Natural resources and bioeconomy studies 80/2021

CropYield—Towards pre-harvest crop yield forecasts with satellite remote sensing

Final report

Maria Yli-Heikkilä, Samantha Wittke, Mirva Kokkinen and Anneli Partala

Luke
NATURAL RESOURCES
INSTITUTE FINLAND

Natural resources and bioeconomy studies 80/2021

CropYield—Towards pre-harvest crop yield forecasts with satellite remote sensing

Final report

Maria Yli-Heikkilä, Samantha Wittke, Mirva Kokkinen and Anneli Partala

Natural Resources Institute Finland, Helsinki 2021

Recommended citation:

Yli-Heikkilä, M., Wittke, S., Kokkinen, M. & Partala, A. 2021. CropYield—Towards pre-harvest crop yield forecasts with satellite remote sensing : Final report. Natural resources and bio-economy studies 80/2021. Natural Resources Institute Finland. Helsinki. 28 p.

The project was funded by the European Union (grant 831735 – 2018-FI-CROPYIELD).

Maria Yli-Heikkilä, ORCID ID, https://orcid.org/0000-0003-1528-7246

ISBN 978-952-380-307-7 (Print)
ISBN 978-952-380-308-4 (Online)
ISSN 2342-7647 (Print)
ISSN 2342-7639 (Online)
URN http://urn.fi/URN:ISBN:978-952-380-308-4

Authors: Maria Yli-Heikkilä, Samantha Wittke, Mirva Kokkinen and Anneli Partala
Publisher: Natural Resources Institute Finland (Luke), Helsinki 2021
Year of publication: 2021
Cover photo: Anneli Partala
Printing house and publishing sales: Juvenes Print, http://luke.juvenesprint.fi

Summary

Maria Yli-Heikkilä[1], Samantha Wittke[2,3], Mirva Kokkinen[1] and Anneli Partala[1]

[1)] Natural Resources Institute Finland (LUKE), Latokartanonkaari 9, FI-00790 Helsinki, Finland
maria.yli-heikkila@luke.fi
[2)] National Land Survey of Finland
[3)] Aalto University, Finland

The general objective of this project was to enhance the crop statistics. To this end, we established a pilot case for an automated process for improved crop yield statistics by merging Earth observation (EO) data, the administrative data, agro-meteorological data and historical crop statistics survey data. The significance of the approach is that the previously very laborious data acquisition process from different sources and the processing of multistep modelling is now by design fully automated, and can thus reduce spending on professional surveying. The main achievement is that a new artificial intelligence-based crop yield forecasting system can produce pre-harvest yield predictions for four main cereals (oats, barley, wheat, rye).

Surveys are very costly in terms of time and expense. The same is true of gathering expert estimates on regional crop yield forecasts. During the last decade, EO systems have been shown to provide an effective means for large-scale crop monitoring and yield estimations. In this sense, this project has fulfilled its promise to establish a pilot case for an automated process of improved crop yield statistics by merging EO data and a data-driven modelling approach. As a result, we can produce several in-season crop yield forecasts, the first already in late June, around the same time as the Joint Research Centre's European-wide forecast. From then on, the forecasts can be published, for example, at 10 day-intervals.

The machine learning models implemented in this project achieved a highly promising level of accuracy in pre-harvest yield predictions for four main cereals (oats, barley, wheat, and rye) when compared to the Joint Research Centre's and LUKE's seasonal forecasts. However, the problem of choosing the best model remains. There was no clear winning model that reliably predicts yields at all times. Therefore, a model comparison will be the most important developmental task ahead.

In the context of agricultural statistics, more accurate in-season forecasts of crop yields benefit sustainable agriculture and food security with better informed political decisions. In addition, reliable crop forecasts have market impacts. Moreover, EO-based applications can be globally applicable. We expect that within a few years our EO-based crop forecasting will be proven to be a sound method to replace in-season regional expert estimates, and in the foreseeable future it will also gradually replace annual farm surveys.

The uptake of EO as a new data source in statistical production was more complex than initially expected. There are a myriad of approaches to monitoring crop yields, the main decisions to make being whether to utilise: i) optical or radar satellite data or both, ii) image mosaics or single images, iii) pixel-based or object-based image analysis. In addition, remote sensing requires specialized expertise, not to mention the specialized expertise needed in predictive modelling. We acknowledged the lack of remote sensing expertise and made a decision at the start to outsource the pre-processing of satellite images. With outsourcing the sustainability of the project may be jeopardized if the know-how outsourced cannot be fully transferred to the statistical production. In this sense, one significant achievement in the project has been the

uptake of EO *knowledge,* with a substantial contribution from the National Land Survey of Finland, which became a sound part of our production system. As a result, we have the in-house readiness to apply EO as a new data source also to other statistical themes.

It was concluded that country-wide forecasts seemed to work already in June, probably due to the inherited sampling weights from the crop production surveys. However, for the regional forecasts the sampling data was inadequate. For regional forecasts, we would need to sample fields to gain an equal spatial coverage. Moreover, for the northern regions the crop forecasting is reasonable only from July on due to the later sowing dates. Therefore, further study is needed to evaluate the best physiologically grounded observation window for each region.

Deploying the forecasting pipeline requires further automatization. Especially at the end of the pipeline the validation of the results needs further scrutiny. Uploading the predictions to statistical production databases requires modifications to existing ICT-systems. In addition, prediction model architectures need to be revised and improved along with the new data from the coming years.

Keywords: Earth observation, remote sensing, crop yield, agricultural statistics, machine learning

Contents

1. Introduction

The Earth observation (EO) systems provide an effective means for large-scale crop monitoring and yield estimation. The current radar and optical remote sensing satellite imagery has a fairly high spatial and temporal resolution and contains a wealth of information on vegetation growth. In the near future new missions are planned for launch. Agricultural monitoring has long utilised satellite remote sensing to estimate product yields. Traditionally, crop yield forecasting has been based on a single or an ensemble of physiological models of crop growth. For example, process-based crop growth models such as the Joint Research Centre's (JRC) MARS Crop Yield Forecasting System utilise satellite-derived data as a complementary source of information.

With the increasing volume and variety of remote sensing data, there is a need to develop a fully data-driven model for crop yield forecasting. In this project, we established a pilot case for an automated process for improved crop yield statistics by merging EO data, administrative data, agro-meteorological data, and historical crop statistics survey data. The novelty of the approach is that the previously very laborious data acquisition from different sources and processing of multistep modelling is now by design fully automated, and thus, can reduce spending on professional surveying. The new artificial intelligence-based crop yield forecasting system can produce pre-harvest yield predictions for four main cereals (oats, barley, wheat, rye).

In Finland, the growing season starts from early May till harvesting time in late August or September. Currently, crop yield forecasts are produced twice: at the end of July and August. The in-season forecasts are based on expert estimates from the regional agricultural advisory bodies. The forecasts are published on a national level. The final statistics based on the Finnish crop production survey are published in February–March the following year. The final statistics are published on a regional level (similar to NUTS3). The Finnish crop production survey is based on stratified sampling and is conducted by the Natural Resources Institute Finland (LUKE). In this project, we produced crop yield forecasts for June, July, August and finally for a complete time series at the end of the growing season in September. With a fully automated forecasting system, we can produce the first forecast in late June, which is around the same time as JRC's European-wide forecast. Forecasts can be published, for example, at 10 day-intervals.

This report describes the forecasting system in detail. In the final phase of the project, we produced forecasts for the year 2020. The model was trained with the data from the previous years: 2016–2019. Here, we report the processing times and amount of data for the year 2020 only. For data storage and computing we used the Puhti supercomputer hosted by the CSC – IT Center of Science, which is a center of expertise in information technology owned by the Finnish state and higher education institutions.

2. Model evaluation

This section describes the model development and results in detail. Figure 1 below serves as a graphical abstract of the model development.

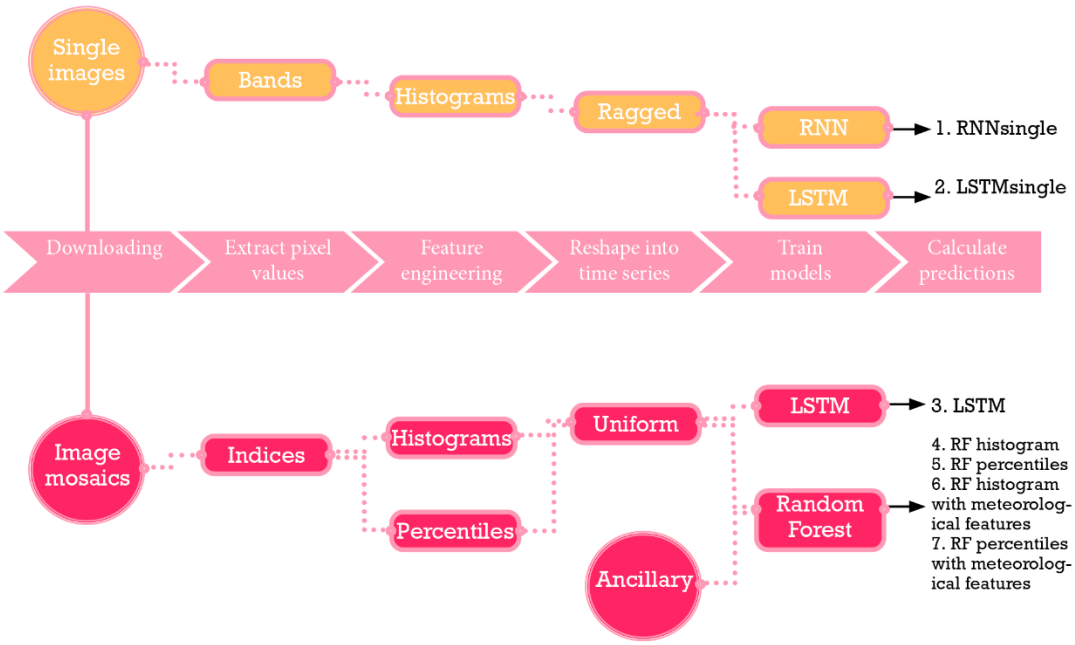

Figure 1. The scheme of the model development.

2.1. Research plan(s)

According to our initial research plan (A), crop yields would be predicted from a sequence of remotely sensed multispectral images, where each image corresponds to a different day of the year (DOY) within a growing season. In the course of the project, we also established a new research line (B) based on mosaicked observations of indices. Figure 2 illustrates the two lines of research.

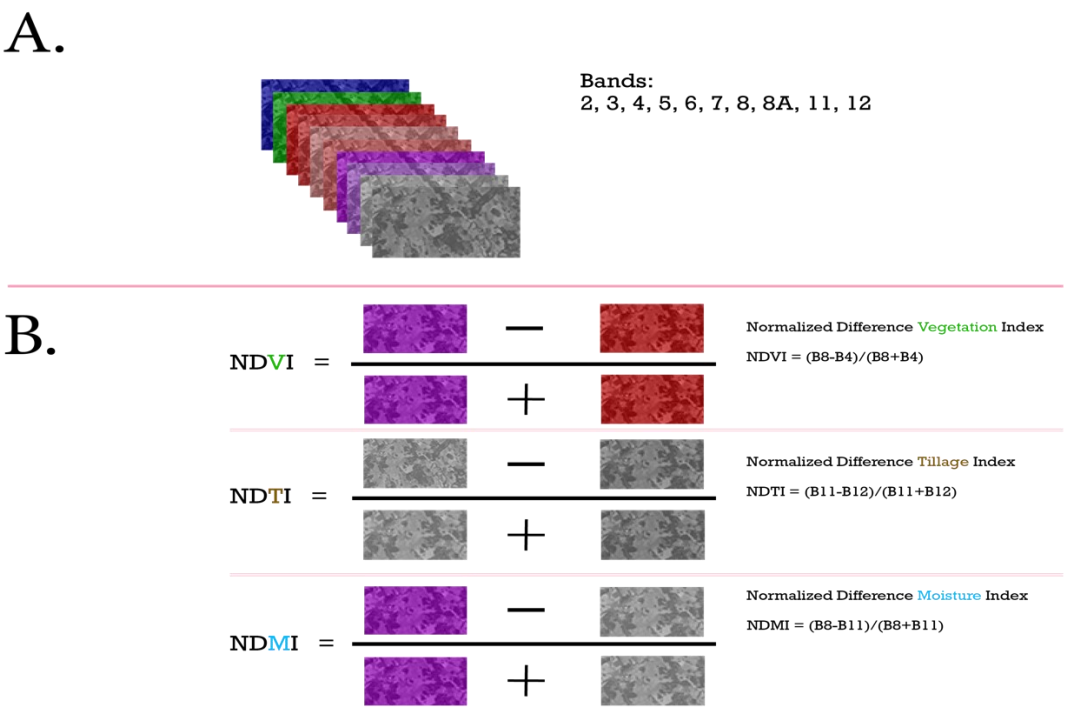

Figure 2. Two lines of research were followed. Plan A is based on single multispectral images from Sentinel-2. Plan B. is based on mosaicked observations of three normalized difference indices.

2.1.1. Reshaping time series into analysis ready data

Multispectral images from Sentinel-2 have 13 bands, of which 10 are usable for crop monitoring. Our initial hypothesis was that data from all available images and all bands should be utilized (plan A) as compared to the traditional approach to compress the band information into indices and to stack observations into fixed-length time windows (plan B). The prevailing assumption in neural network applications is that data should be feed with minimum pre-processing modifications to avoid the loss of information. Hence, the plan A can be labelled a data-driven approach, whereas plan B is a compromise of compressing information into a user-friendly format for downstream modelling. This section describes how satellite and ancillary data were reshaped into an analysis-ready format in both cases.

Single multispectral images

Following the data-driven approach, all relevant spectral bands for land monitoring were used, that is, bands 2, 3, 4, 5, 6, 7, 8, 8A, 11 and 12. The excluded bands were related to observing coastal or inland waters or atmospheric aerosols (band 1), water vapour (band 9), and cirrus reflectance (band 10). We used Level 2A products which are images of geometrically corrected bottom-of-atmosphere reflectances. The product also includes cloud masks, but we did not use them. Instead, again, following the data-driven approach, we preset a 'healthy' range for the pixel values for each band. Namely, by calculating the percentiles of all pixel values within the training set for each crop type, we used the 1st and the 99th percentiles as the 'healthy' minimum and maximum values. Values outside this range were ignored. We believe values

close to zero and at the upper end are either biased reflectance values or represent clouds, and therefore do not contain any relevant information for this study.

The observable unit in the study was a farm because our reference data were the average crop yields on the farm level. More specifically, the spatial observable unit was a field (polygon) or a group of fields (multi-polygon) if a farm was growing the crop in question on several fields. First, pixel values were extracted from a (multi-)polygon in the satellite image. Here we used a 'healthy' range to arrange polygon-wise pixel values into 32 bin normalized histograms. Thus, at each bin, the value is the value of the probability density function normalized such that the integral over the range is 1. For each band we have 32 features.

Finally, for each observable unit (farm) we have multiple observations from the growing season. In our data set we had satellite acquisitions from up to 115 days between 10/5 and 01/09. Due to cloudiness and satellite revisit times, we had significantly fewer observations within a growing season. Therefore, the time series are variable-length (ragged), and the missing time steps need to be padded with zeros. For example, for predicting oats in 2020, we have 13,677 parcels, 115 zero-padded time steps, and as features 32 bins from 10 bands, altogether 320 features. The data shape is three dimensional (13677, 115, 320).

Index mosaics

Image mosaicking is a technique used to assemble multiple overlapping images together. An image mosaic is an integral representation of a scene produced from the most optimal pixel values. We utilized pre-processed 30-day mosaics of Sentinel-2-based indices covering the whole of Finland.[1] The mosaics were computed on the 15th and the last day of month, and therefore, we had two observations available per month. 30 days were compressed into one value based on the maximum NDVI. That is, for all index mosaics, the pixel values were selected from the image with maximum NDVI value with the given time window. Figure 3 illustrates how the mosaics are computed from the time windows.

[1] Sentinel-2 image mosaics are produced by the Finnish Environment Institute SYKE using the CalFin-cluster of Sodankylä National Satellite Data Center, and they were developed as part of the sub-programme "Distribution and Processing of Satellite Imagery" in the "Geospatial Platform of Finnish Public Administration" programme (2017–2019).

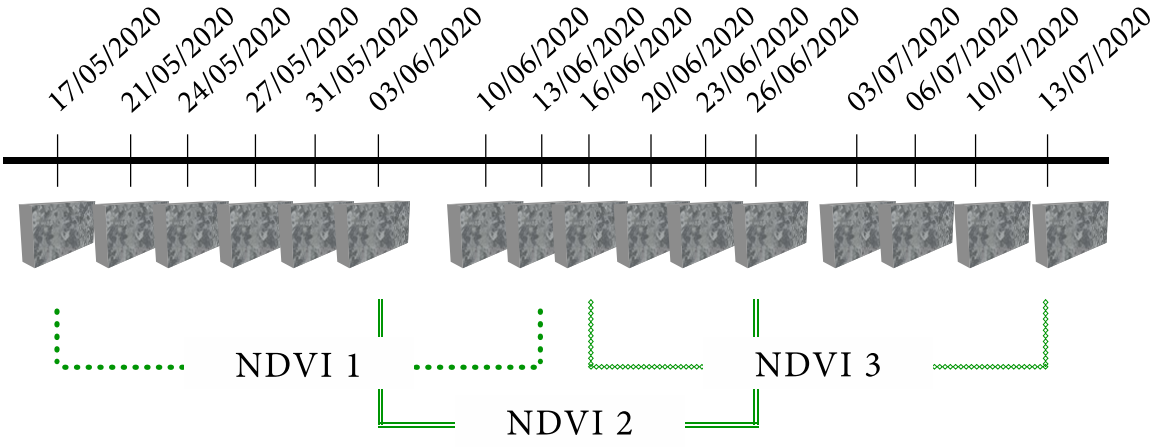

Figure 3. A mosaicked Normalized Difference Vegetation Index (NDVI) is calculated by taking the maximum NDVI value of all available observations within 30 days. The 30-day windows overlap by15 days.

From image mosaics we get a full time series. We used all the available image mosaics the first being from April 01–30/04 and last from October 01–30. Thus, we have 13 time steps.

The Sentinel-2 image index mosaics used in this study were: the Normalized Difference Moisture Index (NDMI), the Normalized Difference Tillage Index (NDTI) and the Normalized Difference Vegetation Index (NDVI). Pixel-wise indices are calculated from Sentinel-2 image bands (B) as follows:

NDVI = (B8-B4)/(B8+B4)

NDTI = (B11-B12)/(B11+B12)

NDMI = (B8-B11)/(B8+B11

Similar feature engineering was carried out as with the single images. Namely, by calculating the percentiles of all the pixel values per index within the training set for each crop type, we used the 1st and the 100th percentiles as 'healthy' minimum and maximum values. Between the 'healthy' range we arranged pixel values polygon-wise into 16 bin normalized histograms. For example, for predicting oats in 2020, we have 13,677 parcels, 13 time steps, and as features 16 bins from 3 indices, making altogether 48 features. The data shape is three dimensional (13677, 13, 48).

In the above, we have considered observations as time series from the growing season. Recurrent neural networks are applicable to time series. However, the decision tree-based machine learning method Random Forest has proven out to be a good and robust method even for time-series tasks, although it cannot grasp the time dimension. We chose to use RF as a baseline learner for neural networks. The input data for RF needs to be 2-dimensional. Therefore, we stacked the time steps and features into one dimension resulting in a data shape of (13677, 624) with image mosaics.

2.1.2. Ancillary meteorological data

Meteorological data are considered significant predictors in traditional crop growth models. Therefore, we tested whether the RF model improved with ancillary meteorological features. For each (multi-)polygon, a centroid was calculated. Based on the centroid, the cumulative precipitation, solar radiation, and temperature were fetched from a weather database for the following dates: 01/06, 15/06, 01/07, 15/07, 01/08, 15/08, 01/09. This resulted in 13 time steps x 3 indices x 16 bins + 7 time steps x meteorological features. Accordingly, for predicting oats in 2020, the data set has a shape of (13677, 645). See Section 2.3.2. for results.

2.1.3. Alternative feature engineering approach

In the above, we used normalized histograms with a predefined range as compressed features from pixel values. A common method to prepare data for modelling is to take only the mean value per observable unit, or alternatively the median, 25^{th} and 75^{th} percentiles. We tested whether histograms outperformed the percentile/mean approach. From all the pixels per (multi-)polygon we extracted the mean, median, the 10^{th}, 20^{th}, 30^{th}, 40^{th}, 60^{th}, 70^{th}, 80^{th}, and 90^{th} percentiles. We only tested this with image mosaics for the RF. Instead of 16 bins per index, we now have 10 features. Thus, the shape of the data is (3400, 13x30). See Section 2.3.2. for results.

2.2. Modelling

Recurrent neural networks (RNNs) (Rumelhart et al., 1986) were chosen for processing the sequential data from a chain of satellite images. RNNs can learn to recognize temporally extended patterns in time series. Long-Short Term Memory networks (Hochreiter and Schmidhuber, 1997) are a special type of RNNs. We compared RNN methods to a Random Forest (RF) (Breiman, 2001) decision tree-based machine learning method. The fundamental difference between RNNs and RF is that while RNNs can consider the temporal dynamics in the data, RF does not explicitly capture that.

2.2.1. Vanilla Recurrent Neural Network (RNN)

Neural networks consist of connections from the input layer to the output layer. In between, there are usually one or more hidden layers. In the hidden layers, the Vanilla RNN stores an internal state and updates it after each time steps. In the end, the hidden state is a function of all previous hidden states. This way, the network carries information in time. We can say it remembers the earlier phases.

2.2.2. Long-Short Term Memory Network (LSTM)

An LSTM is a variant architecture of an RNN that often outperforms the Vanilla RNN especially in long time series. An LSTM has the ability to carry a long-term memory by slowly changing weights. At each time step it considers whether to update the internal state or not with the help of four gatekeepers.

2.2.3. Random Forest (RF)

RF is a powerful and robust decision tree-based learning approach. Regression RF is an ensemble method, where the forest prediction for a particular test point (farm) is the average of 500 tree-based predictors. Note that the RF approach does not particularly capture the time

dimension. Each time step is merely added to the input data as a new independent static variable. While RNNs require lot of developmental work to build a network architecture which is suitable for the prediction task and to fine-tune the hyperparameters, RF performs quite well with simple 2D data and with default parameters. We used RF as a baseline method for more complicated RNNs, but also to test alternative feature engineering methods (histograms or percentiles), and whether meteorological features improved the predictive performance.

2.3. Results

In this section we present results from our research lines A and B (see Figure 1 for the research line schemes). From single satellite images we processed all available Sentinel-2 data, i.e., from 2016–2020. From these data sets we trained two RNN models: RNNsingle and LSTMsingle.

The openly available pre-processed image mosaics covered the years 2018–2020. From these image mosaics we produced the LSTMmosaic and the RFmosaic models. Finally, we trained the models with data from the years 2016–2019 (single) or 2018–2019 (mosaic) and tested the models with year 2020 data.

2.3.1. Seasonal forecast for the year 2020

We developed prediction models for four main crop types. We have separate models for winter wheat, spring wheat, rye, feed barley, malting barley, and oats. To compare the performance of the prediction models, we collected public forecasts for reference. JRC publishes forecasts for all the aforementioned crops except for oats. Forecasts are published in the JRC MARS Bulletin in the middle of June and at the end of July and August.

LUKE publishes crop forecasts in the middle of July and at the end of August. The preliminary crop statistics based on a farmer survey are published at the end of November. In this section, we present figures forecasting the crop yields for 2020. 'Final' means predictions based on all the satellite data for the growing season. The final LUKE forecast are the preliminary crop statistics published in November. This can be considered as the ground truth and is denoted here with the red horizontal line. Predictions for June, July and August are based on the satellite images acquired by the 15th of the passing month.

Here we present results for spring wheat and feed barley. Figure 4 shows how the LSTMsingle approach was closest to the Final spring wheat yield estimate in June. The RFmosaic performed best in July but was over optimistic again in August. This year the JRC and LUKE forecasts, as well as mosaic predictions seemed to be in consensus for a higher yield. For the feed barley forecasts in Figure 5 there was a consensus between all the rivals, except that the RNNsingle approach predicted overly low yields. However, surprisingly, its companion LSTMsingle was very close to the Final yield estimate already for June and throughout the season.

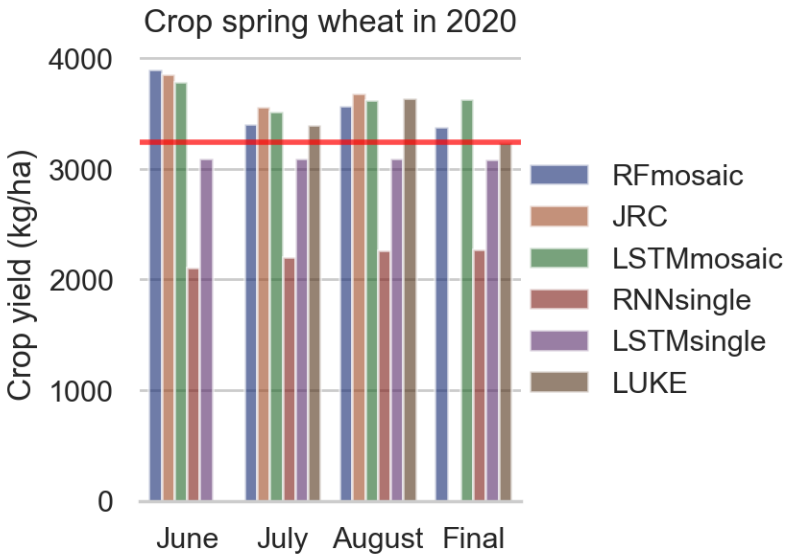

Figure 4. In-season and post-harvest crop yield forecast for spring wheat in 2020.

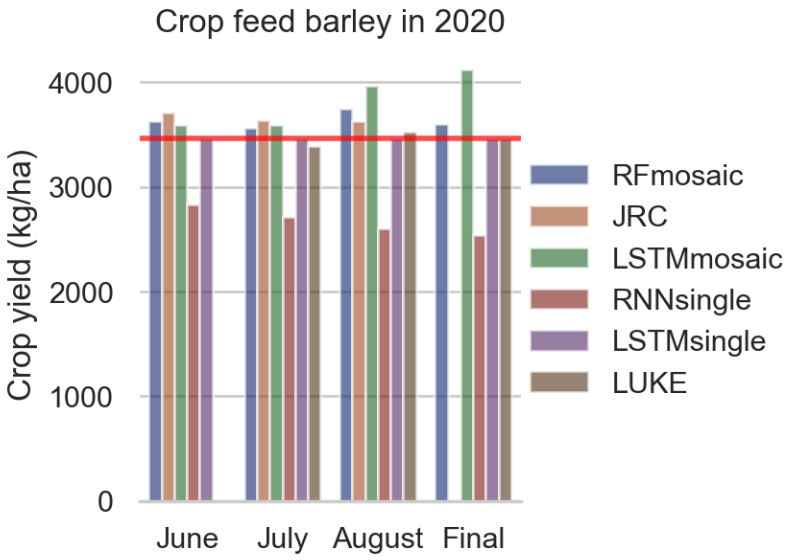

Figure 5. In-season and post-harvest crop yield forecast for feed barley in 2020.

2.3.2. Testing alternative feature engineering and meteorological features

In this section, we present results for two crop type, spring wheat and feed barley. We tested how two different feature engineering methods (histograms vs. percentiles) performed, and how adding meteorological features effected the model accuracy. The meteorological features consisted of the cumulative precipitation, solar radiation, and temperature (see Section 2.1.2. for details). The metric for accuracy was the root mean squared error (RMSE), and this means how much on average the prediction per farm deviated (+ or -) from the yield the farmer had declared. The results are from the test set. The data set was pooled from the years 2018–2019.

In the following Figures 6 and 7, all RF models were trained from image mosaic data for spring wheat and feed barley, respectively. RF-percentilesMeteo means that from each (multi-)polygon all the pixel values are compressed into percentiles (see Section 2.1.3. for details) and means. These and the meteorological features are the features for the model. RF-histoMeteo

means that from each (multi-)polygon all the pixel values are compressed into histograms (see Section 2.1.1. for details), and these plus the meteorological features are the features for the model. The RF-percentiles model was trained with only percentiles+mean features, The RF-histo model was trained with histogram features. The LSTM model was trained from image mosaics data with histogram features, and it is shown here just for comparison.

Figure 6 shows that in June, the spring wheat yield forecasts from RF-histoMeteo and RF-histo were slightly superior to rivals with percentiles. The number of farms in the training set was 2072. The R^2 for the Final RF model was quite rather high evenly for all model variants:

- percentilesMeteo 0.61
- histoMeteo 0.60
- percentiles 0.60
- histo 0.58

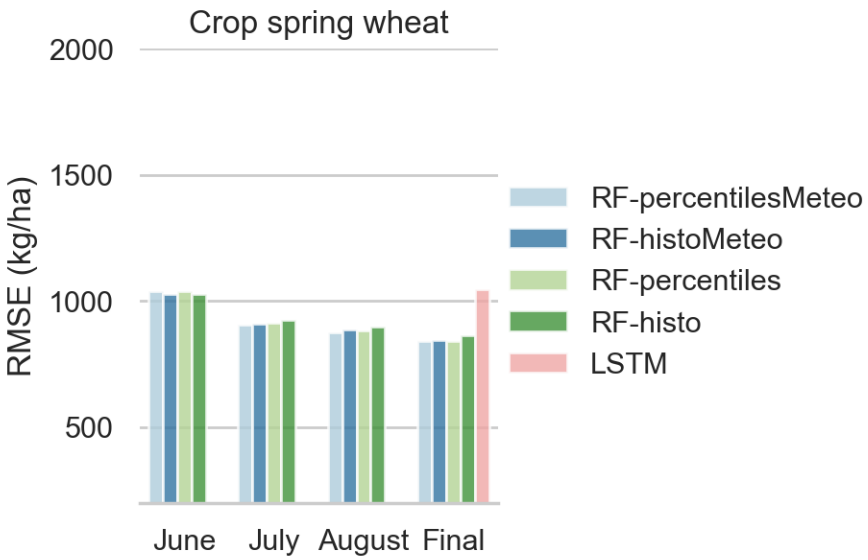

Figure 6. Model accuracies in in-season and post-harvest crop yield prediction for spring wheat in 2020.

Figure 7 shows that in June, the feed barley yield forecasts from RF-histoMeteo and RF-histo models were slightly superior to the rivals with percentiles. However, the performance of the histogram model variants did not improve later in the season, whereas the percentile models made performance improvements by the end of the season. The R^2 value for the Final RF model was higher for the percentile model variants compared to the histogram models:

- percentilesMeteo 0.45
- histoMeteo 0.29
- percentiles 0.45
- histo 0.27

The number of farms in the training set was 3,190. Adding meteorological features did not seem to improve predictive performance with any model variant.

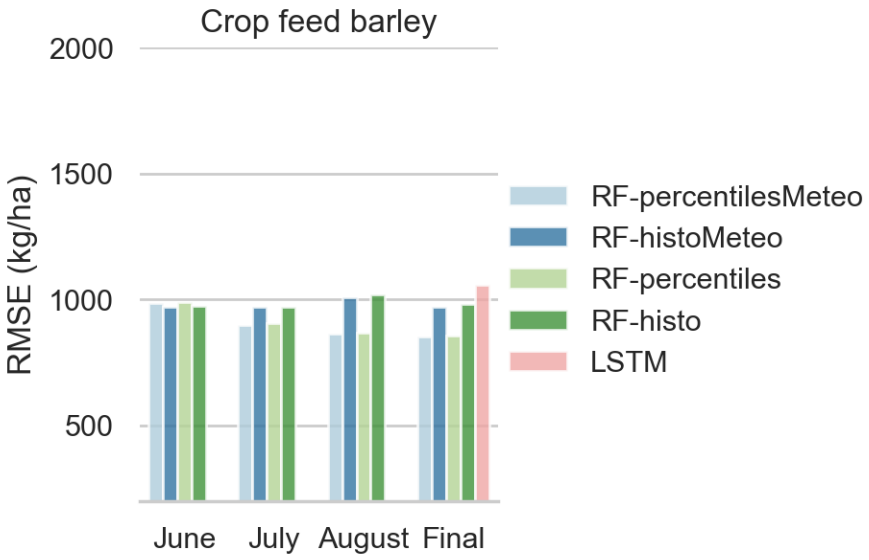

Figure 7. Model accuracies in in-season and post-harvest crop yield predictions for feed barley in 2020.

In the light of results from all the crop types under scrutiny, we conclude that neither the histogram nor the percentile approach yielded a superior performance, except that the percentiles method performed better for feed barley. Adding meteorological features may improve the model, but only slightly. We also saw a decreasing trend in the model error from the first predictions in June to the last prediction (Final). This was also expected as in June there was less information available, and at the end of the season the learner benefited from the full time series from the growing season.

2.3.3. Model comparison

For a model comparison on an aggregated country-level we can calculate crop yield predictions over multiple years and compare these predictions to averaged crop yield statistics. In Table 1 we have calculated the mean crop yield from the crop statistics over 2016–2020 in comparison to the post-harvest mean crop yield predictions from LSTM and RNN models trained on single satellite images from the same years. In Table 2 we have calculated the mean crop yield from the official crop statistics over 2018–2020 vis-à-vis test set predictions from the same years. These models (RF and LSTM) were trained on image mosaics. We can see that in this light LSTM performs better than RNN with a single image approach. However, in all cases, models trained on image mosaics performed better. The RFmosaic model performed better than the LSTMmosaic model in 5 out of 6 cases.

Table 1. Mean crop yields (kg/ha) over 2016–2020 in comparison to the mean post-harvest LSTMsingle and RNNsingle predictions and the difference (Δ) of the predictions from the crop statistics for the same period.

Crop type	Crop statistics	LSTMsingle	ΔLSTMsingle	RNNsingle	Δ RNNsingle
Winter wheat	4,140	3,553	-586 (-14%)	2,244	-1895 (-46%)
Spring wheat	3,650	3,244	-405 (-11%)	3,222	-427 (-12%)
Rye	3,648	3,234	-413 (-11%)	1,667	-1980 (-54%)
Feed barley	3,704	3,412	-291 (-8%)	2,730	-973 (-26%)
Malting barley	3,880	3,700	-179 (-5%)	1,639	-2240 (-58%)
Oats	3,510	2,969	-540 (-15%)	2,674	-835 (-24%)

Table 2. Mean crop yields (kg/ha) over 2018–2020 in comparison with the mean post-harvest RFmosaic and LSTMmosaic predictions and the difference (Δ) of the predictions from the crop statistics for the same period.

Crop type	Crop statistics	RFmosaic	Δ Rfmosaic	LSTMmosaic	Δ LSTMmosaic
Winter wheat	4,186	4,240	54 (+1%)	4,078	-108 (-3%)
Spring wheat	3,446	3,547	100 (+3%)	3,836	390 (+11%)
Rye	3,660	3,660	0 (0%)	3,698	38 (+1%)
Feed barley	3,650	3,536	-113 (-3%)	4,125	475 (+13%)
Malting barley	3,713	3,603	-109 (-3%)	3,770	57 (+2%)
Oats	3,466	3,424	-41 (-1%)	3,701	234 (+7%)

2.4. Seasonal forecasting assessment

Here we note a few key points for assessing the seasonal forecasts. With single images, we should be able to detect seasonal dynamics, whereas with mosaic composites, tracking physiological dynamics is somewhat lost. However, the image mosaics still show smooth curves of NDVI, NDTI, and NDMI, whereas the band intensities from the single images are hard to interpret. Figures 8 and 9 show how the data looks at the farm level.

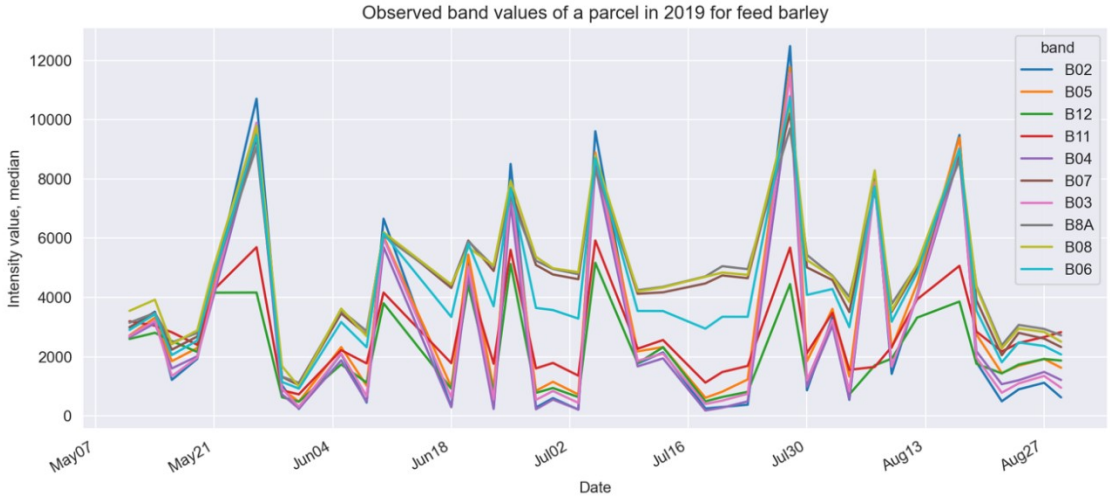

Figure 8. The median intensity values of ten Sentinel-2 bands from a parcel growing feed barley.

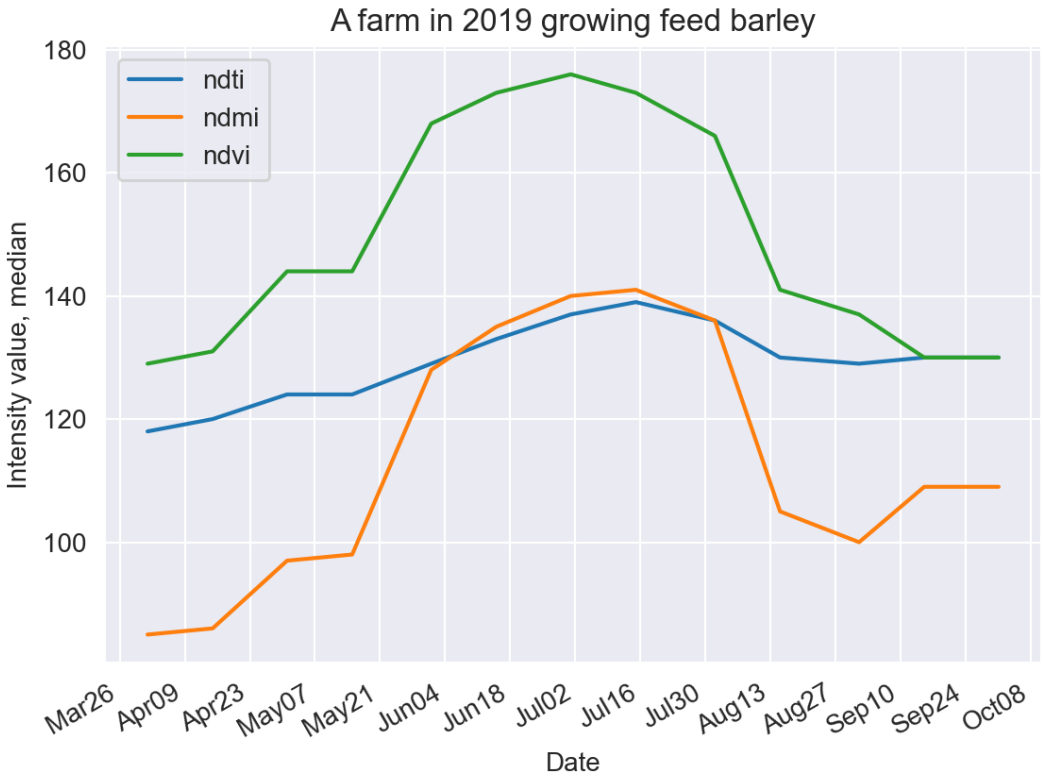

Figure 9. The median intensity values of NDVT, NDMI, and NDVI from the parcels of a farm growing feed barley.

The greenness index NDTI and the moisture index NDMI in Figure 9 form nice bell curves peaking in June and July, as expected. The tillage index NDTI is related to cellulose and lignin absorption features of the vegetation and presents here as a surprisingly flat bell curve. The NDTI is to detect crop residues and senescent, nonphotosynthetic physiological phase of the crop in the autumn. In this data, the NDTI shows only a moderate increasing trend towards the end of the season.

The single images with ten bands should detect seasonal dynamics, such as physiological stress and photosynthetic activity, on a full scale. However, surprisingly, no clear trend can be detected in this case. More-in-depth analysis of the growing season agro-meteorological conditions could be conducted to illuminate what is happening in the Figure 8. Nevertheless, the reflectance data is most probably suffering from pervasive cloudiness and shadow contamination, as no cloud filtering was applied. This also explains why the farm-level mean square errors are quite high, whereas the large-scale predictions is fairly accurate.

3. Crop production survey as a reference

This section describes how we utilized data from crop production surveys for forecasting. The Finnish crop production survey is based on stratified sampling and is annually conducted by the Natural Resources Institute Finland. The total number of agricultural holdings in Finland is about 45,000 and around 6,000 (13%) participate in the farm survey each year.

3.1. Pre-processing

First, we applied some pre-processing steps. First of all, at the model training time the statistical unit used was a farm because as a reference we had average yields per farm for six crops. At forecasting time, the observational unit used was a field.

In satellite images, the smallest observational unit is a pixel. Sentinel-2 images have a resolution of 10 by 10 meters. Therefore, we decided to filter out all fields smaller than 1ha. This ensured we had enough pixels overlaying a field. We also checked whether all the geometries were valid. Table 3 shows how the area and number of fields in the sample was divided between crops.

Table 3. Area of crops and the number of fields and their share in the crop production sample in 2020.

Crop	Total area in the sample (ha)	No. of fields in the sample
Autumn wheat	5,545 (3%)	1,157 (2%)
Spring wheat	37,903 (19%)	8,483 (17%)
Rye	3,789 (2%)	860 (2%)
Feed barley	78,511 (40%)	20,334 (41%)
Malting barley	12,850 (7%)	2,665 (5%)
Oats	56,774 (29%)	15,665 (32%)
Total:	195,373 (100%)	49,164 (100%)

3.2. Representativeness of the reference data

The sampling design for the crop production survey follows multistage weighted sampling. Most of the weight is determined by the regional share of total harvested area for the main crops in Finland. Therefore, we do not have equal spatial coverage of farms in Finland, but most of the farms come from the high agricultural productivity regions. Other variables determining the weights in the sampling design are the production type and economic size of the farm.

When testing the prediction models, we used the sample for the crop production survey for the test year (2020). Therefore, we have inherited the weighting from the sample design into the predictions as well. The weighted sampling explains why we have quite an accurate average yield prediction for the whole country. However, for producing regional aggregates, the weighted sampling fails. There are very few farms in the sample from some areas (see for example the numbers of malting barley parcels in Table 4). Therefore, for regional predictions we should increase the number of observations, and not follow the crop production sampling plan.

Table 4. Number of field parcels in the testing set in 2020 per crop type in the regions. Please refer to the region number on the map in Figure 11.

Region	Autumn wheat	Spring wheat	Rye	Feed barley	Malting barley	Oats
1	181	1,270	91	569	552	1,017
2	381	2,155	190	2,209	847	1,915
3	74	1,166	124	1,743	218	1,551
4	111	751	110	1,160	658	1,613
5	74	702	49	1,190	90	1,594
6	92	497	75	796	117	957
7	10	116	16	532	0	447
8	10	100	14	1,786	4	522
9	0	75	20	547	13	430
10	13	92	15	627	6	552
11	100	840	80	3,372	71	2,330
12	35	403	27	2,985	79	1,281
13	30	175	12	2,443	2	1,236
14	0	27	5	151	0	112
15	0	1	0	64	0	20

Table 5. The starting and ending windows of regional sowing dates in 2016–2020 for all crops. Source: LUKE, 2021.

Region	Start	Finishing
Uusimaa	20.4.–7.5.	29.5.–2.6.
Southwest Finland	15.4.–7.5.	27.5.–8.6.
Satakunta	25.4.–10.5.	25.5.–10.6.
Pirkanmaa	20.4.–10.5.	25.–31.5.
Häme	22.4.–9.5.	26.5.–1.6.
Kymenlaakso	24.4.–10.5.	19.5.–5.6.
South Karelia	24.4.–12.5.	24.5.–1.6.
South Savo	25.4.–8.5.	6.–7.6.
North Savo	7.–16.5.	25.5.–10.6.
North Karelia	11.-18.5.	31.5.–10.6.
Central Finland	29.4.–17.5.	23.-31.5.
South Ostrobothnia and Ostrobothnia	23.4.–17.5.	23.5.–5.6.
North Ostrobothnia	5.–29.5.	3.–20.6.
Kainuu	4.–14.5.	19.–25.6.
Lapland	23.–29.5.	5.–30.6.

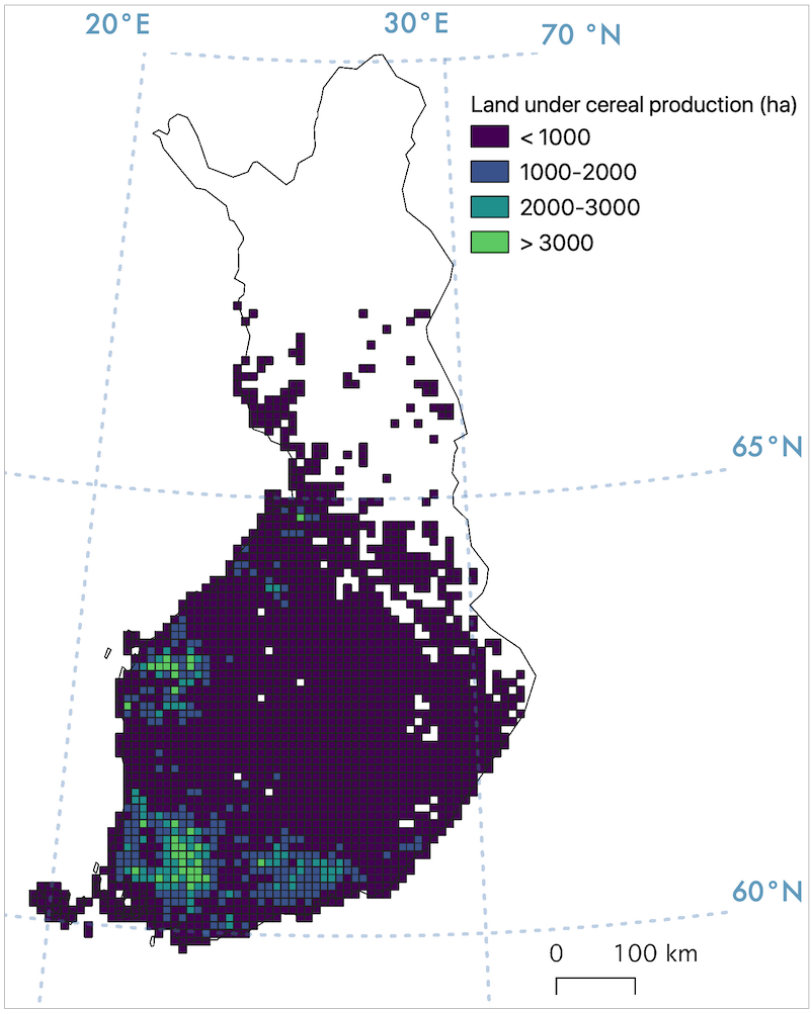

Figure 10. Land under cereal cultivation (ha) in 2020 based on Finnish Food Authority's Agricultural Parcel Registry.

Another issue to bear in mind is the wide difference of the cropping systems in the south of the country compared to the north. The majority of Finnish arable land is located between the latitudes of 60 and 65°N (see Figure 10).

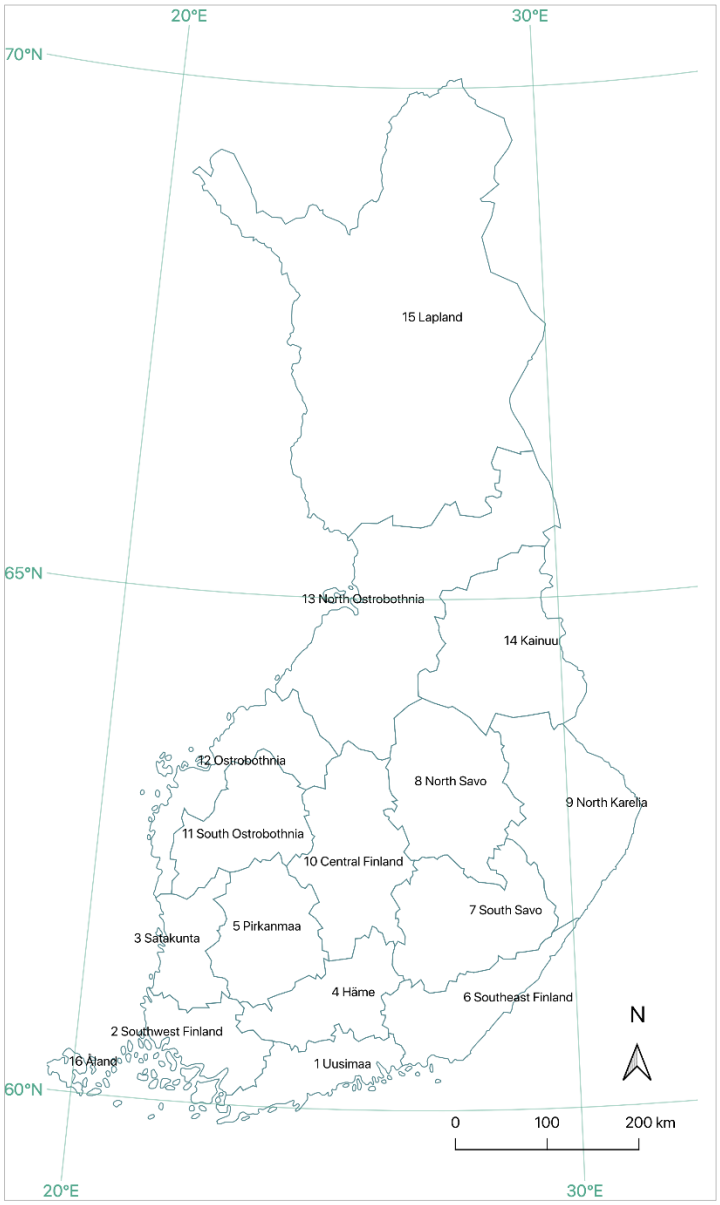

Figure 11. The regions as a reference to Table 4 and 6. Source: The Municipal Division dataset, National Land Survey of Finland 2020.

Table 6 shows how regional forecasts from Random Forest predictions in June 2020 deviate from the final crop survey-based regional forecasts for all cereal crops. The mean of error in kg/ha and percentage is region-wise and crop type-wise. The error is large in Lapland (no. 15 on the map) and Kainuu (no. 14 on the map). Firstly, it is not uncommon that crops are sown only in June in the northern parts of Finland. Secondly, we only have a few observed fields in Lapland. Whereas, in the southwestern Finland the regional forecasts seem to be quite accurate. This region is also considered to be the granary of Finland. Crops are sown there already at the end of April or early May. See Table 5 for the regional sowing dates of all crops.

With this in mind, we can conclude that country-wide forecasts seemed to work already in June, probably due to the inherited sampling weights from the crop production survey. However, for the regional forecasts the sampling data is inadequate. For regional forecasts, we should sample fields to have an equal spatial coverage. Moreover, for the northern regions the crop forecasting is reasonable only from July onwards due to later sowing dates.

Table 6. Deviation (error) from the final regional crop yield forecasts (kg/ha) when predicting with the Random Forest model in June 2020. Please refer to the region numbers on the map in Figure 11.

Region	Autumn wheat	Spring wheat	Rye	Feed barley	Malting barley	Oats	Mean
1	504	-563	358	-671	-634	-30	-173 (-6%)
2	32	-929	114	-629	-500	110	-300 (-10%)
3	-333	-905	7	-652	-840	-12	-456 (-14%)
4	166	-456	122	-391	-369	173	-126 (-4%)
5	-101	-814	315	-575	-402	-389	-328 (-11%)
6	-126	-660	7	-851	-866	-51	-425 (-13%)
7	-1,605	-443	397	-149	NA	-22	-364 (-10%)
8	-722	-449	207	-23	NA	-515	-300 (-11%)
9	NA	-907	55	-141	-2,489	-389	-774 (-36%)
10	-109	-566	-223	-428	-553	-208	-348 (-12%)
11	-484	-109	500	21	-44	80	-6 (0%)
12	121	-61	751	220	-237	223	170 (+4%)
13	-192	191	91	431	-1,864	-175	-253 (-13%)
14	NA	-1,243	-850	-716	NA	-1,267	-1,019 (-46%)
15	NA	NA	NA	-428	NA	-2,746	-1,587 (-133%)
Mean	-237 (-8%)	-565 (-19%)	132 (+1%)	-332 (-11%)	-800 (-32%)	-348 (-23%)	

4. Data processing pipeline

In this section we evaluate how much time, computing resources, and data storage is required for processing the data into an analysis ready format.

4.1. Single images

Creating the data set from single images starts with downloading the Sentinel-2 Level 2A images from the Copernicus Open Access Hub (Scihub). In the summer 2020, we downloaded images in two sets. The first images from 10.5.–20.6. were downloaded with a customized script on the 21st of June. There were 1496 zip-files occupying 1Tb of storage space. The downloading took 16 hours. Unzipping the zip files took another 2 hours. In the autumn, we downloaded the rest of the images from 21.6.–1.9.2020. We got 2,673 zip-files, which occupied 1.8Tb of storage space, and the download took 34 hours. Unzipping took 3 hours. The images are from 64 tiles covering the whole of Finnish arable land. After unzipping we had SAFE folders occupying 3.1Tb. Each SAFE folder includes many small files, mostly meta data files. Therefore, also the number of files is considerable, amounting to 600,000 in total.

After the image data was stored, we ran a Python program to extract all the pixels from our observational fields. The program is run for each image separately, but the tasks run in parallel as an array job. However, six crops are run separately, one after another, to maintain sanity. Although we downloaded the data in two sets, we processed all of them only in the autumn. Therefore, all the processing times here relate to the whole data set in 2020.

There were 41,570 images (JPEG2000) to process. Note that one satellite image consists of 13 bands, and each of these bands are stored in a separate JPEG2000-file. A field may be included in two or even three overlaying tiles, and therefore, some fields may have more observations. This processing step was the most time and computing expensive step. Table 7 shows the computing time needed, the number of images and observed fields per crop. For some crops, the arable area spans multiple tiles, therefore has more images. For example, oats and feed barley are cultivated up to the north of the country, but malting barley is cultivated only in the south.

Table 7. The computing time, the number of images and observed fields per crop when extracting pixel values from Sentinel-2 images.

Crop	Fields	Images	Computing time
Winter wheat	2,037	20,511	27min
Spring wheat	7,289	25,340	3h 28min
Rye	2,279	22,920	24min
Feed barley	11,762	29,630	7h 45min
Malting barley	2,301	16,670	1h 5min
Oats	13,677	27,200	6h 20min
Total:	39,345	142,271	19h 29min

4.2. Image mosaics

When following the research plan with image mosaics, we started with image files locally stored in the Puhti supercomputing server. Additionally, extraction of intensity values could be done in parallel for all sets. Table 8 shows the computing time needed, the number of images and observed fields per crop.

Table 8. The computing time, the number of images and observed fields per crop when extracting pixel values from Sentinel-2 image mosaics.

Crop	Fields	Images	Computing time
Winter wheat	998	48	parallel
Spring wheat	7,384	48	parallel
Rye	728	48	parallel
Feed barley	18,470	48	parallel
Malting barley	2,214	48	parallel
Oats	13,876	48	parallel
Total:	43,670	288	1h 23min

After extracting the pixel values, both the single image and image mosaic research line followed the same processing steps. Namely, from pixel values, histograms or percentiles were computed. After that the histograms were reshaped into an analysis-ready format. These steps take at maximum tens of minutes for each data set.

4.3. Processing time for forecasting

Here we estimate, how much time it takes to produce country-wide crop forecasts in June. First of all, with single images, it takes about 18 hours to download and unzip the data. This can be started already earlier in June and the latest images can be downloaded just at the start of the forecasting time. For extracting the pixel values from the images, we will need parcel geometries from the Finnish Food Authority, which are possibly available in mid-June. Extracting pixel values may take 10 hours. Downstream processing will take max. 2 hours. Producing forecasts from single images will thus take about three days. Producing forecasts from image mosaics will take a few hours or a maximum of one day.

To conclude, deploying the forecasting pipeline requires further automatization. Especially in the end of the pipeline, the validation of the results needs further scrutiny. Uploading the predictions to statistical production databases needs modifications to existing ICT-systems. In addition, the prediction model architectures need to be revised and improved along with the new data from the becoming years.

5. Practical implications to statistical production

The uptake of EO as a new data source in statistical production was more complex than initially expected. There are a myriad of approaches to monitoring crop yields, the main decisions to make being whether to use: i) optical or radar satellite data or both, ii) image mosaics or single images, iii) a pixel-based or object-based image analysis. In addition, remote sensing requires specialized expertise, not to mention the specialized expertise needed in data engineering and predictive modelling.

We acknowledged the lack of remote sensing expertise and made a decision at the start to outsource the pre-processing of the satellite images. With outsourcing the sustainability of the project may be jeopardized if the know-how outsourced cannot be fully transferred to the statistical production. In this sense, one significant achievement in the project has been the uptake of EO *knowledge,* with substantial contribution from the National Land Survey of Finland, as to a sound part of our production systems.

Although we aimed at fully automated processing pipeline to be deployed in statistical production, the project was initially explorative. In order to develop an existing crop yield statistics to be more detailed and timelier, we have explored new data sources and developed methodology for preparing first experimental statistics to evaluate its potential. We declared a need for improvement in the existing statistics, we had the innovative minds and courage to take the necessary steps to formulate a project plan and we had a vision how to fulfil the promise of economically sound improvements in the statistical production to the decision makers. We did not have a dream team of specialized expertise in remote sensing, IT systems, data engineering, data science and machine learning. However, during the project, we learned to master the core skills needed. At the same time, we have inherited the well-known downsides of explorative pilot projects: the skills and knowledge of a new system sits tight within the key workers of the project. We will need to work out how to transfer the knowledge within the production unit. At this stage the risk of departing experts would severely hamper efficient utilization of the achievements of the project in the future.

Another well-known obstacle with explorative projects is the bottleneck from developmental phase to the regular statistical production. How to facilitate the use of new data source to the domain statisticians remains a challenge and will require further reallocation of human resources and update of skills before the new data source becomes operationally mature part of the statistical production.

However, we are fully motivated to drive this change, since there is a lot at stake. In the context of agricultural statistics, more accurate in-season forecasts of crop yields benefit sustainable agriculture and food security with better informed political decisions. We expect that within few years our EO-based crop forecasting is proven a sound method to replace in-season regional expert estimates, and in the foreseen future also gradually replaces the annual survey. At the same time, we have in-house readiness to apply EO as a new data source also in other statistics themes.

As EO-based applications can be transnationally and even globally applicable, we should also foster international cooperation with international statistical organizations and national statistical institutes. We should also seek partners for research initiatives (inside or outside academia) for example on topics such as how reliable crop forecasts impacts the market, or what are the impacts of more dense-in-time crop forecasting for society at large.

6. Conclusions

Currently in Finland crop production statistics are based on both farm surveys and expert estimates combined with the data gathered from registers. The collection of the individual data is resource intensive. During the last decade, EO systems have been shown to provide an effective means to deliver large-scale crop monitoring and yield estimations. In this sense, this project has fulfilled its promise to establish a pilot case of an automated process for improved crop yield statistics by merging EO data and an advanced data-driven modelling approach. This report showed how, with a fully automated forecasting system, we can produce the first forecast in late June, around the same time as the JRC's European-wide forecast, with higher accuracy. The forecasts can be calculated, for example, at 10 day-intervals.

The machine learning models implemented in this project achieved a highly promising level of accuracy in pre-harvest yield predictions for four main cereals (oats, barley, wheat, and rye). This new method will eventually reduce data collection costs, but in this initial phase, the annual crop production survey is still a necessity for collecting more ground truth data for model development. At this stage, we have developed prediction models for the four main crops but there is still a lot of work needed to enlarge the method to cover all crops under statistical production. We foresee that with further development of the prediction models, the reliability of the predictions can be improved in the coming years and can be enlarged to include other crops.

On a large scale, more accurate in-season forecasting of crop yields will benefit sustainable agriculture and food security with better informed political decisions. In addition, reliable crop forecasts have a market impact. The societal impact of more crop forecasting needs to be evaluated.

References

Breiman, L. 2001. Random Forests. Machine Learning 45. p. 5–32.

Hochreiter, S. & Schmidhuber, J. 1997. Long Short-Term Memory. Neural Computation 9. p. 1735–1780.

LUKE, 2021. Statistics on Utilised Agricultural Area [Online]. Sowing dates available at: https://stat.luke.fi/sites/default/files/kevatkylvot_2000-2021_0.xls (Accessed: 21.9.2021).

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. 1986. Learning representations by back-propagating errors. Nature 323. p. 533–536.

Natural Resources Institute Finland
Latokartanonkaari 9
FI-00790 Helsinki, Finland
tel. +358 29 532 6000