Scottish Crop Map 2019

May 2021

An ‘Experimental Statistics’ publication from Rural and Environmental Sciences and Analytical Services (RESAS), a division of the Scottish Government.

RESAS

Rural & Environmental Science and Analytical Services

Scottish Government
Riaghaltas na h-Alba
gov.scot

Contents

Quick Guide

The Scottish Crop Map comes in 3 parts including this publication, as well as an interactive map with summary statistics, and the published open source code and associated data.

Publication

The Publication is this document and contains a summary of:

- the data sources used to build the map;
- information about how the crop map has been built;
- statistics and resulting analysis from the map;
- an assessment of the quality of the map;
- information on the data in the map and license conditions of its re-use; and
- future steps for the project.

Interactive Map and Statistics

The interactive map can be found by clicking the link below:

https://scotgov.maps.arcgis.com/apps/dashboards/f9216efc72e44b7e9093cfae08f6f861

The interactive map contains:

- field boundary layer categorised for the main crop types
- underlay map of Scotland
- contact information to provide the development team with corrections
- probability statistics for each field
- regional statistics of crop production
- comparison statistics with the 2019 Cereal Harvest – Official Statistics

Open Source Code

The open source 'Scottish-Crop-Map-2019-Publication' code repository is accessed via GitHub, providing code used to develop the statistics including:

- code for generating zonal statistics
- code for creating the random forest model
- Supporting documentation to support general re-use of the codes

There are also repositories for code used to develop previous iterations of the crop map however these repositories and supporting documentation have not been quality checked and are for general reference.

For further support on using any of these products email agric.stats@gov.scot.

Introduction

This report is part of a larger set of publications released as Experimental Statistics and called the 'Scottish Crop Map'. It uses novel modelling techniques to develop a map of agricultural fields in Scotland categorised by likely main crop types or grassland which were grown in 2019. The summary statistics provide production areas of crop values for the main types of crops as well as areas used for grassland.

To produce the map, spatial data collected from satellites and the real world are inputted into a machine learning model to provide probabilities of the most likely crop growing in every field in Scotland. This document explains how the model is developed, data is collected and the summary statistics produced from the model.

The Objectives of the Scottish Crop Map

Currently the official statistics publications for cereal production and harvest statistics relies on industry intelligence for initial estimates and a follow up large scale survey of farms with significant arable farming output to produce final estimates.

The main objective of the modelling used to create the Scottish Crop Map is to reduce the number of surveys and the associated survey burden for farmers. The modelling will allow earlier predictions of crop production values and also improve the initial estimates by reducing the reliance on industry intelligence.

The initial estimates are verified against the anticipated growing values recorded in the agricultural 'Single Application Form' (SAF) submitted by farmers to Scottish Government in their annual application for support payments. However both initial and final estimates are susceptible to human biases and often rely on best estimates rather than recorded data.

The development of the Scottish Crop Map can introduce a more robust method of collecting, verifying and producing estimates as well as a systematic way of collecting data.

Other objectives of the Scottish Crop Map include:

- Develop the data science techniques to use satellite images to produce other agricultural statistics; and
- Explore other uses of the techniques, data and model to develop new metrics on land-use and the positive and negative impacts of agricultural land-use.

Approach to Development

The Scottish Crop Map has been in development for a number of years. It began with a feasibility study which examined the possibility of creating a crop map using machine learning. The feasibility study found that:

- A crop map could be developed using radar images and could be augmented with visible images
- Using field boundaries was critical to the mapping product and that a hexagonal ‘mosaic’ style map would not create a high enough resolution
- The specifications of the computing requirements to develop a crop map required a higher specification than is available from desktop computing
- The methods required to develop the crop map requires a multi-skilled team including statisticians and Geographic UInformation Systems (GIS) specialists
- The coding required to develop the model was within the grasp of specialist statisticians skills

A further study was developed to look at the models and the data required to conduct the modelling. This study concluded:

- Either a ‘neural network’ or ‘random forest classification’ (RFC) were the suitable models for conducting the large scale modelling
- The first iteration of the map should concentrate on the RFC model.
- The model should concentrate on identifying the major crop types as there is not enough information to create a ‘training dataset’ for minor crops

The study also recommended a series of variables to be trialled in the model and the suitable timeframes to be included

The first model that was trialled was a collaboration with EDINA, the specialist data and analysis service at the University of Edinburgh. This collaboration included trialling a model which had been developed for the Department for Environment, Food, Rural and Agriculture (DEFRA) and used the ‘analysis ready dataset’ for Sentinel-1 data images for Scotland provided by the Joint National Committee for Conservation (JNCC).

This iteration of the model is regarded as the ‘Alpha’ Version. This means that the data and the model have been developed to a high enough standard for full deployment, but that the methods and results should be used with caution and future revisions will be implemented in future publications.

Satellite Images

The satellite images used in the Scottish Crop Map come from the Sentinel-1 earth observation satellite, which is part of the European Space Agency (ESA) Copernicus Programme. The data from these images was provided by JNCC.

More information can be found following the links below

ESA - ESA - Copernicus

JNCC - JNCC - Adviser to Government on Nature Conservation

Experimental Statistics

The statistics from the map are designated as ‘experimental’ because the methods used to assign the crop types are novel and are under review.

Experimental Statistics is a UK Statistics Authority classification for Official Statistics released by an Official Statistics producer, in this case the Scottish Government. This classification is used for statistics which are still in the testing phase and not yet fully developed. The reasons for publishing them ahead of a finalised publication are:

- Consultation – to get informed feedback from potential users
- Acclimatisation – this is an alternative version of the existing series of Cereal Harvest Production Statistics and is released in the current state to help users adapt to the method and presentation of the data
- Use – as an experimental series, these statistics can provide useful information for users, however caution should be exerted when re-using the statistics provided.

Consultation

The crop map statistics and associated publications are being produced partway through a well-defined development programme. Also, the statistics are new but still subject to testing in terms of their volatility and ability to meet customer needs. Before moving onto the next stages of development the development team are using this publication as an opportunity to align the current product with those needs.

The modelling used to produce these statistics has been developed in conjunction with a wide range of experts and has been developed to a standard where there is a fairly good degree of confidence in the accuracy of the results.

Before these statistics can be released as National or Official Statistics, the process which will follow the publication of the statistics and the consultation period following this publication will be used to help design future iterations and fully validate the measures to the standard expected of National Statistics.

Acclimatisation

Publishing the data in this form may also help users prepare for future publications for their own uses of the statistics and data.

Use

As a new measure of cereal harvest statistics, they may have a component that has immediate value to users, users should be aware of the statistics' theoretical quality and can use them before all operational testing has been conducted.

Users should be aware that the current format in which the data is presented may be changed in future iterations. Future releases may also have additional information on

agricultural crops in Scotland. Improvements to the model in future releases may include data on harvest yields, more timely data as the model is refined and improvements to the mapping and summary statistics that are presented.

National Statistics Designation

Statistics classified as “Experimental Statistics” are only made National Statistics following assessment by the Office for Statistics Regulation.

For this to happen, there are four stages that must be followed:

- Stage 1: self-assessment by business area
- Stage 2: methodological review by Strategy and Standards Directorate methodologists and/or business area methodologist
- Stage 3: recommendation by the Strategy and Standards Directorate, ONS Director General's office and the Statistical Policy Committee (SPC)
- Stage 4: assessment by the Office for Statistics Regulation

For Official Statistics, removing the Experimental Statistics label does not require Office for Statistics Regulation involvement.

Published documentation

This document explains the techniques and data used to develop these statistics and the methods that are being tested. This is a novel approach to developing statistics and instead of relying on traditional methods such as survey forms which are sent to farmers, these statistics rely on satellite images from the European Space Agency Copernicus Programme.

Statisticians, mapping specialists and data scientists have developed and tested a new methodology to interpret the vast data from the satellite images using high-powered computing referred to as machine learning.

The machine learning technique uses statistics and computer algorithms to analyse a dataset of radar backscatter time series for fields in Scotland, where it is known what crop was grown in each field. By understanding key characteristics of this satellite-derived dataset, a computer algorithm was written to then predict the main crop growing in all other agricultural fields (with unknown crop types). These predictions are then used to develop statistics to estimate the total growing area for all main crop types in Scotland.

All information used to develop this map and associated products have been made available online including:

- An interactive map of all crop types in Scotland
- The computing code used to build the crop map (access via GitHub)
- The satellite images used to develop the crop map.

Other data used in the development of the map including the field boundaries used are available on request and are released under licence from Ordnance Survey.

Data sources

A number of data sources are collected or are developed from primary sources of data to input into the crop map model. All of them have a spatial element to them, that is the data collected is known or tagged for a geographic point. All data collected and developed is available either freely to the public or, as is the case with Field Boundaries data, is available under license and on request.

Some summary information on the data used is included in Table 1 below.

Table 1: Data sources used in the Scottish Crop Map model.

Data	Purpose	Format	Time-frame	Primary source	Secondary source
Satellite	To produce zonal statistics from radar backscatter used in the machine learning model	tif	Mar-Oct 2019	JNCC	Ceda Archive
Ground Truth	Used to train and test the machine learning model	shp	2019	Rural Payments and Inspections Division (RPID)	N/A
Land parcels (field boundaries)	To assign areas within radar images to field areasto produce zonal statistics for individual fields.	shp	2019	Ordnance Survey	N/A
Spatial Data Masks	Exclusion of land parcels which intersect with the National Forest Inventory.	shp	2018	National Forest Inventory	N/A

For any further queries please contact **agric.stats@gov.scot**

Methodology

The methods presented here provide an overview of the steps taken to produce crop predictions in Scotland in 2019. The methods are broken down into pre-processing, processing, and modelling steps. To explore the methods further please visit the GitHub repository or email the team.

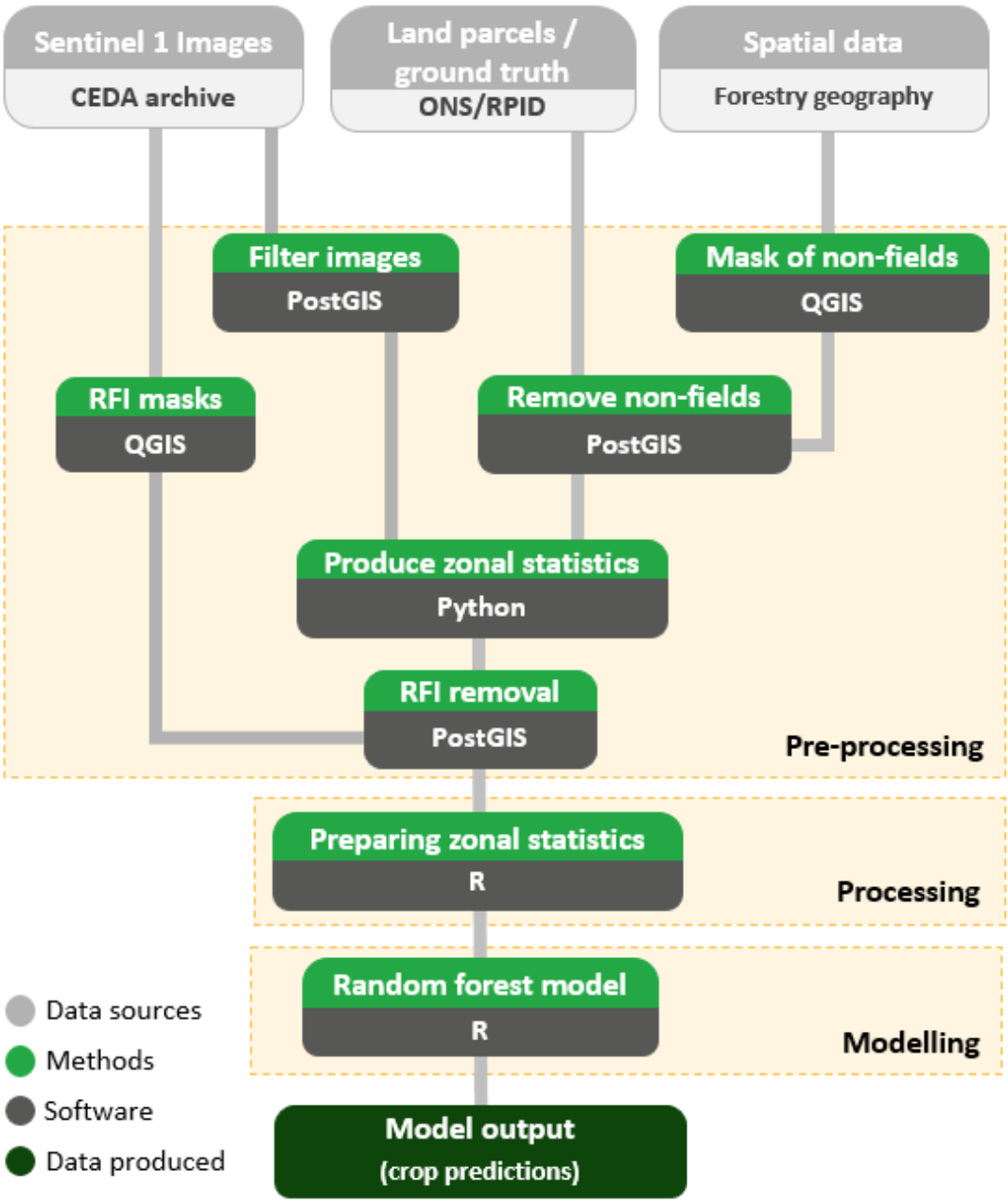
Product Overview of the Scottish Crop Map

The Crop Map of Scotland is a polygon vector dataset containing a subset of crop types in Scotland. The dataset contains 385,028 fields classifying Scotland into 4 main crop types, grassland, or unknown crop types (where no prediction was made).

Non-agricultural land is excluded from our analysis. Permanent grassland is included in the analysis as it contributes to a large proportion of agricultural fields. Crop types with a smaller amount of fields, including temporary grassland and crops grown in polytunnels, are excluded from the analysis due to not enough ground truth data being available to train the random forest model on.

A supervised classification algorithm (Random Forest Model) was used on data acquired from Sentinel-1 radar images during the period March to October 2019. The results were quality-assured against ground truth data (using Rural Payments and Inspections Division (RPID) inspection data) as well as summary data provided by the Scottish Agricultural Census Team and external stakeholders.

Figure 1: Overview of data processing and software used in the project



Pre-processing

Pre-processing steps were required to clean and combine data sets ready for analysis. Figure 1 shows the steps which were undertaken in this project. JNCC produced and supplied us with analysis-ready Sentinel-1 radar images. These were accessed through the CEDA archive in JNCC's virtual machine (JASMIN).

Mask of non-fields

Spatial data was collected and used to create a mask of non-field areas in QGIS software. A maximum field size threshold was also applied to remove very large

areas likely to be used for rough grazing and a spatial dataset was produced of the mask area. A minimum field size threshold (an area under 1,300m^2) was also applied to avoid the use of small areas which did not produce viable zonal statistics.

Remove non-fields

The mask identifying non-field areas, and extremely large or small fields, was then applied to the land parcels data set (in the software PostGIS) resulting in only land parcels we identified as fields being taken forward.

RFI masks

Areas suffering from Radio Frequency Interference (RFI) were manually identified in QGIS and masks were produced. A lookup containing field and date combinations for these RFIs was also created.

Filter Sentinel-1 images

Sentinel-1 images provided to us by JNCC were filtered to only those covering our particular selection of Scottish fields. This was done in PostGIS.

Produce zonal statistics

Python scripts were written (functioning within JNCC's JASMIN environment) to access these analysis-ready Sentinel-1 radar images within the CEDA archive (covering only Scotland). Once retrieved, radar images and known field boundaries (excluding extremely large and small fields) were used to summarise pixel values within each field into the mean values of the radar backscatter (for VH and VV polarisation).These values are referred to as the 'zonal statistics' for each field.

RFI removal

Fields containing RFIs were filtered out once the zonal statistics were produced using a lookup file of date-field combinations which had RFIs in them. Some areas were seriously affected by radar interference – this included the Dumfries area where we were unable to collect enough satellite data to use within the model, resulting in these fields being unclassified.

Processing

Preparing zonal statistics

Zonal statistics (which now excluded RFI areas) were averaged to form six-day blocks (rather than near-daily observations) to consolidate temporally close points into one variable. Other processing steps were performed to prepare the data for use within the random forest model. This was done in R.

Modelling

Random forest model

The zonal statistics were used to train a random forest model on only those fields with known crop types (ground truth data).

Once the zonal statistics were read into R, interpolation was carried out. This dealt with the remaining missing data so that more fields could be predicted by the model. Subsets were created to contain known fields (that have been assigned a crop type) and all (known and unknown) fields.

Some data preparation was conducted on the known fields dataset – any fields that still contained missing values (after averaging into six-day blocks and interpolating) alongside non-crops were removed. Crop types with a small amount of fields (< 25) were excluded from the known dataset and the model. This resulted in the model only including spring barley, winter barley, winter wheat, spring oats and (permanent) grassland.

Afterwards, the known dataset was split into training and test datasets using a 60/40 ratio. This was used to ensure there was enough data for smaller crop types in the test dataset. The training dataset was used to create a random forest model and model improvements were made using accuracies from the test dataset. These were: variable selection, selecting the number of trees (ntree), selecting the number of variables used at each split in a tree (mtry) and identifying the number of training fields to use for each crop (sampsize).

Once the model had been evaluated on the full dataset (containing all fields), a probability cut-off was selected based on the predicted probabilities. A probability cut-off of 0.48 was set for the model which allowed for the best balance between under- and overestimating crop area. For those fields that didn't meet the cut-off, it was assumed the model was less confident and thus were predicted to be NA (missing).

Code

The code used in (a) generating the zonal statistics and (b) running the random forest model are available on GitHub in the following repository:

https://github.com/cropmapteam/Scottish-Crop-Map-2019-Publication

Results and Key Findings

The results presented here are on areas of production. The results are provided as a guide to the breakdown of results which are produced from the Scottish Crop Map Model. The Official Statistics are available at Cereal and oilseed rape harvest: 2019 final estimates.

Scotland

In 2019 a total of 347,887 hectares (ha) were modelled and classified as areas growing main type crops in Scotland. In total, the model estimated the crop production in 385,028 fields in Scotland shown in Table 2.

The Scottish Crop Model identified and classified fields as growing either:

- Grassland;
- Spring barley;
- Spring oats;
- Winter barley; or
- Winter wheat.

Other crop types are not available from the current model and fields growing mixed crops are also not included, but may be classified as whole fields.

Table 2: Total fields classified by crop type

Crop Type	No. of Fields Predicted	Percentage of Total (%)
Grassland	303,026	78.7
Spring Barley	20,935	5.4
Spring Oats	169	> 0.0
Winter Barley	4607	1.2
Winter Wheat	8731	2.3
Unclassified	47560	12.4
Total	**385,028**	**100**

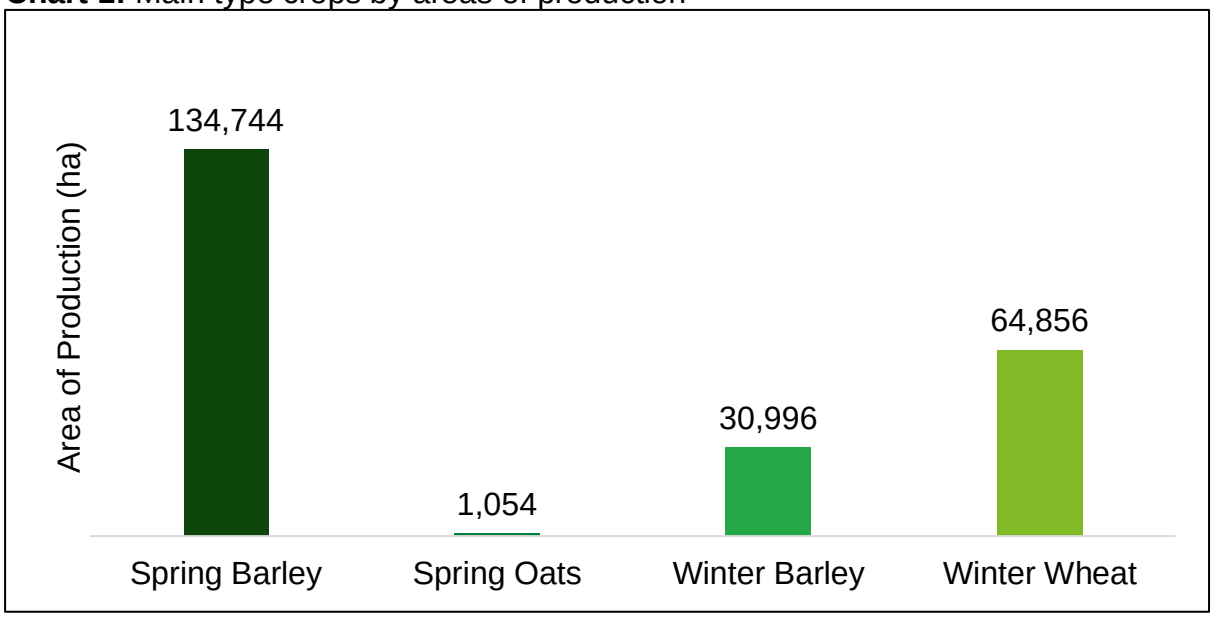
Source: Scottish Crop Map 2019, RESAS

The predominant cereal crop grown in Scotland is Spring Barley with an estimated growing area of 134,744 ha. Over half the area, at just over 58%, of land used for cereal production was used for Spring Barley which is shown in Chart 1.

The second most popular crop production, totalling 64,856 ha, is used for winter wheat in terms of area for 2019. At 28% this is just over a quarter of all land used for cereal production. Winter Barley production areas totalled 30,996 ha accounting for 13% and was the third most popular. Spring Oats production made up less than one per cent of all area with only 1,054 ha.

Of the total agricultural land available for cereal production, most land is used for grassland and pasture with over 1.8 million ha set aside for grassland.

Chart 1: Main type crops by areas of production



Source: Scottish Crop Map 2019, RESAS

Regional Breakdown

The largest areas for crop production are in the North East Scotland and Eastern Scotland regions. Both of these regions combined account for nearly two thirds of the total cropping area of Scotland and a total of 156,730 ha

Table 3: Main crop type by area and region not including grassland

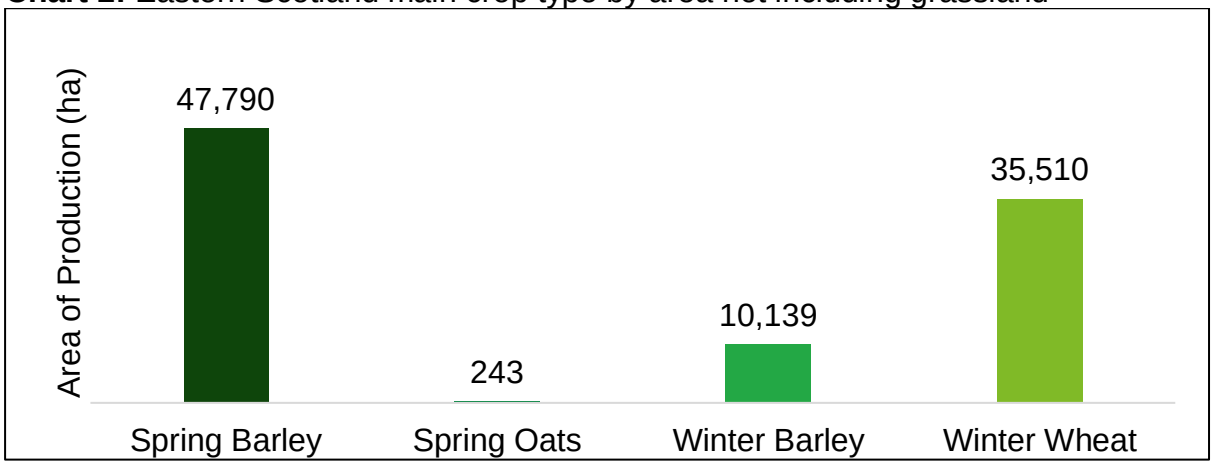
Region	Area (ha)	Percentage of Total (%)
Eastern Scotland	93,683	40
Highlands and Islands	25,847	11
North East Scotland	63,047	27
Southern Scotland	47,849	21
West Central Scotland	1,224	1
Total	**231,650**	**100**

Source: Scottish Crop Map 2019, RESAS

Eastern Scotland and Southern Scotland predominantly grow winter wheat as shown in Charts 2 and 5, with 35,510 ha and 21,861 ha respectively.

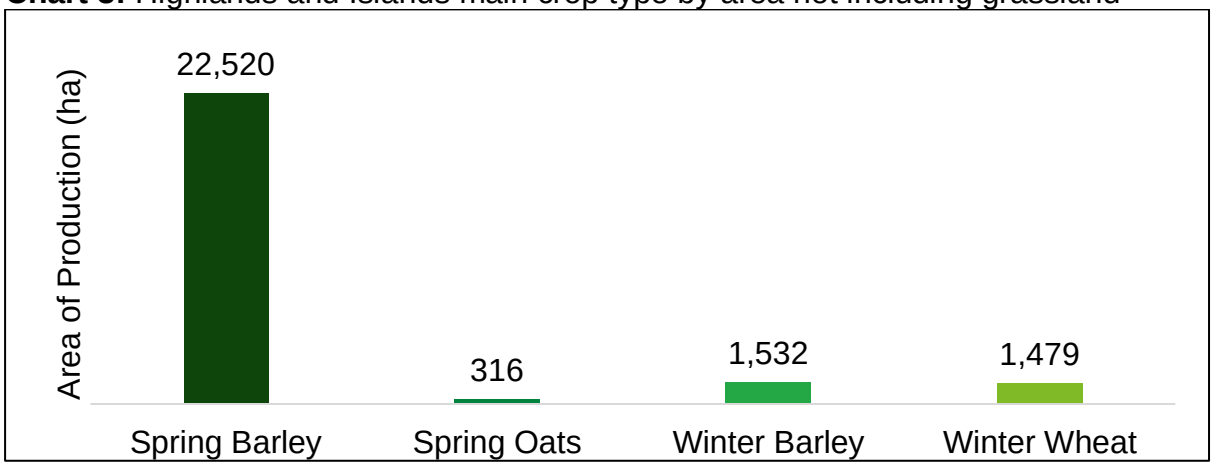
West Central Scotland has the lowest growing area of all five regions, however this region covers the Glasgow City area and conurbations.

Chart 2: Eastern Scotland main crop type by area not including grassland



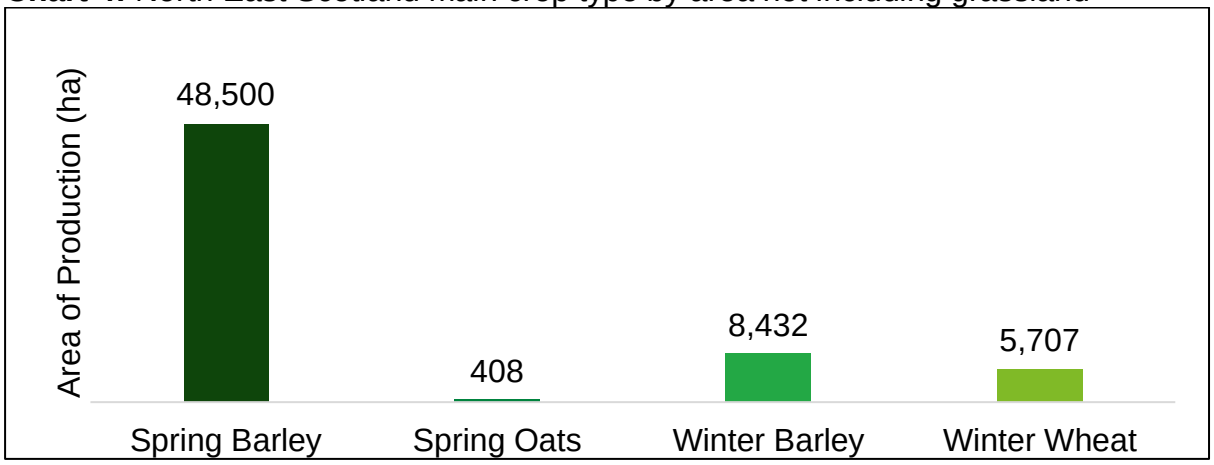
Source: Scottish Crop Map 2019, RESAS

Chart 3: Highlands and Islands main crop type by area not including grassland



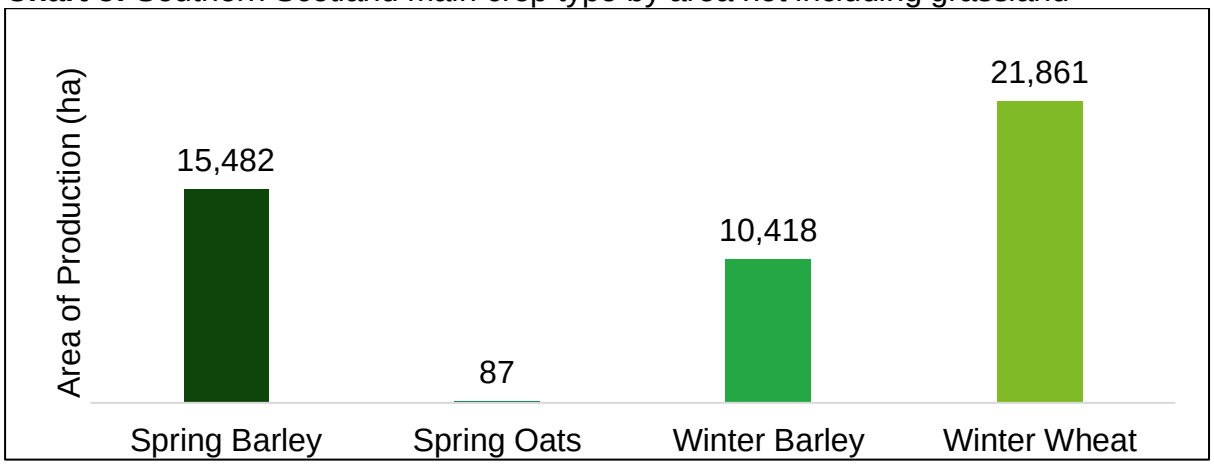
Source: Scottish Crop Map 2019, RESAS

Chart 4: North East Scotland main crop type by area not including grassland



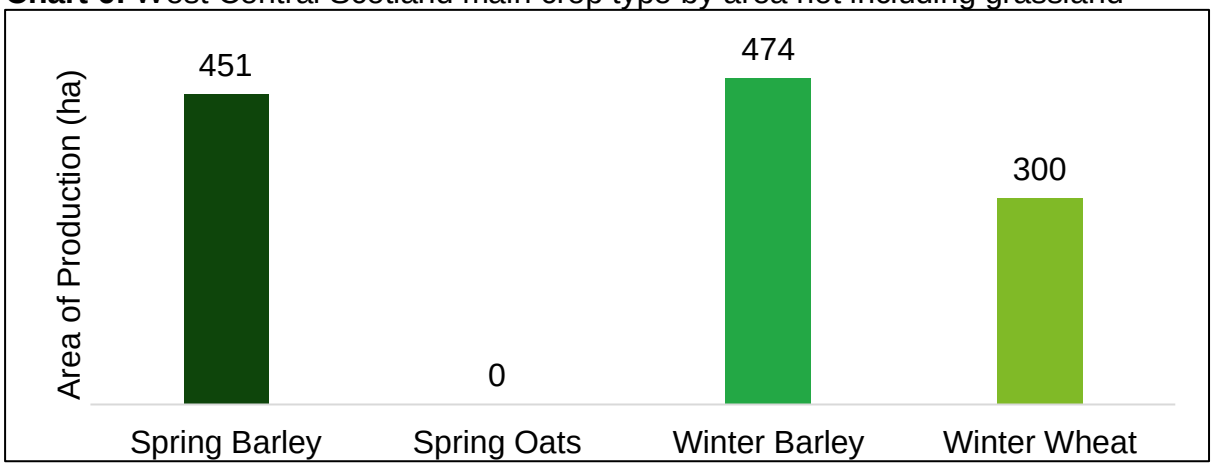
Source: Scottish Crop Map 2019, RESAS

Chart 5: Southern Scotland main crop type by area not including grassland



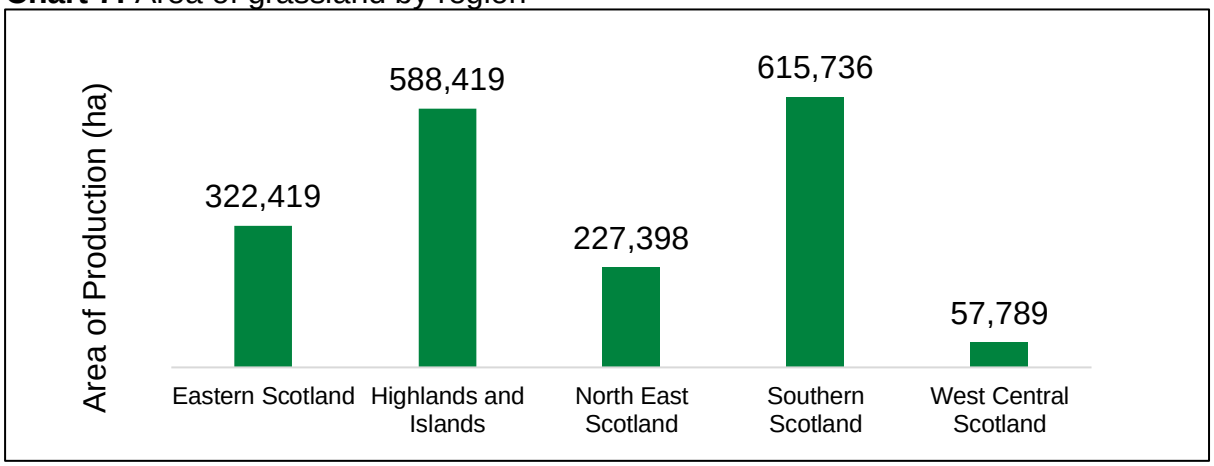
Source: Scottish Crop Map 2019, RESAS

Chart 6: West Central Scotland main crop type by area not including grassland



Source: Scottish Crop Map 2019, RESAS

Chart 7: Area of grassland by region



Source: Scottish Crop Map 2019, RESAS

Map 1: Scottish Crop Map 2019, RESAS

Grassland, featured as darker green on Map 1 (left), is a predominant feature of fields and the Scottish landscape in general. Most grassland is found in areas where there is a higher proportion of Less Favoured Area (LFA) land. LFA can be defined as the presence of poor land of poor productivity, which is difficult to cultivate and with a limited potential which cannot be increased except at excessive cost.

Therefore this land will be used to cut sillage and other forage for beef and dairy cattle and sheep.

Of the fields included in the Scottish Crop Map, 1.8 million hectares was used for grassland.

The highest areas used for grassland is the Highlands and Islands and Southern Scotland. In total 1.1 million hectares of fields were used for grassland, which is over two thirds. This is shown on the charts on the previous page.

Other grassland not designated within field boundaries in the crop map may also be used for keeping livestock outdoors, such as hill and other inaccessible areas.

Quality Assurance

The quality assurance steps taken here are to identify any weaknesses within our model and to provide greater understanding of how well different crops are being predicted. Within the QA exercise limitation of masking non-fields are outlined, accuracy of the predictions within the model were explored and comparisons were made to census data. Caution must be taken when comparing our data to census findings as the census figures will have associated caveats. The knowledge gained from this exercise will support further development of the crop map model.

Masking non-field areas

Areas of potential or young forestry may be included as fields if not accounted for within the spatial dataset at the time the mask for this project was created.

Model accuracies

The model produced an overall accuracy score of 98.86% with 95 % confidence intervals of 98.45% to 99.19% for the test data. The Kappa statistic (which compares the observed to the expected accuracy) was 0.89.

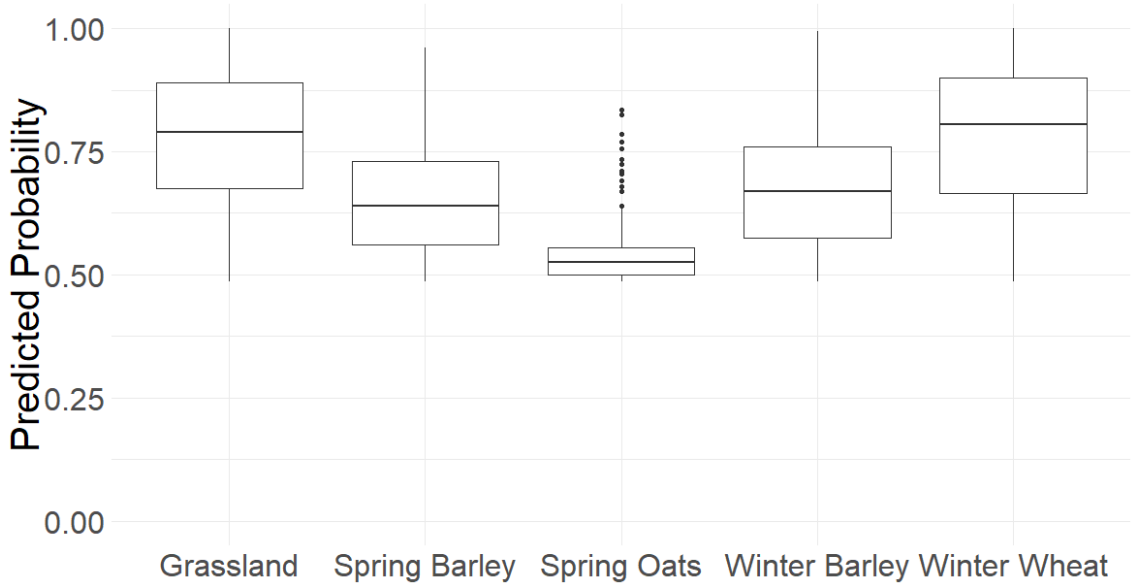
Table 4: Crops included in the model and their accuracy score

Crop	Producer accuracy
Spring Barley	94.12
Winter Barley	94.44
Winter Wheat	100.00
Spring Oats	41.18
Grassland	99.38

Source: Scottish Crop Map 2019, RESAS

Four out of five groups classified had accuracy scores of over 90%. Spring oats likely underperformed due to being under-represented in the ground-truth data and similarities in its appearance to other crops (Table 4).

Chart 8 : Boxplot of probability predictions for crops within the model



Source: Scottish Crop Map 2019, RESAS

As seen in Chart 8 there is variation between the crops in how well fields are being predicted. The probability cut off for the model was 0.48. Grassland, spring barley, winter barley and winter wheat have median probabilities above 0.625. Spring oats median fell close to the cut off, this suggests the model had difficulty predicting spring oats and had lower confidence in each field that was predicted as spring oat compared to the other groups.

Table 5: Confusion matrix (predictions made on the full dataset, unclassified predicted fields are excluded)

		Actual				
		Grassland	**Spring Barley**	**Spring Oats**	**Winter Barley**	**Winter Wheat**
Predicted	**Grassland**	7599	6	3	0	0
	Spring Barley	2	265	0	0	0
	Spring Oats	1	0	14	0	0
	Winter Barley	3	1	0	44	0
	Winter Wheat	3	0	1	0	67

Source: Scottish Crop Map 2019, RESAS

To test how well ground truth data lined up with model predictions a confusion matrix was used. Table 5 shows the confusion matrix for the dataset. The field predictions for the crops with ground truth data performed well for all groups. While spring oats has the highest proportion misidentified it also has the lowest number of fields.

Model predictions compared to known census areas

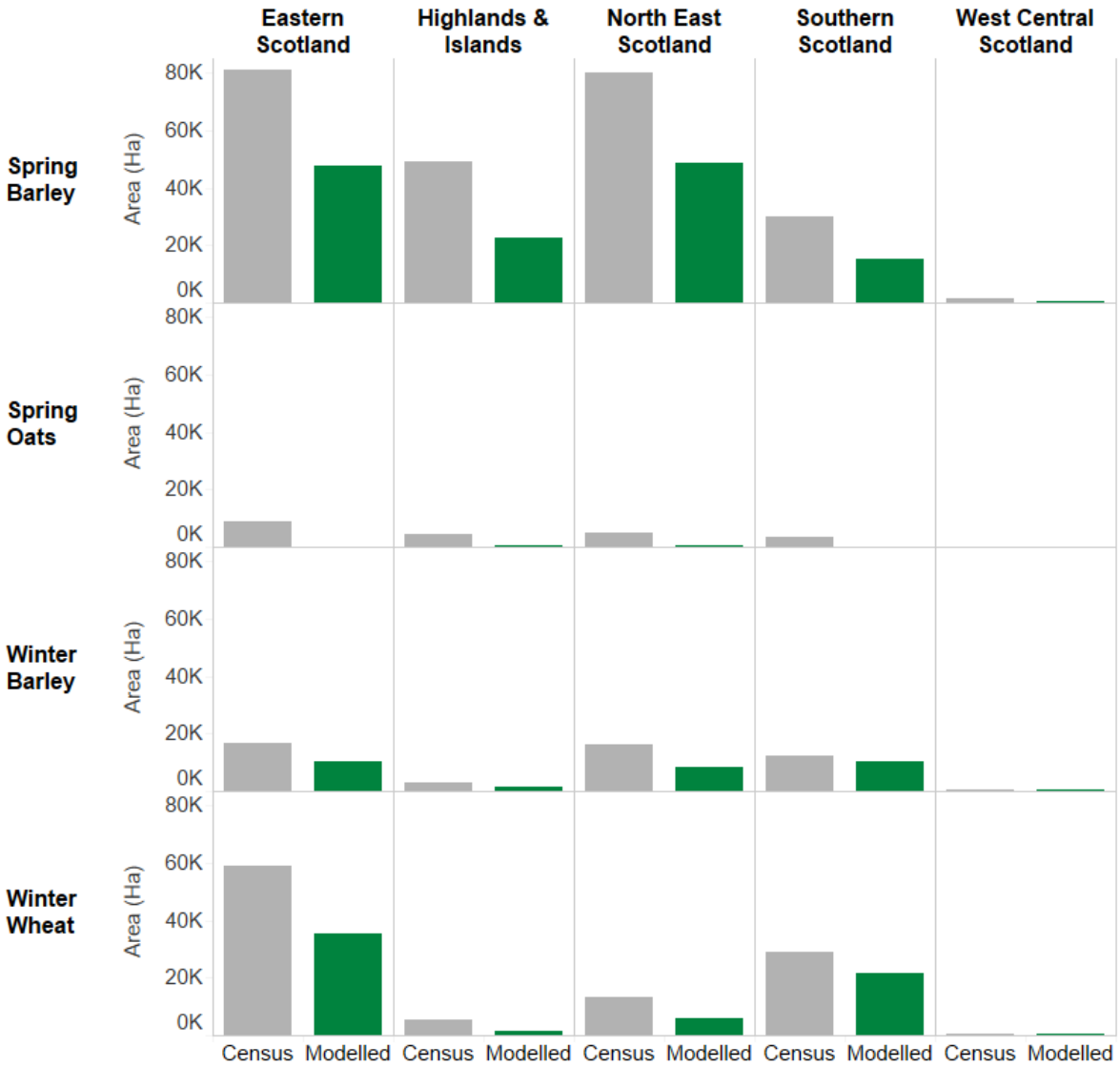
The model currently includes zonal statistics produced from radar images and the size of the fields as variables, even with such few variables some predicted crops areas are close to the census areas. Winter barley and wheat performed best in the model when comparing the total areas of each crops to the census areas, however spring oats again performed particularly badly (Table 6).

Table 6 Crops included in the model against their modelled area and area collected from the Scottish agricultural census

Crop	Modelled area (Ha)	Census area (Ha)
Spring Barley	134,744	242,090
Winter Barley	30,996	48,802
Winter Wheat	64,862	* 107,480
Spring Oats	1,054	* 21,905
Grassland	1,811,859	1,126,627

** Census areas for wheat and oats combines both spring and winter*
Source: Scottish Crop Map 2019, June Agricultural Census, RESAS

Chart 7: Modelled areas per crop type to census areas for NUT2 regions in Scotland



Source: Scottish Crop Map 2019, June Agricultural Census, RESAS

When comparing the predicted areas per crop to census areas we find that the model underestimates area for all crop types and regions, as seen in Chart 7. The closest predicted areas were for winter barley in Southern Scotland which were 15 per cent different to the census area. Winter Wheat in Southern Scotland had a predicted area which was 25 per cent different to the census area in that region. Spring barley consistently underestimated with differences from the census areas ranging from 40 to 63 per cent.

Next Steps

Improving the model

The published results from the Scottish Crop Map 2019 come from the first iteration of the machine learning model. Work will continue to improve the model in light of the limitations discussed in this paper.

The next step is to start including other data sources into the model to provide different types of signal for each crop. This should help to improve the predictions for some of the crops we have already included in the model but allow us to include smaller crop types which couldn't be included in this first iteration.

There are a number of data sources we will start to incorporate soon into the model. This includes a day of cloud-free Sentinel-2 satellite images for Scotland and weather variables (including hours of sunlight, average temperature and rainfall).

More timely crop predictions

Looking forward we aim to produce predictions in a more timely manner. To support us in this we have produced a guidance document and worked to streamline our methods. We have also carried out preliminary testing of the current model to identify whether all months of radar images are necessary. Early findings suggest the Autumn months are not required for our current predictions.

Estimating production values and yields

Estimating production values and yields is an important step in the development of this project, as these could become more timely and granular than what can be supplied through the current method. We are looking to collaborate with other organisations on this work and include data such as soil and weather variables.

Communicating with the team

Please contact us to query any of our work or to discuss collaboration projects. You can contact us at agric.stats@gov.scot.

Key	
CEDA	(The Natural Environment Research Council's Data Repository for Atmospheric Science and Earth Observation) Centre for Environmental Data Analysis
Grassland	Only permanent grassland is included in our model
Ground truth	Information (data) collected from the location and therefore are known.
JASMINE	JNCC's data analysis virtual machine available through the Simple ARD service.
JNCC	Joint Nature Conservation Committee
LPIS	Land Parcel Information Service, part of Scotland's rural payments and pervices
Mask	A dataset that defines which locations/ objects will be excluded from a geospatial file
Neural network	A machine learning algorithm that attempts to learn from data in a way that mimics the human brain
Polarisation	The orientation of the radar signal. Radars can transmit and receive horizontally polarized (H) or vertically polarized (V) signals
Polygon vector	A coordinate-based geographic data model that uses points, lines and shapes to represent features
Producer accuracy	The number of fields classified as a particular crop type by the model, as a proportion of the number of fields known to be that crop
Random forest	A machine learning algorithm based on using data to make repeated decisions that lead to an optimal classification

RFI	Radio frequency interference
SAF	Single Application Form, the form used to claim payments under a number of agricultural support schemes
Supervised learning	Any of a class of machine learning algorithms that use existing knowledge of categories or values to train a model to make new decisions using the same categories or values
Zonal statistics	Summary statistics about individual spatial areas

Scottish Government
Riaghaltas na h-Alba
gov.scot

This publication is available at **www.gov.scot**

Any enquiries regarding this publication should be sent to us at

The Scottish Government
St Andrew's House
Edinburgh
EH1 3DG

ISBN: 978-1-80201-000-8 (web only)

Published by The Scottish Government, May 2021

Produced for The Scottish Government by APS Group Scotland, 21 Tennant Street, Edinburgh EH6 5NA
PPDAS880027 (05/21)