire country is challenging due to a limited number of "flyable days" caused by frequent cloud cover throughout the year. Additionally, the pixel size of 10 meters poses a problem for areas with fragmented agricultural structures. Consequently, we focused on research towards maximizing the simplification of the control methodology. We were inspired by the publication (Maponya et al. 2020) and decided to test a single registration of Sentinel-2 imagery taken four weeks before harvest.

METHODS:

During the research work, a **two-stage approach** (from general to specific) was adopted. The first stage involved examining the possibility of training a **neural network for the classification of selected land cover types in agricultural areas**. We hypothesized that such a selection of classes (particularly the defined crop classes: 4 and 9) would allow training the network in a way that enables its use for classifying any Sentinel-2 image, regardless of the registration date and location in the country, without the need for additional training fields. The chosen model for the network was Unet, which was implemented using the PyTorch framework and the Python programming language.
In the second stage, we investigated the possibility of recognizing crops at the plant level. We selected two test areas that are representative of two different agricultural structures in Poland:

- Świdwin in northwestern Poland, in the West Pomeranian Voivodeship – characterized by large, regularly shaped agricultural fields.
- Kolbuszowa in southeastern Poland, in the Subcarpathian Voivodeship – characterized by small agricultural fields with complex, often elongated shapes.

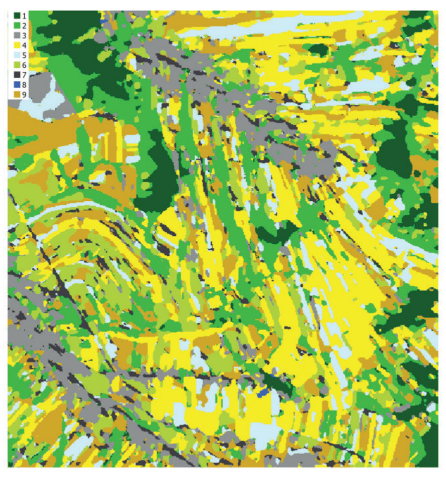


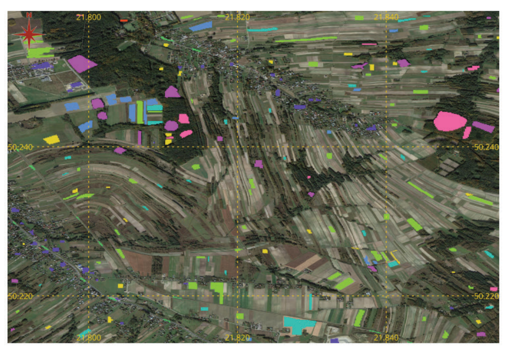
Świdwin | Kolbuszowa

In the Świdwin area, we utilized farmers' declarationsas well as Sentinel-2 images obtained from the ESA Hub and performed classification using the Random Forest method with custom Python scripts.
In the Kolbuszowa area, we performed classifications using the CART, Random Forest, and SVM methods with training fields acquired during on-site visits using Sentinel-2 images and hyperspectral aerial images from the HySpex camera. All calculations were conducted in the Google Earth Engine (GEE) cloud-based solution.
For both test areas, we analyzed the classification results of multiple time-series images and compared them to the results of a single registration classification. The accuracy metric used was overall accuracy, which is the number of correctly classified pixels divided by the total number of examined pixels.

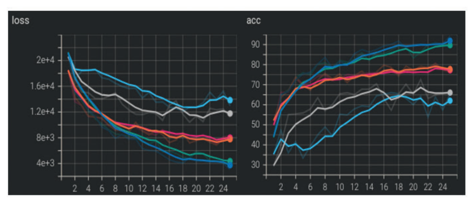
RESULTS:

The result of the first stage is a preliminary classification of agricultural areas which allows for the delineation of "masks" of non-agricultural areas and areas covered by crops divided into four groups: permanent green cover, crops in the intensive vegetation phase, agricultural fields before harvest or in the spring before agricultural operations, and bare soil. The accuracy of the Unet model's classification in the Kolbuszowa test area was 91%, while the RF (Random Forest) method in OpenCV achieved 78%, and ANN (Artificial Neural Network) achieved 80%.
In the second stage, in the Świdwin area, the classification result for three multi-temporal images was 81%, while for one pre-harvest image, it was 79%. In the Kolbuszowa area, the highest accuracy was obtained for multi-temporal images using the RF method, with a result of 82%. However, for single aerial registration using the CART method, the accuracy was 77%. Single registration of Sentinel-2 images did not yield results above 50% in any case, with the highest value achieved just before harvest being 39%.





OpenCV RF	0.78
OpenCV ANN	0.89
UNET model	0.91



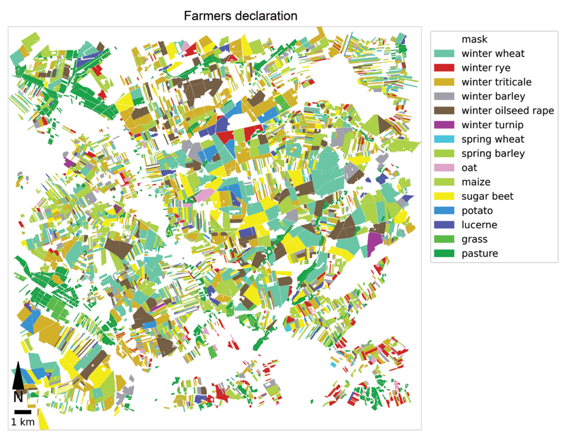
1 coniferous forest
2 mixed forest
3 buildings, industrial areas
4 crops: mature cereals (before harvest) or in the spring before performing agro-technical treatments
5 bare soil
6 permanent grassland
7 roads
8 waters
9 crops: in intensive vegetation phase

OA = 79% | OA = 81%

single registration 20190526 | multitemporal registration

mask | no_diff | diff

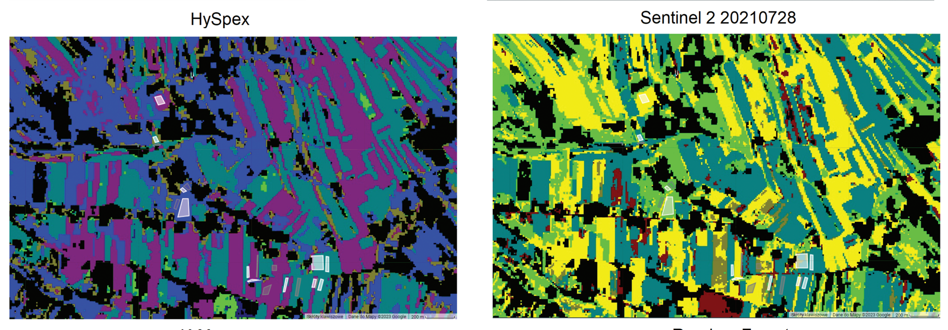
Farmers declaration



Id	Crop	Count_Control	Area_Control	Count_Test	Area_Test
1	winter wheat	288	23,749	279	23,754
2	winter rye	117	4389	137	5617
3	winter triticale	402	22,998	452	22,493
4	winter barley	147	6377	178	8652
5	winter oilseed rape	121	10,564	119	15,126
6	winter turnip	3	251	5	1188
7	spring wheat	8	310	11	466
8	spring barley	141	5423	122	3883
9	oat	32	1100	51	2163
10	maize	402	25,213	447	23,961
11	sugar beet	120	11,289	125	11,194
12	potato	11	1059	22	4128
13	lucerne	53	3206	48	2925
14	grass	54	2969	62	2834
15	pasture	291	12,352	328	12,870
	Total	2190	131,250	2386	141,253

Parameter	Information
Satellite:	Sentinel-2
Level:	2A (after geometric, radiometric and atmospheric correction)
Selected bands:	10 bands: B2, B3, B4, B5, B6, B7, B8, B8a, B11, B12
Number of images:	3
Dates:	28 September 2017, 26 May 2018, 20 July 2018
Images:	S2A_MSIL2A_20170928T100021_N9999_R122_T33UXT_20200104T144027 S2A_MSIL2A_20180526T100031_N0208_R122_T33UXT_20180526T161700 S2B_MSIL2A_20180720T100029_N0208_R122_T33UXT_20180720T142656
Short name:	S2_20170928, S2_20180526, S2_20180720

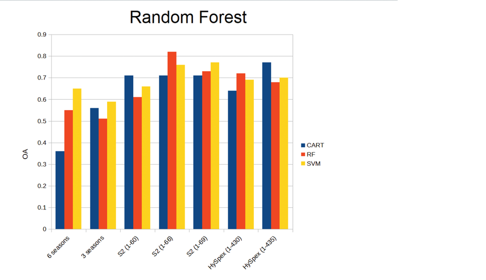
HySpex | Sentinel 2 20210728



K-Means | Random Forest

Data	S2 ID
27.03.2021	S2B_MSIL2A_20210327T093039_N0214_R136_T34UEA_20210327T120034
11.04.2021	S2A_MSIL2A_20210411T093031_N0300_R136_T34UEA_20210411T122810
09.05.2021	S2B_MSIL2A_20210509T094029_N0300_R036_T34UEA_20210509T120133
25.07.2021	S2B_MSIL2A_20210725T093039_N0301_R136_T34UEA_20210725T115620
28.07.2021	S2B_MSIL2A_20210728T094029_N0301_R036_T34UEA_20210728T125908
06.09.2021	S2B_MSIL2A_20210906T094029_N0301_R036_T34UEA_20210906T113414

Id	Crop
2	soil
5	oats
6	wheat rye
7	wheat winter
8	grass



S2 20210728 OA 39%

DISCUSSION:

The discussion covers the topic of accuracy metrics. Currently, due to the use of machine learning methods, there has been some confusion in publications regarding the use of different accuracy metrics, often used interchangeably. Specifically, accuracy and overall accuracy (OA) are frequently reported as the same metric. An example is provided to illustrate the difference. In our case, we could report the highest accuracy for Kolbuszowa using RF as 93% instead of OA as 82%.

TABLE III. ACCURACY METRIC

Name	Formula
Producer accuracy (PA) Sensitivity True positive rate (TPR)	$\frac{TP}{TP+FN}$
Specificity True negative rate (TNR)	$\frac{TN}{TN+FP}$
User accuracy (UA) Precision Positive predictive value (PPV)	$\frac{TP}{TP+FP}$
Accuracy (ACC)	$\frac{TP+TN}{TP+TN+FP+FN}$
F1 score	$\frac{2TP}{2TP+FP+FN}$
Overall accuracy (OA) Percent of correct precision	$\frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n}(TP_i+TN_i+FP_i+FN_i)}$

TABLE IV. ERROR MATRIX (TEST SET)
LEFT – S2 (1-66 BANDS, RF), RIGHT - HYSPEX (1-435 BANDS, CART)

	ground true									
ID	2	5	6	7	8	2	5	6	7	8
2	3	6	24	0	0	239	22	62	0	61
5	0	52	0	0	0	0	707	17	9	12
6	2	6	28	2	0	7	36	76	289	14
7	4	8	0	70	0	73	83	68	764	0
8	0	0	0	0	88	0	31	10	0	927

TABLE V. METRICS S2 (1-66 BANDS, RF), **OA=0.82**

TPR/PA	TRN	PPV/UA	ACC	F1
0.33	0.89	0.09	0.88	0.14
0.72	1.00	1.00	0.93	0.84
0.54	0.96	0.74	0.88	0.62
0.97	0.95	0.85	0.95	0.91
1.00	1.00	1.00	1.00	1.00
0.71	**0.96**	**0.74**	**0.93**	**0.70**

TABLE VI. METRICS HYSPEX (1-435 BANDS, CART), **OA=0.77**

TPR/PA	TRN	PPV/UA	ACC	F1
0.75	0.95	0.62	0.94	0.68
0.80	0.99	0.95	0.94	0.87
0.33	0.89	0.18	0.86	0.23
0.72	0.91	0.77	0.85	0.75
0.91	**0.98**	**0.96**	**0.96**	**0.94**

SUMMARY:

For large agricultural fields, a single registration of Sentinel-2 imagery can be used to achieve crop classification accuracy of around 80%, which is not significantly lower then analyzing multi-temporal images. For small agricultural fields, similar accuracy can be achieved using multi-temporal Sentinel-2 images and single aerial registration. However, single registration of Sentinel-2 images does not allow for obtaining results above 50%. Definitely, in accuracy analysis, one should not use the ACC (accuracy) metric instead of overall accuracy (OA).

Maponya, M.; van Niekerk, A.; Mashimbye, Z. Pre-harvest classification of crop types using a Sentinel-2 time-series and machine learning. Comput. Electron. Agric. 2020, 169.