Services for Earth-Observation-based statistical information for agriculture

**Keywords:** agriculture, satellite, crop type maps, crop yields, machine learning, data fusion

# Introduction

The Copernicus programme with novel Sentinel satellites has opened a new chapter in monitoring agriculture production on a country level. The projects *Agriculture Poland: Services for Earth Observation-based statistical information for agriculture* (EOStat funded by the European Space Agency) and *Satellite identification and monitoring of crops for the purposes of agricultural statistics* (SATMIROL funded by the National Centre for Research and Development) presented here exploit these new capabilities, aiming at creation of a system of Earth-Observation-based support for acquisition of statistical information for agricultural production. The projects strive to provide a suite of evaluated and documented tools for automatic processing of satellite data and derivation of products relevant for agricultural production assessment such as crop type maps and crop yield estimates. The main user of developed data products is Statistics Poland (Central Statistical Office) currently exploring innovative data sources based on satellite Earth Observation to support official statistics.

# Methods

## Crop type mapping

The crop type mapping approach builds on the random forest classifier that synergistically uses radar and optical satellite data acquired by sensors onboard Sentinel-1 and Sentinel-2 to derive a set of spatial and temporal features. These consist of radar data derivatives such as the time series of dual polarization, entropy, H/α decomposition as well as normalized time ratios that describe the evolution of the plant during the growing season. The features derived from optical data improve the delineation of spring and winter crops. The algorithms have been trained using thousands of in-situ samples: gathered during the annual nationwide field campaigns carried out by the Statistics Poland, and provided by the farmers applying for the subsidies (greening payments) under the Common Agricultural Policy. The satellite data processing for 2019 and 2020 growing seasons was performed on the CreoDIAS platform.

## Crop yield forecasting

Crop yield forecasting system estimates expected yields for main crop types in Poland based on satellite-derived and agrometeorological parameters derived from various sources. We use MODIS data to develop a long-term time series of vegetation indices starting in 2000. For operational in-season forecasts the system employs Sentinel-3 data. The agrometeorological data consist of meteorological measurements of air temperature and precipitation as well as satellite-derived solar radiation and soil moisture. The time series of the yield explanatory parameters are transformed from a standard time domain to the domain of the crop development stage. This allows for a comparison of crop biophysical status along years in the same moment of the crop development, which can be shifted from year to year due to weather conditions. Finally, the machine learning approach has been built that uses Extreme Gradient Boosting to forecast the crop yields at NUTS-2 and LAU-2 levels. The forecasting system has been implemented in R.

# Results

For 2020, the crop type map (Figure 1) provides information for 12 million parcels above 0.2 hectares that can be analysed using the Sentinel data. These represent 74% of the parcels, and 83% of the arable land. We distinguished 13 crops including 8 types of cereals. Figure 2 presents the performance assessment of the 2020 crop map based on 20% of reference parcels left for validation. The overall classification accuracy is 0.83. The worst classified class is spring rapeseed, which had very small training sample in comparison with other classes. The known problem of distinguishing summer crop types is visible in our validation for spring triticale and spring wheat, but spring barley and oat are correctly recognised.



Figure 1. The crop type map for the 2020 growing season

Table 1. Confusion matrix of classification done fusion approach based on synergistic use of optical and microwave data (2020)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **W. wheat** | **Maize** | **Grassland** | **W. barley** | **S. rapeseed** | **W. rapeseed** | **S. wheat** | **W. triticale** | **S. barley** | **Buckwheat** | **Oat** | **Rye** | **S. triticale** |
| **W. wheat** | 178 | 1 | 2 | 0 | 3 | 0 | 38 | 6 | 1 | 0 | 0 | 0 | 0 |
| **Maize** | 0 | 146 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 11 | 0 | 1 | 3 |
| **Grassland** | 0 | 1 | 185 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 2 | 2 | 3 |
| **W. barley** | 2 | 0 | 0 | 191 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 2 | 0 |
| **S. rapeseed** | 0 | 0 | 0 | 0 | 9 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| **W. rapeseed** | 0 | 0 | 0 | 0 | 45 | 201 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| **S. wheat** | 4 | 0 | 2 | 0 | 1 | 0 | 82 | 4 | 10 | 2 | 10 | 1 | 9 |
| **W. triticale** | 9 | 0 | 0 | 4 | 1 | 0 | 7 | 172 | 4 | 0 | 0 | 8 | 33 |
| **S. barley** | 1 | 0 | 0 | 0 | 0 | 0 | 13 | 3 | 161 | 0 | 5 | 0 | 1 |
| **Buckwheat** | 0 | 4 | 2 | 0 | 1 | 0 | 1 | 0 | 1 | 141 | 1 | 0 | 4 |
| **Oat** | 2 | 0 | 3 | 0 | 0 | 0 | 18 | 0 | 10 | 1 | 161 | 2 | 10 |
| **Rye** | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 6 | 1 | 0 | 0 | 174 | 8 |
| **S. triticale** | 0 | 0 | 1 | 0 | 1 | 0 | 12 | 8 | 4 | 2 | 15 | 6 | 92 |
| **Producer's accuracy accuracy** | 0.91 | 0.96 | 0.95 | 0.98 | 0.14 | 0.98 | 0.47 | 0.85 | 0.83 | 0.88 | 0.83 | 0.89 | 0.56 |
| **User's accuracy** | 0.78 | 0.90 | 0.94 | 0.97 | 0.69 | 0.81 | 0.66 | 0.72 | 0.88 | 0.91 | 0.78 | 0.91 | 0.65 |
| **F1\_score** | 0.84 | 0.93 | 0.95 | 0.97 | 0.24 | 0.89 | 0.55 | 0.78 | 0.85 | 0.90 | 0.80 | 0.90 | 0.61 |
| **Overall accuracy** | **0.83** |  |  |  |  |  |  |  |  |  |  |  |  |
| **Kappa** | **0.81** |  |  |  |  |  |  |  |  |  |  |  |  |

Figure 2 presents the results of crop yield forecasts for 2020 at NUTS-2 level. It can be seen that machine-learning-based forecasts significantly differ from the estimates provided by Statistics Poland. Overall, the yields are expected to be higher than in 2019.

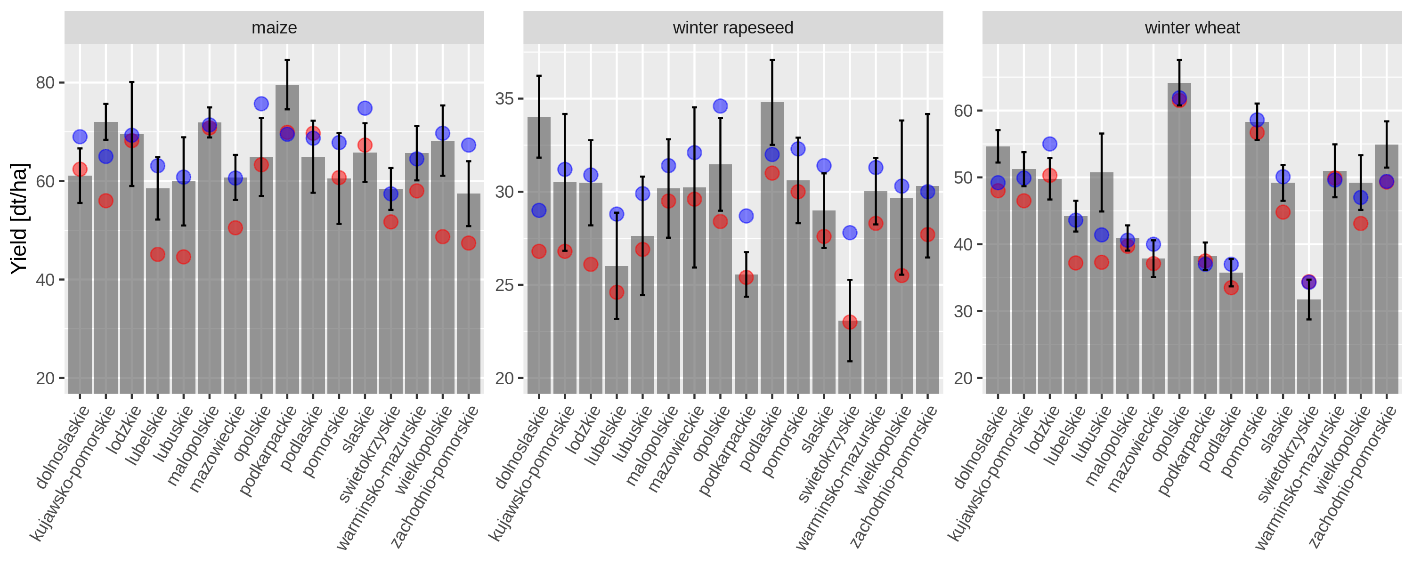


Figure 2. Crop yield forecasts for 2020 at NUST-2 level (grey bars), forecast uncertainty (vertical black lines) juxtaposed with traditional expert-based forecast of Statistics Poland (blue dots). Red dots indicate the offically reported yields for 2019.

The accuracy of the forecasting model (shown as uncertainty in Figure 2) was estimated by means of leave-one-year-out procedure. The relative root mean square error varies among NUTS-2 and crop types (Figure 3). The mean relative RMSE is 7%, 10% and 11% for winter wheat, winter rapeseed and maize, respectively.

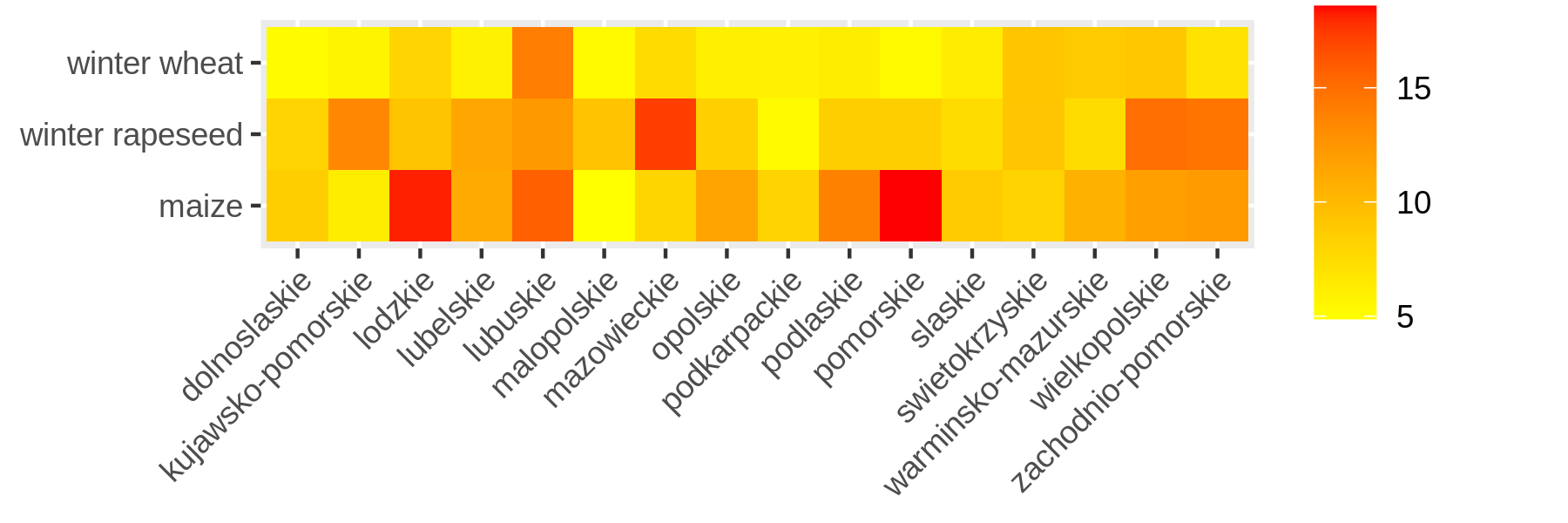


Figure 3. Relative root mean square error (%) of crop yield forecasts based on leave-one-year-out procedure for years 2003-2019.

# Conclusions

The EOStat and SATMIROL projects address specifically the needs of the Statistics Poland which is the only official national source of agricultural statistical information transmitted to the European Commission (Eurostat). The satellite-based tools developed within the projects allow for a measurement of acreages of each crop type that can replace the expert-based estimates. The concurrent yield forecasts can further lead to better approximations of the crop production on a country level and its spatial variability. We demonstrate that a fusion of satellite data, machine learning and cloud computing facilities can lead to accurate information on agricultural statistics. The methodology should significantly decrease the costs of gathering statistical information with traditional way.