Sentinel data fusion for official agriculture statistics in Hungary

Daniel Marton KOVACS[1], Istvan Peter KOVACS[1], Luca GATTI[2], Levente RONCZYK[1], Agnes PATAY [3]

[1] University of Pecs, Department of Cartography and Geoinformatics, Hungary (kdanielm@gamma.ttk.pte.hu)

[2] sarmap SA, Via Stazione 52, 6987 Caslano, Switzerland l.gatti@sarmap.ch

[3] Hungarian Central Statistical Office - Agriculture Statistics Section - Agnes.Patay@ksh.hu

Abstract: Field level crop mapping is a rather conventional technology today, utilizing a wide range of satellite imagery fused and validated with ground-level datasets. However, there are certain land-cover categories and crop types that are difficult to distinguish from one another, even with the ever-developing tools of satellite remote sensing. The primary purpose of this study was to demonstrate the capability of fused Sentinel-1 and Sentinel-2 data and explore the possibility of aggregating the results into official statistics. During our project, we intended to detect and quantify grasslands, and two of the most frequently sown winter cereals: wheat, and barley. These are difficult domains of crop mapping, and coincidently, the area where regular and reliable data is rather difficult to acquire for various reasons in Hungary.

1. Introduction

The main objective of our project was to demonstrate the crop mapping capabilities of Sentinel data, and its possible applications for the Agricultural Statistics Section of the Hungarian Central Statistical Office (HCSO) in a joint research effort between the *University of Pecs*, the *HCSO*, and *sarmap SA* from Switzerland. Our aim was to elaborate an adequate solution for grassland detection, and winter crop separation in a selected region of Hungary (HU23 - Southern Transdanubia Region) (Figure 1). HU23 region consists of three subdivisions: Baranya, Somogy and Tolna counties, with a total land area of 14197.86 km², of which roughly 11634.43 km² is under some kind of agricultural use, according to recent data from the HCSO. Surveying grasslands and croplands on large scales are important for many different reasons. Grassland ecosystems covering a vast section of the world's terrestrial area. These territories are prone to severely deteriorate if tampered with, for instance by extensive grazing, invasive species, or changing climate, and Hungarian grasslands are no exceptions. Protecting these areas are at the forefront of nature conservation efforts, as grasslands are representing vital ecosystem functions. Hence, many grassland areas are protected natural assets around Europe as well as globally and monitoring and quantifying these areas is an important topic of land cover change studies [1], [2]. Wheat on the other hand is one of the primary food sources of humanity. During recent years in Hungary, almost one-third of the total arable land was used to grow wheat and barley. Keeping precise records of the lands and monitoring the crop types over time is of paramount importance not only for direct economic reasons or statistical data collection and analysis, but is also fundamental for soil conservation efforts as well.

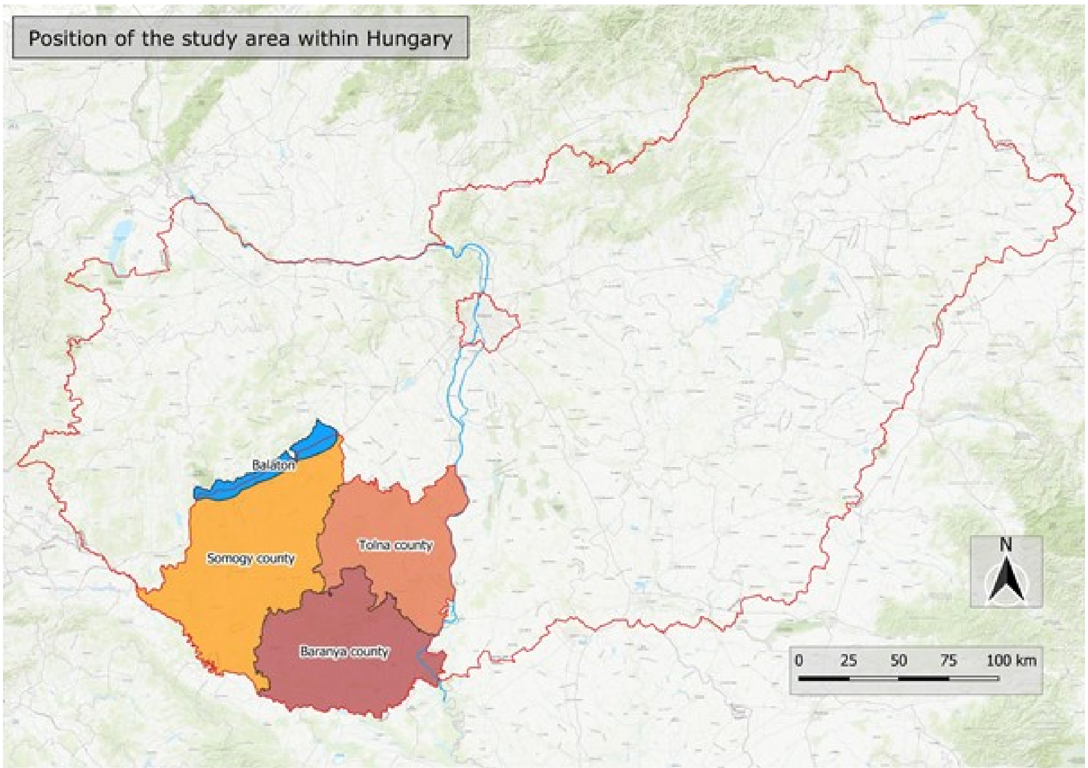

Figure 1 - The study area.

2. Data and Methods

To identify land cover types and cereal fields throughout the study area, we used various Sentinel-1 (S1-SLC and GRD) and Sentinel-2 (1C) imagery, produced between September 2019 and October 2020. The borders of the research area were appointed in conformity with NUTS data, while the water bodies and populated areas were masked out by layers according to the refined version of the CORINE database (also known in Hungary as CLC50). All processing was carried out on the Hungarian HPC system using the MAPscape 5.2 software, which is a well-proven solution for large scale monitoring [3]. We designed a semi-automated workflow that was responsible for downloading, preparing, and processing the images without human intervention, according to pre-set details of time-frames and required parameters.

2.1 Utilized data in detail:

- *Sentinel-1 data:* Sentinel-1A/1B IW SLC and GRD product datasets, acquired from Copernicus Open Access Hub according to the following time-frames:
    1. Intensity (GRD)
        a. 2019/02/09. - 2019/13/12.
        b. 2020/06/01. - 2020/10/07.
    2. Coherence (SLC)
        a. 2019/02/09. - 2019/20/10.
        b. 2020/06/03. - 2020/28/06.
- *Sentinel-2 data:* MSI Level-1C data, acquired between 1st of March 2020 and 31st of August 2020.
- *NUTS region borders:* Providing the official frame for the project, we used the NUTS 2 and NUTS 3 level borders of the Southern Transdanubia region (HU23), as well as individual county extents of Baranya (HU231), Somogy (HU232), and Tolna (HU233) counties in .shp format
- *CLC50 - CORINE Land Cover:* Refined version of the original CLC100 inventory, produced by the Hungarian Institute of Geodesy, Cartography and Remote Sensing for the entire extent of Hungary. The dataset was used to mask out water bodies and populated areas during the early phase of data processing.

During the workflow in MAPscape environment, after the standard preparation of slant range mosaicking, co-registration, speckle filtering, terrain geocoding and calibration, finally anisotropic filtering we produced the Dual Polarization SAR Vegetation Index (DPSVI) for S1. As for the S2 optical imagery, after simple masking and mosaicking, the vegetation index (NDVI) could be calculated. The DPSVI and NDVI components then were fused at a uniform 20-meter resolution and into multi-temporal features that were combined according to Boolean logic and to be implemented in decision trees. At this point, it is essential to have knowledge of the signal interactions of different crop types, to select the most discriminative multi-temporal features for successfully separating grasslands and different crop fields from their surroundings. As these outputs were scattered with small, agriculturally insignificant plots, a final pixel-based filtering led to the generation of the end products. The simplified, general classification scheme is presented in Figure 2 below. Postprocessing and refinements were performed in QGIS, and on mobile GIS applications (QField, Mergin) during the accuracy assessment phase. Due to the time frame of the project, the accuracy assessment of the wheat and barley fields were performed post-harvest, and according to private datasets and maps provided by selected producers across the study area.

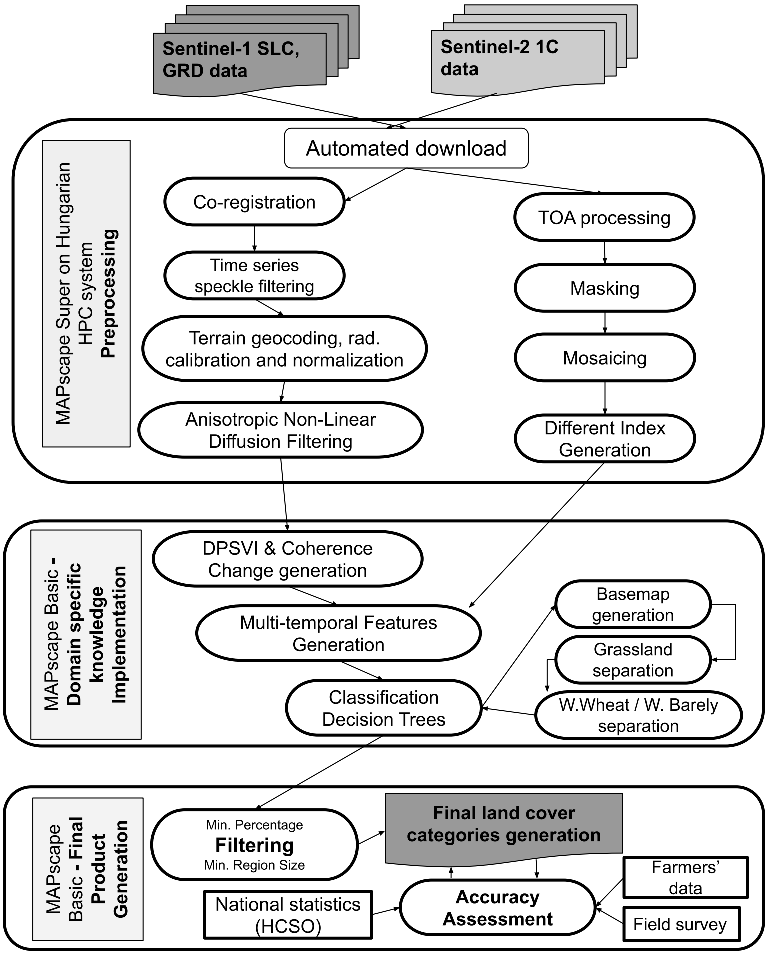

Figure 2. General classification scheme (by L. RONCZYK)

3. Results

To identify our objectives, the first step was to produce a base map that later would be used to mask out the non-essential land cover types. It was calculated from different multi-temporal features such as VH maximum value, coherence pair minimum values have played a significant role during the classification schemes. Important to highlight that the averaged NDVI from the Sentinel-2 images, effectively enabled the extraction of the permanent land cover types (settlements, forests, water bodies) from the croplands and grasslands. Unfortunately, by its static nature, the CORINE dataset tends to be inaccurate in covering the dynamic changes of cropland and grassland areas, therefore we just implemented it to built-up area masking. During the following steps, we separated winter and summer crops, sorting the data according to the spectral differences caused by the different cultivation times of the individual fields, until mostly grasslands remained unmasked.

3.1 Winter crops

The first step of the winter crop identification was to discriminate between the summer and the winter crop fields. It can be performed in different ways. In our case, we used NDVI mean values from April and July as a temporal descriptor. The following task was the main challenge, to split winter barley and winter wheat fields. The key detail to this was derived from a simple piece of information provided by the farm operators about their crop cultivation practices, as their general routine is to sow the barley two weeks earlier than the wheat. This temporal advantage is detectable in the phenology phase of the crops and the management practice on each field as well. Based on this consideration we fused Sentinel-1 and Sentinel-2 images in both periods of germination and ripeness respectively, and by doing so, successfully separated the cereals to the required level. During the process, we separated rapeseed fields as well - which is also a very commonly sown winter crop in the region - but it was only a by-product. Its separation was relatively easy from the cereals due to its different plant structure, which caused much stronger backscatter values. The results are represented in Figure 3, as a winter crop map of the Southern Transdanubia region.

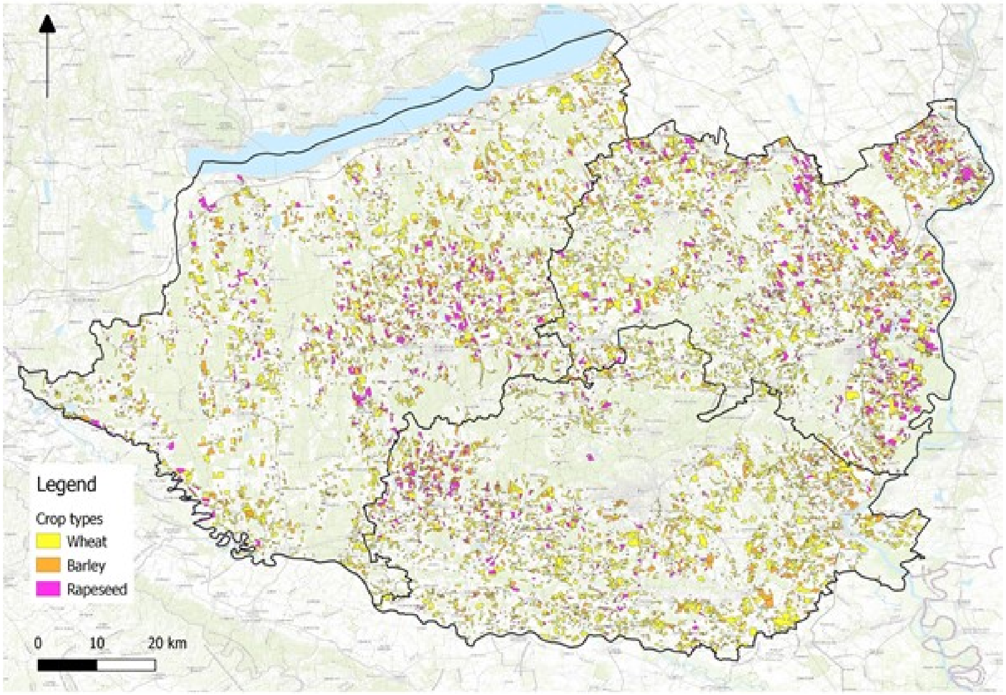

Figure 3: Winter crop map of the Southern Transdanubia region.

Quantifying the outcome of the winter wheat classification, our results yielded only a few per cent difference compared to the fresh HCSO statistics for 2020. Generally, an 8% of extra area was detected in Somogy county, 3.9% in Tolna county while a 5.9% deficit presented itself in the data for Baranya county. These figures are averaging a slight, 1.89% of positive difference in our calculations compared to the data provided by the HCSO for the region.

Concerning the winter barley fields, in our results, 6.4% extra area were detected in Somogy county, 8.9% in Baranya county, while an 18.4% extra area was detected as winter barley fields in Tolna county. These numbers are averaging out to a 10.01% positive difference compared to the same official data source. However, considering the data acquiring methods of the HCSO, and the methodology used in our project, some differences were expected, and accounted for during the work. Both approaches have some weaknesses, and chances for inaccuracy in very different ways. Details of the results in hectares, alongside the official data from the HCSO, are presented in Figures 4 and 5 below.

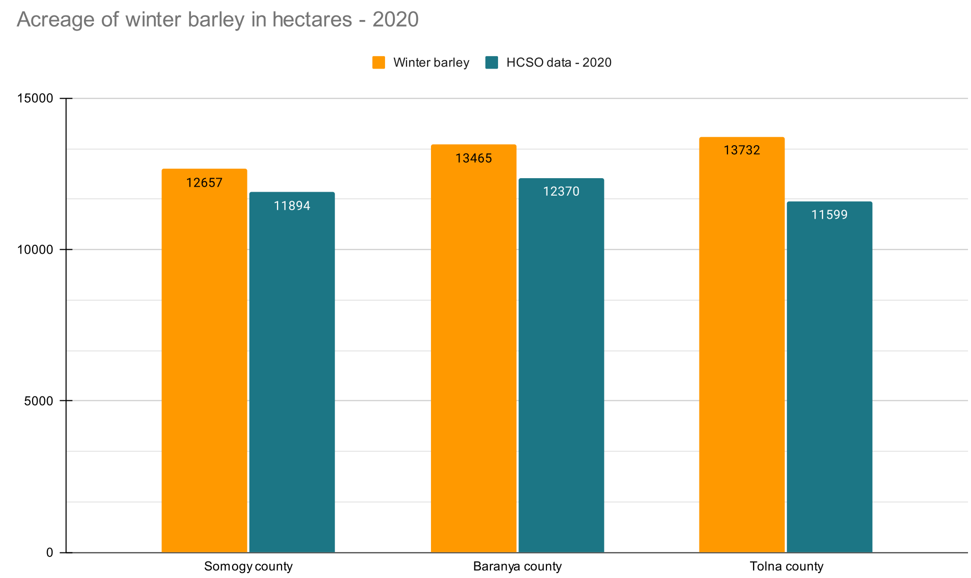

Figure 4. County-level comparison of our calculation of winter barley with the HCSO data (in hectares)

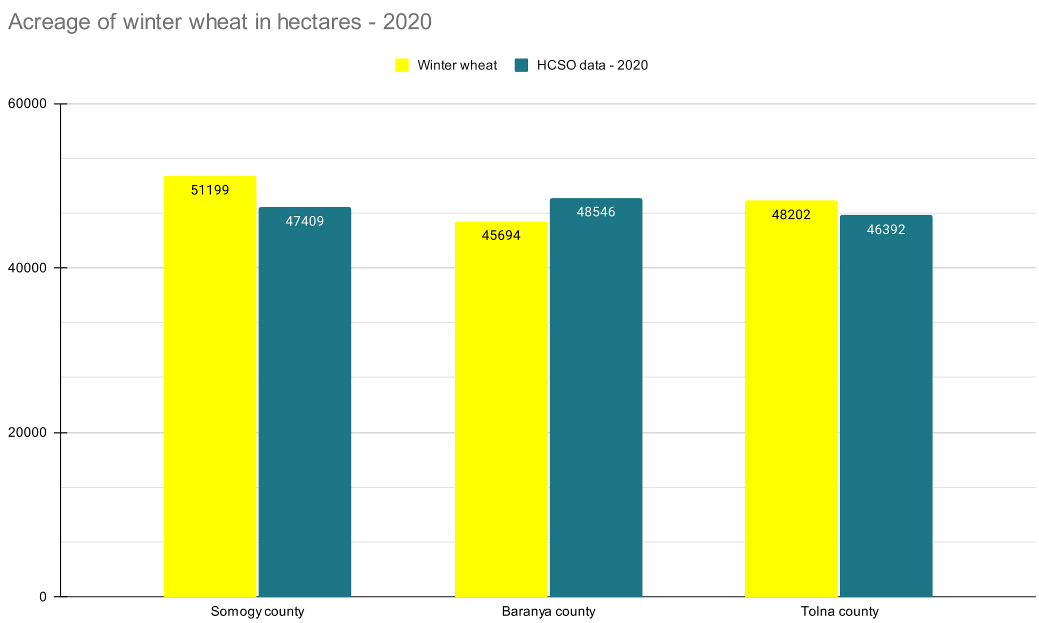

Figure 5. County-level comparison of our calculation of winter wheat with the HCSO data (in hectares)

3.2 Grassland classification

The primary aim of the grassland classification was only to locate and quantify the grasslands across the Southern Transdanubia region. The HCSO has quantified and well-maintained data of the grasslands, but this data lacks spatial attributes, so the whereabouts and distribution of these areas are unknown within individual counties. In Hungary, grasslands are very broadly interpreted, and as such traditionally many different land-use types were generally considered as grassland. Thus, the actual land cover types falling into the official grassland statistical category are very diverse. Consequently, during our project, we also classified meadows, grazing lands, fallow lands, and abandoned fields into this class. The final product covered 64081 hectares and needed heavy filtering (min 50 pix.), because of the elongated shapes of grassland patches alongside the road-, train, and waterways network (Figure 6).

Taking into account the complexity of the category our results yielded fairly successful figures. In Somogy county we detected 7.1% less grassland area, in Baranya county 6.1% less, while interestingly in Tolna county the results were 14.5% more than the official statistics for 2020 provided by the HCSO. These figures averaging to an altogether satisfactory result of approximately 2% positive difference in our calculations considering the whole study area. However, as for the winter crops, the methodologies of our work and the practices of the HCSO are so different in terms of approaches, that this small overall variation is not entirely conclusive.

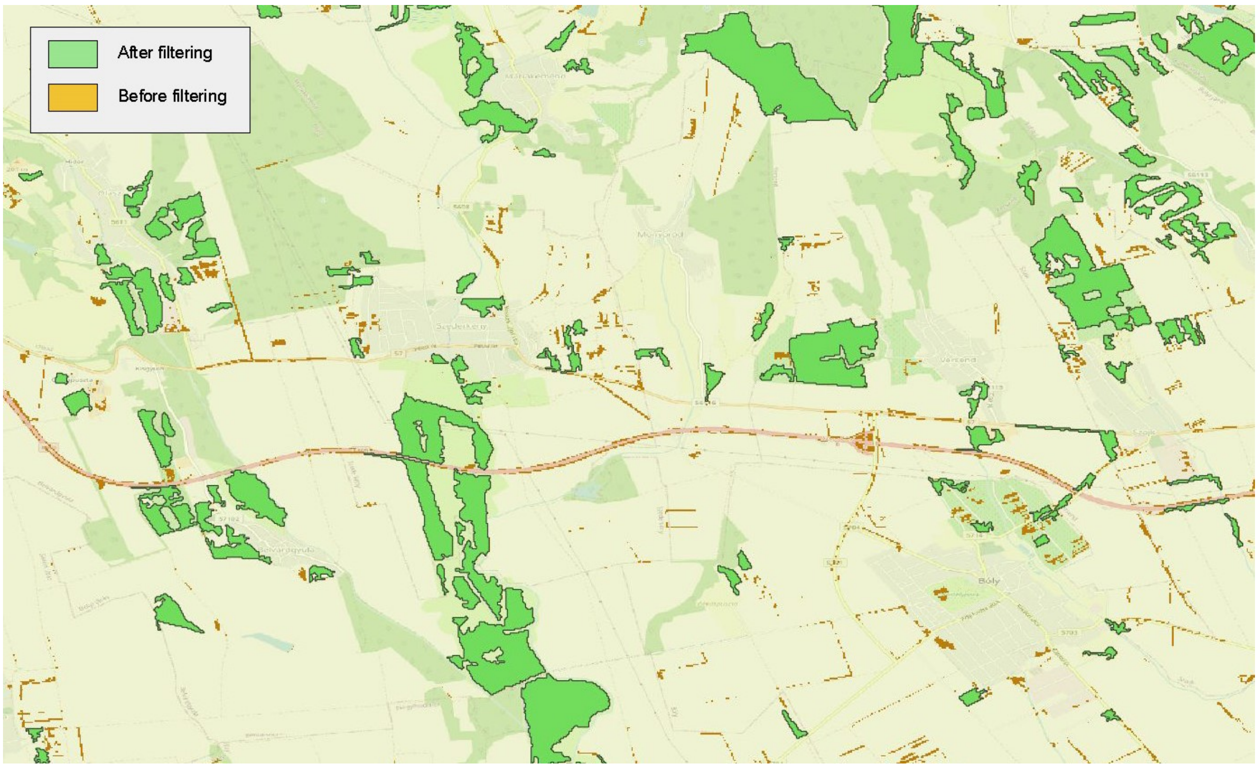

Figure 6. Grassland classification before and after the filtering process.

3.3. Accuracy assessment

The accuracy assessment of the cereals is based on the crop management data of regional farm operators currently cultivating croplands over two counties (Baranya and Tolna) in the study region. Based on this data set we could execute user accuracy assessment, during which a total of 68 barley fields were checked. From the 68 parcels, 48 fields were precisely classified. 10 fields were categorized as winter wheat fields. Two meadows also were misclassified as barley, and we had a very diverse category, where fallow lands and abundant fields were dominant and classified as barely.

In any case, it is very important to consider that the research was evaluated post-harvest by the short-term nature of the project and its time frame (Autumn of 2020). Thus, it was hard to find any other solution for accuracy assessment at the time. Additionally, the field data that we have been using could also contain some percentage of error or any other misleading information. Probably an earlier research activity, conducted in April or May would have fostered both the quality of the final product and the implementation of the accuracy assessment as well.

As for the accuracy assessment of the grasslands, 100 random samples were selected and individually validated on-site as a field survey, on which the actual land-use type subcategories (mowed meadows, fallows, etc.) were also recorded. Our method yielded an 83% overall accuracy, with alfalfa causing the most common misclassification. Comparing our quantified results with the official 2020 HCSO data for the region, the overall difference is less than 2% or 1500 acres. This difference is augmented to a 9% average on individual county levels, but that is still an acceptable margin of error considering the resolutions of the input data of the imaging sensors, and the complexity of the target.

4. Conclusions

Considering that our work covered only a fragment of the agricultural statistics and was only conducted in a single region of the country, the following general consideration could be noted:

- For a more complex approach, it is essential to harmonize the identifiable land cover categories with the nomenclature used by the HCSO to integrate our results with their existing methodology and terminology.
- The project was too short to establish a solid frame for integrating remote sensing and satellite data into the daily work of the HCSO. In general, it can be stated that the pilot was a promising initial step but setting up an operational service needs much more time and human resources.
- To improve the final product quality further, additional auxiliary datasets from farmers are essential, consisting of precise information on their cultivated plots. Not only the final

validation, but the general classification, and the preparation of the decision trees could also benefit greatly from these extra data.

- Multi-year datasets would improve the quality of the results, and in essence, the accuracy and the reliability of the classification. Especially grassland areas would benefit from this improvement, in which case, it eventually could solve most of the misclassification issues caused by herbs and alfalfa.
- Due to the varying cloud cover - which affects the optical imagery - there cannot be two results alike, which further individualize each and every processing.
- Implementing dynamic crop mapping would be an appropriate approach to further improve the results, especially the numbers associated with the barley and wheat fields.
- Important to note, that the validation phase is at least as time-consuming as the product generation itself for any given area.
- Of course, in any case, the spatial resolution of Sentinel sensors also needs to be considered. This particular limitation means that most fields under 40-50 pixels are hard to identify accurately.

Evidently, our methodology can be improved significantly with the implementation of the aforementioned solutions and findings. On the other hand, utilization of Machine Learning (ML) techniques also could provide more accurate results. Since our project ended, there have been some experiments in that regard between our partners. Certain ML algorithms performed remarkably on selected Hungarian test sites. This approach could provide new directions and effectively substitute the manually configured decision-tree based technique, even so without the utilization of optical imagery, solely deriving the results from SAR data.

In summary, we would like to point out that the research presented here, was the first project of this kind, where we tested the capabilities of Sentinel datasets in this particular domain and set the results against the high demands of the Hungarian Central Statistical Office. It was a promising start, with sufficient improvement opportunities, but in general, it might lead the way for future applications in the field of agricultural statistics in Hungary.

5. Acknowledgement

The research presented in this paper was produced as part of an ESA financed tender, identified as 'Sentinel data integration into the official statistics of the Hungarian Central Statistical Office (HCSO)'; (Contract No. 400123350/18/I-LG). The conducted work was an excellent and instructive possibility to work jointly with such a knowledgeable company as sarmap SA, to introduce a modern approach on agricultural statistics to the HCSO through the detailed utilization of the expansive data offered by ESA through the Copernicus programme. Still, we would be unable to properly contextualize our results and their importance without valuable instructions and datasets from the HCSO. We also would like to express our thankfulness to Dr Francesco Holecz and his team from sarmap SA for providing reliable software background and valuable insights to the project.

6. References

[1] Felix Stumpf F., Schneider M. K., Keller A., Mayr A., Rentschler T., Meuli R. G., Schaepman M., Liebisch F., (2020) Spatial monitoring of grassland management using multi-temporal satellite imagery, *Ecological Indicators*, Vol. 113, 106201.

[2] Abdel-Hamid A., Dubovyk O., Greve K. (2021) The potential of sentinel-1 InSAR coherence for grasslands monitoring in Eastern Cape, South Africa, *International Journal of Applied Earth Observation and Geoinformation*, Vol. 98, 102306.

[3] Pagani V. , Guarneri T, Busetto L., Ranghetti L., Boschetti M., Movedi E., Campos-Taberner M., Garcia-Haro F., Katsantonis D., Stavrakoudis D., Ricciardelli E., Romano F, Holecz F., Collivignarelli F., Granell C., Casteleyn S., Confalonieri R (2019), A high-resolution, integrated system for rice yield forecasting at district level, *Agricultural Systems*, Vol. 168, 181-190 .

[4] Milisavljević, N., Collivignarelli, F., Holecz, F. (2014) Estimation of Cultivated Areas Using Multi-Temporal SAR Data in *Holecz F., Pasquali P., Milisavljević N., Closson D.(edit.), (2014) Land Applications of Radar Remote Sensing*, Intech, Rijeka, Croatia DOI: 10.5772/55833.pp 95-120.

[5] Kellndorfer, J., Cartus, O., Bishop, J., Walker W., Holecz, F., (2014) Large Scale Mapping of Forests and Land Cover with Synthetic Aperture Radar Data in *Holecz F., Pasquali P., Milisavljević N., Closson D.(edit.), (2014) Land Applications of Radar Remote Sensing*, Intech, Rijeka, Croatia DOI: 10.5772/55833. pp 59-94.

[6] Blaes, X., Vanhalle L., and Defourny, P. (2005) Efficiency of crop identification based on opticaland SAR image time series," *Remote Sensing of Environment* vol. 96, 352-365.