Crop Map of England
Product Specification- v.2020.2

Non-Sensitive Information

Contents

Abbreviations

CROME – Crop Map of England

ESA – European Space Agency

GRD – Ground Range Detected

LUCODE – Land Use Code

RGBNIR – Red, Green, Blue, Near-Infrared

SNAP – Sentinel Application Platform

VH – Vertical-Horizontal Polarisation

VV – Vertical-Vertical Polarisation

VV/VH – Vertical-Vertical/Vertical-Horizontal dual Polarisation

1 Introduction

1.1 Document Control

1.1.1 Revision History

Date	Author	Version	Change reference
16/11/2017	Sanjay Rana	0.9	V0.8 revised to include new LUCODES and Labels, also information on 2017.2 ground truth.
12/12/2018	Sanjay Rana	0.10	V.09 revised to include confusion matrix for CROME 2018.2, revised methodology
18/10/2019	Russell Oliver	0.11	V.10 revised to include confusion matrix for CROME 2019.2, revised methodology
08/01/2021	Russel Oliver	0.12	V.11 revised to include confusion matrix for CROME 2020.2, revised methodology, added Random Forest probability information to attribute table

1.1.2 Reviewers and Contributors

Name	Role	Team
Russell Oliver	Lead Author	GI Tech Team
Valentin Louis	Co-Author	GI Tech Team
Sanjay Rana	Co-Author	GI Tech Team
Andrew McClune	Co-Author	GI Tech Team
Liann Rafter	Contributor	GI Tech Team
Sarah Murray	Reviewer	GI Tech Team
Rupert Waite	Reviewer	GI Tech Team
Alex Irving	Reviewer	GI Tech Team
Brian O'Toole	Reviewer	GI Tech Team

1.2 Purpose Scope

This document defines the product specification for the Crop Map of England (CROME) layer delivered by the Rural Payments Agency (RPA).

CROME is a tessellated multi-temporal visualisation of the type and distribution of land covers identified using remote sensing techniques. The purpose of this specification is to define that representation and the technical details for the reference dataset structure, data format and delivery. This document includes the technical specification and a general overview of the methodology used to produce the CROME layer.

The CROME layer covers most of the land in England (including some small isles) into approximately 32 million hexagon cells. The CROME Layer is expected to be revised once a year during September-October, and then released shortly afterwards.

It is important to note here that CROME is basically a scientific product and not a conventional cartographic product. Therefore, minimal efforts have been taken to:

- generalise the thematic variation of crop and land cover types;
- revise the land use types to homogenise land cover over an area;
- improve the cosmetic appearance of the layer.

1.3 Document Ownership and Distribution

This document is owned by the GI Technical Team within the RPA Operational Delivery Directorate. The document is for use by RPA staff, its third-party suppliers and end user organisations. There are no restrictions on internal Defra Organisation employee access to this document, or to contractors/consultants, third parties and any other agency or body with access to Defra Organisation assets or data handling facilities.

1.4 Assumptions, Dependencies and Constraints

The document describes the derivation of a technical geospatial data product. It is not written for the layman and thus assumes that users of this document are knowledgeable in:

- Geographic Information Systems
- Earth Observation concepts

The product is released under the Open Government Licence. This requires a reference as follows:

2 Data Creation

2.1 Feature Classification

The main sources for the crop classification are satellite images from the Sentinel constellation together with Ground Truth data for land cover types. A combination of radar and multispectral imagery from Sentinel sensors was used for automatic classification. Sentinel-1 with its C-band sensor transmits and receives microwave radiation in the horizontal (H) and vertical (V) polarisation. Sentinel-2 captures multispectral imagery in 13 spectral bands at varying spatial resolutions. More information regarding the specification of each satellite can be found on the following ESA websites: Sentinel-1 and Sentinel-2.

2.2 Geometry

Unlike conventional land cover classification maps, the classification of crops and land cover in CROME is represented as hexagonal cells. Traditional thematic visualisation of geospatial information (e.g. distribution of crop types and other land use types) partitions the land into either a regular grid (e.g. CropScape-Cropland Data layer) or an agricultural/administrative polygonal representation of the land use extent (e.g. CEH Land Cover plus –Crops 2015; onesoil.ai). However, the CROME layer is based on the hexagonal grid, which provides a superior opportunity to model and visualise the arbitrary arrangement of land use forms and locations in comparison to a regular grid. Furthermore, it avoids using third party vector products (with any associated licence constraints) that represent agricultural/administrative polygons.

Each hexagon cell covers an area of 4156 sq. m., or 0.41 hectares. The hexagon cells in the CROME layer are spatially distinct units and the CROME layer does not provide any spatial adjacency information. The vertices of adjacent cells are mostly coincident; therefore, the CROME layer provides a continuous representation of the land use. The hexagon cells are not constrained by any topographic features, except the extent of the land.

Figure 1. Example of CROME hexagonal classification cells

2.3 Attributes

In the creation of the CROME layer, attribution is generated from the associated imagery and the feature classification process.

3 Reference Layer Features

3.1 Representation

The classification is provided as hexagonal polygon cells, with each cell being attributed a land cover classification, according to the classification in Annex A. An example of the landscape and the classification cells, labelled with the land cover code in Annex A, is shown in Figure 1. Unlike previous versions, CROME 2018 (and onwards) data production did not involve any climate zones, as internal experiments revealed that Random Forest Classification performed with equivalent accuracy with or without regional labelling. The distribution has also been revised to be now based on ceremonial counties because user feedback have suggested that zones used in previous versions were too large.The ceremonial county boundaries are based on the boundary line product by the OS (link). CROME 2020 is distributed on 46 ceremonial counties. Table 1 lists the number of counties and their respective codes and Figure 2 shows the map of the counties.

Ceremonial County	County Code	Ceremonial County	County Code
Bedfordshire	BED	Leicestershire	LEI
Berkshire	BER	Lincolnshire	LIN
Bristol and Somerset	BRS	City and Greater London	LON
Buckinghamshire	BUC	Merseyside	MER
Cambridgeshire	CAM	Norfolk	NOR
Cheshire	CHE	Nottinghamshire	NOT
Cumbria	CMB	Northamptonshire	NRM
Cornwall	COR	Northumberland	NRT
Derbyshire	DER	North Yorkshire	NYO
Devon	DEV	Oxfordshire	OXF
Dorset	DOR	Rutland	RUT
Durham	DUR	Shropshire	SHR
East Riding of Yorkshire	ERY	Staffordshire	STF
Essex	ESS	Suffolk	SUF
East Sussex	ESX	Surrey	SUR
Gloucestershire	GLO	South Yorkshire	SYO
Greater Manchester	GMN	Tyne & Wear	TAW
Hampshire	HAM	Warwickshire	WAR
Herefordshire	HER	Wiltshire	WIL
Hertfordshire	HRT	West Midlands	WMD
Isle of Wight	IOW	Worcestershire	WOR
Kent	KEN	West Sussex	WSX
Lancashire	LAN	West Yorkshire	WYR

Table 1. CROME data distribution zones with their zone codes and cell counts respectively.

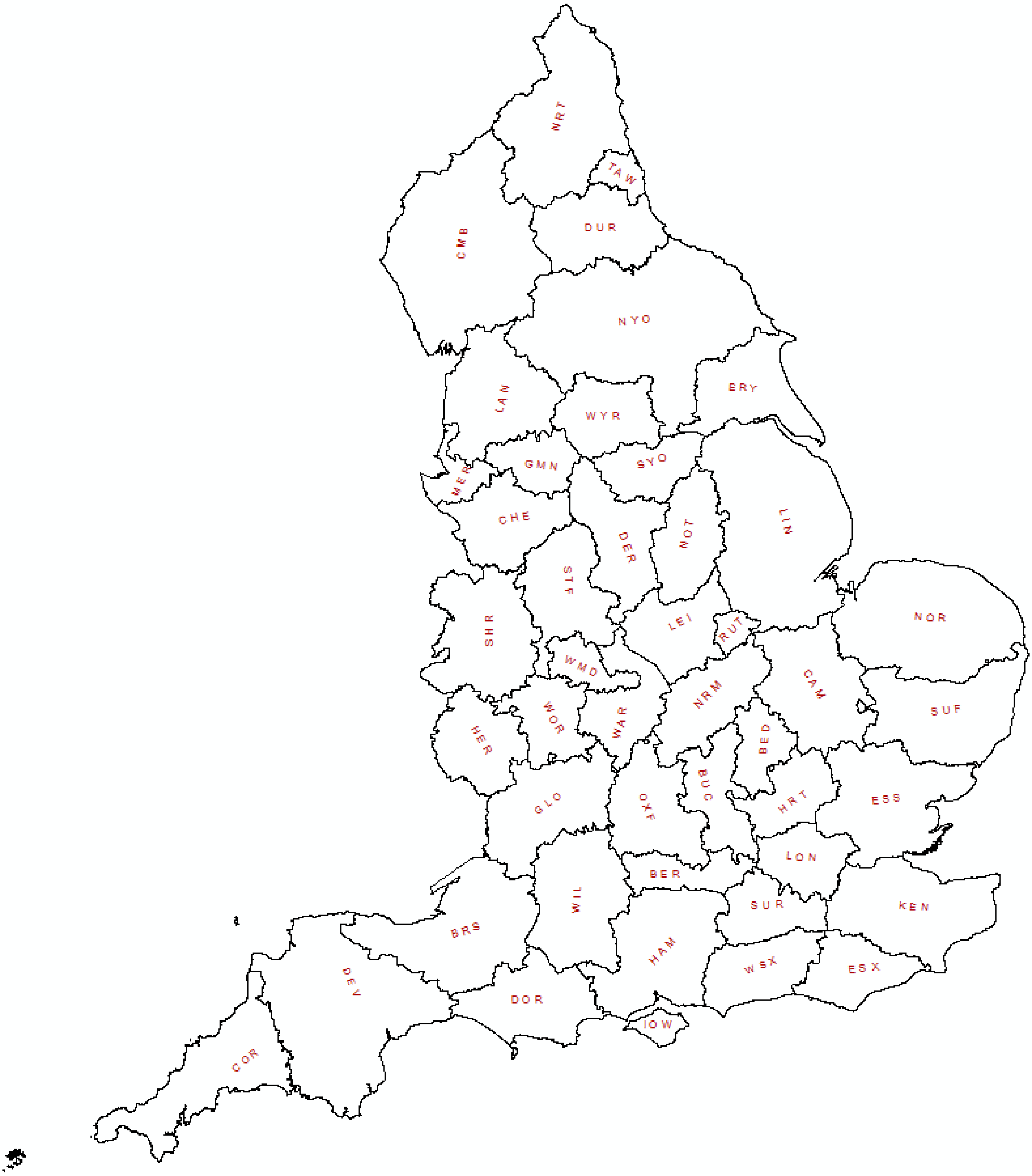

Figure 2. Ordnance Survey Ceremonial Counties in England (source: Ordnance Survey). See Table 1 for full names for each county codes.

4 CROME Classification

The classification of the topography was performed using the Random Forest classifier, a supervised machine learning approach. More information on Random Forest Classification can be found on the authors' homepage. The approach learns the classification of specific crop types by associating backscatter characteristics of each radar polarisation (VV, VH, VV/VH) and multispectral (RGBNIR) signatures at 10m spatial resolution to known land covers, previously collected by field inspectors. The ground truth points, radar and multispectral images were fed into the workflow, outlined in Figure 3, making up the four main work packages described below in further detail.

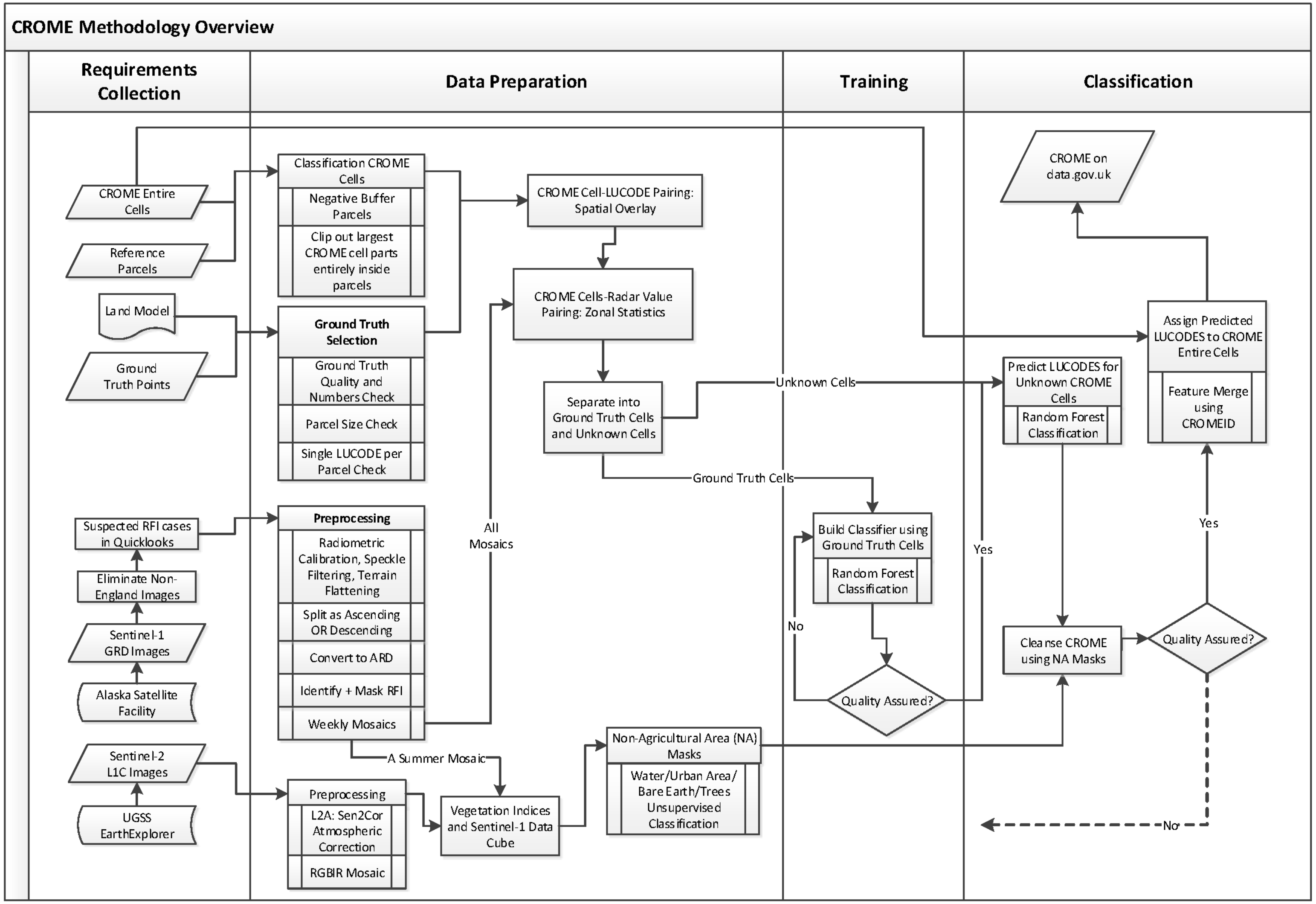

Figure 3. Workflow for CROME classification

4.1 Requirements Collection

This step involves the creation and collection of the input datasets and specifications required to perform the classification. The step involves five main inputs:

a) CROME Cells: These are hexagon cells that form the spatial framework of the CROME layer. These are essentially Voronoi polygons and were created using a combination of ESRI ArcMap and FME applications.
b) Land Model (current version 1.7.1): This is an internal RPA document that contains a list and specification of the relevant land cover and crop types which are eventually assigned to the CROME cells.
c) Ground Truth Points (GTP): These are a record of the spatial location and type of crop and land use found by the RPA field inspectors and Cyient during late-spring and summer period of 2020. The ground truth collection process involves verification and, where required, correction of the crop and other land cover types declared by the farmers against each parcel registered under their name. In order to minimise classification errors, only the ground truth data and parcels that matched the following conditions were used:
    a. Parcels only contained a single GTP crop/land use type.
    b. Parcel area after applying a negative buffer of 20 m (to exclude contamination from hedges and other boundary features) was greater than 0.32 ha (i.e. it would cover 80% of a CROME cell area).
    c. The crop/land use type was present in at least 10 ground truth parcels.

    In addition, the numbers and varieties for non-agricultural areas were boosted by collecting additional ground truth information, such as man-made surfaces and structures, from other RPA datasets e.g. inspections.
d) Sentinel-1 GRD images: These are Sentinel-1 Radar Ground Range Detected (GRD) dual polarisation (VV, VH) images for the period of January till September 2020, covering the main land masses of England. These can be downloaded from the Copernicus Open Access Hub or the Alaska Satellite Data Facility.
e) Sentinel-2 Level2A images: These are a collection of Sentinel-2 RGBNIR images, mosaiced to a cloud-free product for the period of March till June 2020. These can be downloaded from the Copernicus Open Access Hub.

4.2 Data Preparation

The objective of this step is to attribute the CROME Cells with the required LUCODE (only for cells used during supervised training) and radar backscatter and multispectral surface reflectance values for the purposes of classification. It involves the following main processes:

a) Ground truth data points are filtered by applying a combination of simple attribute and spatial filtering using farm field polygons.
b) The Sentinel-1 GRD images are processed using the Sentinel Application Platform (more information available on SNAP homepage) software to transform the raw images into georeferenced and radiometrically corrected Sigma-0 (backscattering coefficient) images.
c) The Sentinel-2 L2A images are downloaded and mosaiced to cloud free composite.
d) Zonal Statistics analysis is performed to calculate the weekly mean of VV, VH and VV/VH-ratio image values for the area covered by each CROME cell. Each CROME cell (32 million in total) is assigned a pair of triplet values i.e. mean VV, VH and VV/VH, based on an England-wide weekly composite of scenes. Statistics were calculated of each polarization combinations

(VV, VH and VV/VH) from suitable weekly mosaic radar scenes captured from January to September 2020. Weekly statistics allowed a common temporal reference axis for all cells because although satellite pass dates over cells are broadly predictable as per ESA guidance, i.e. within 5-10 days revisit, these still varied considerably across England thereby creating data gaps.

e) Zonal statistics (mean) for Sentinel-2 L2A RGBNIR 10m are extracted only from one cloud-free mosaic due to persistent cloud cover.

f) Spatial overlap is tested between the selected ground truth subset points and all the CROME cell polygons to assign the known LUCODES to the training-CROME cell polygons.

Classification rules were formulated based on the known LUCODES, established in step (f), and the weekly radar backscattering coefficient values derived in step (d) and the surface reflectance values derived in step (e) from the training CROME cells.

4.3 Training

The objective of this step is to automatically build the classification rules that can be used to predict the LUCODE of a CROME cell, given its associated weekly radar backscatter coefficient values and the multispectral values.

This work uses the Random Forest classification technique, a supervised machine learning approach, available in the R application. The unique advantage of Random Forest classification is that given a large number of ground truth points, it is able to perform robust cross-validation internally by generating numerous decision trees using a subsample of the training data for each tree using a technique called bagging (bootstrap aggregation). Furthermore, the Random Forest classifier estimates the variable importance and calculates the probability of each feature (e.g. hexagon cell) belonging to a specific class / LUCODE by taking advantage of the majority voting strategy.

The training process involves randomly combining input variables and deriving a permutation of input variables that provides the most instances of accurate match between the known LUCODE and a predicted LUCODE. The software internally uses 2/3rds of the input variables for developing and training the ruleset, and the remaining 1/3rd for testing the classification, using the out-of-bag (OOB) error.

4.4 Classification

Each CROME cell is then applied a LUCODE from the Random Forest classification ruleset. A random visual check using 2020 Sentinel-2 images was performed to detect obvious misclassifications. Simple cases of misclassifications (e.g. slightly rough areas of manmade surfaces conflicted with grass, water, urban areas misclassified as trees, and vice versa) are reduced by using the 10m RGBNIR Sentinel-2 mosaic.

5 Features Types and Attribution

5.1 Attribution

Table 2 shows the attribute for each CROME cell:

Name	Type	Properties	Comments
CROMEID	TEXT	Primary Key	Unique across all supply years
REFDATE	NUMBER	Not Null	The date of the classification was performed.
LUCODE	TEXT	Not Null	The land use code. See Annex A for lookup tables.
SHAPE	GEOMETRY	Not Null	Polygon representing the extent of the classified land use object.
SHAPE_area	FLOAT	Not Null	Area of the hexagon cell in sq. m
SHAPE_length	FLOAT	Not Null	Circumference of the hexagon cell in m
PROB	FLOAT	Not Null	Random Forest derived probability of a hexagon cell belonging to a specific class (majority voting) with a range from 0 to 1

Table 2. Attribute Schema of CROME layer

5.1.1 CROMEID

The CROMEID is the unique identifier assigned to each cell, consisting of the letters ‘RPA’ and a twelve-digit number, which are the Easting and Northing coordinates of the centroid of the cell.

5.1.2 REFDATE

The REFDATE is the date when the classification was performed. The date format used is YYYYMMDD e.g. 20181121. The value is expected to remain constant for each cell in a CROME release.

5.1.3 LUCODE

The Land Use Code is alphanumeric code of up to 5 characters long and is intended to record land use information. At present, all non-agricultural areas will be assigned a generic non-agricultural land use code. These land use codes are based on the land model used by the RPA as part of the claim purpose. The table in Annex A contains the valid Land Use Code (Column 3) that will be used in the current release of the CROME.

5.1.4 PROB

Random Forest derived probability of a hexagon cell belonging to a specific class. This probability is derived from the majority voting process used when using the Random Forest classifier i.e. the actual numbers of trees that resulted in the final class/LUCODE compared to the total number of trees.

5.2 Data Format and Naming

The dataset is supplied in the ESRI shapefile format, together with a projection file (*.prj) that defines the coordinate system as British National Grid. The data for a given year shall be supplied either as a single shapefile or as a series of shapefiles that collectively cover all CROME zones.

The dataset format shall be consistent with the specification in Section 5.1, with all fields whose source is marked as shapefile being included in the supplied dataset.

The shapefile name shall be as follows:

CROME_<YYYYMMDD>_<ZONECODE>.shp

where:

<YYYYMMDD> is the release date of the dataset

<ZONECODE> is the mnemonic for each ceremonial county data distribution zone covered by the shapefile (see Table 1).

Each shapefile will be accompanied by a metadata xml file in the accordance to the UK GEMINI v 2.2 standard.

Please note that the final packaged filenames (see Table 1) may be different due to other requirements.

6 Thematic Accuracy

Thematic accuracy is concerned with the degree to which the classification and attribution of features correctly match the real-world features that are being modelled within the dataset.
The quality of CROME land code classification was assessed by comparing the crop/land use types predicted by the Random Forest Classification against the ground truth data collated by the field inspectors from Rural Payments Agency. The comparison is reported in the form of a confusion matrix, with the overall accuracy and Kappa Coefficient presented in Table 3 below. The confusion matrix was measured on the majority of CROME cells within each RPA land parcel. CROME v2020.2 has an overall accuracy of 70.5% and a Kappa coefficient of 0.66.

The Ground Truth data did not provide information on non-agricultural land covers such as woodland, water bodies and general non-agricultural areas (e.g. manmade surfaces) therefore the confusion matrix does not provide accuracy numbers for each land use. Samples for these land uses were collected from other reliable sources e.g. non-Remote Sensing inspections and were only used during the training to minimise confusion with similar classes.

6.1 Known Accuracy Challenges

Due to strong correspondence in the physical morphology of Permanent Crops ground truths and other trees, TC01 cells also cover areas of other trees e.g. along parcel boundaries, roads. Due to lack of any definitive morphological definition for fallow land ground truth in general, FA01 cells also cover areas that are mostly bare soil to partially grass. Similarly, due to variability in growth of grass, PG01 cells cover areas that are declared as fallow land by the customers. The seasonality of the suitable Sentinel-2 image used for cleansing can also influence the attribution of otherwise vegetated parcels to NA01 if the usable image was taken when the field had been harvested or still bare soil.

7 Dataset Specification

7.1 Format

The CROME datasets are in ESRI shapefile format.

7.2 Topological Consistency

No topological consistency errors are known to exist with CROME, i.e. each feature is valid according to the OGC specification.

7.3 Horizontal Positional Accuracy

Not applicable

7.4 Currency

The map is to be updated annually, after August to coincide approximately with end of harvesting and cropping season.

7.5 Format Consistency

N/A

7.6 Domain Consistency

N/A

7.7 Temporal Consistency

There are no known temporal consistency issues with CROME.

7.8 Temporal Validity

There are no known temporal validity issues with CROME.

7.9 Attribute Completeness

The version of the dataset has 100% attribute completeness.

7.10Spatial Completeness

The version of the dataset has 100% spatial completeness, as verified by visual inspection.

7.11Lineage

Crop types were derived by using Random Forest Classification on Sentinel-1 radar data and refined using surface reflectance images from Sentinel-2 optical data.

7.12 CROME Layer Metadata

A metadata file will be supplied with each product supply. Metadata will conform to ISO 19115 and be UK GEMINI discovery level metadata.

		Ground Truth																			
		AC01	**AC03**	**AC16**	**AC17**	**AC19**	**AC20**	**AC32**	**AC44**	**AC63**	**AC65**	**AC66**	**AC67**	**FA01**	**LG03**	**LG07**	**LG11**	**LG14**	**LG20**	**PG01**	**Total**
CROME	**AC01**	**227**	1		15			6	1	2		9		2	4	4				11	**282**
	AC03	3	**17**		2				1			2									**25**
	AC16			**8**								2									**10**
	AC17	4			**41**							2		3	4	1		2	3	14	**74**
	AC19	1		1		**24**					1	2				1					**30**
	AC20						**2**														**2**
	AC32	3			3	3		**36**				21		1	1					5	**73**
	AC44	4				1			**30**	1		1		1	1	1					**40**
	AC63	1	2		3			1	2	**76**		5	1	1						14	**106**
	AC65					2					**4**	1	1			1				2	**11**
	AC66	18			10			3	1	8	2	**184**	3	3		2		3	1	19	**257**
	AC67	9			4			1		2	3	6	**102**		2					1	**130**
	FA01				4									**4**	2			1		6	**17**
	LG03	1			3			2				2		5	**60**	1			7	3	**84**
	LG07				2									1		**22**				1	**26**
	LG11											1		1			**7**			8	**17**
	LG14										1							**1**		2	**4**
	LG20	1			5	1				1		4	1		2				**13**		**28**
	PG01				1									6			1		1	**19**	**28**
	Total	**272**	**20**	**9**	**93**	**31**	**2**	**49**	**35**	**90**	**11**	**242**	**108**	**28**	**76**	**33**	**8**	**7**	**25**	**105**	**1244**

Table 3. Confusion Matrix of CROME v 2019.2

Land Cover	LUCODE	Accuracy	
		User	Producer
Barley - Spring	AC01	80%	83%
Beet	AC03	68%	85%
Linseed -Spring	AC16	80%	89%
Maize	AC17	55%	44%
Oats - Spring	AC19	80%	77%
Onions	AC20	100%	100%
Wheat - Spring	AC32	49%	73%
Potato	AC44	75%	86%
Barley - Winter	AC63	72%	84%
Oats - Winter	AC65	36%	36%
Wheat - Winter	AC66	72%	76%
Oilseed - Winter	AC67	78%	94%
Fallow	FA01	24%	14%
Field beans - Spring	LG03	71%	79%
Peas - Spring	LG07	85%	67%
Lucerne	LG11	41%	88%
Clover	LG14	25%	14%
Field beans - Winter	LG20	46%	52%
Permanent grassland	PG01	68%	18%

Table 4. User and Producer accuracies of land cover.

Annex A – Land Cover Codes

Land Cover Description	LUCODE	Land Use Description
Cereal Crops	AC01	Spring Barley
	AC03	Beet
	AC04	Borage
	AC05	Buckwheat
	AC06	Canary Seed
	AC07	Carrot
	AC09	Chicory
	AC10	Daffodil
	AC14	Hemp
	AC15	Lettuce
	AC16	Spring Linseed
	AC17	Maize
	AC18	Millet
	AC19	Spring Oats
	AC20	Onions
	AC22	Parsley
	AC23	Parsnips
	AC24	Spring Rye
	AC26	Spinach
	AC27	Strawberry
	AC30	Spring Triticale
	AC32	Spring Wheat
	AC34	Spring Cabbage
	AC35	Turnip
	AC36	Spring Oilseed
	AC37	Brown Mustard
	AC38	Mustard
	AC41	Radish
	AC44	Potato
	AC45	Tomato
	AC50	Squash
	AC52	Siam Pumpkin
	AC58	Mixed Crop-Group 1
	AC59	Mixed Crop-Group 2
	AC60	Mixed Crop-Group 3
	AC61	Mixed Crop-Group 4
	AC62	Mixed Crop-Group 5
	AC63	Winter Barley
	AC64	Winter Linseed
	AC65	Winter Oats
	AC66	Winter Wheat
	AC67	Winter Oilseed
	AC68	Winter Rye
	AC69	Winter Triticale
	AC70	Winter Cabbage
	AC71	Coriander
	AC72	Corn gromwell
	AC74	Phacelia
	AC81	Poppy
	AC88	Sunflower
	AC90	Gladioli
	AC92	Sorghum
	AC94	Sweet William
	AC100	Italian Ryegrass

	CA02	Cover Crop
Leguminous Crops	LG01	Chickpea
	LG02	Fenugreek
	LG03	Spring Field beans
	LG04	Green Beans
	LG06	Lupins
	LG07	Spring Peas
	LG09	Cowpea
	LG08	Soya
	LG11	Lucerne
	LG13	Sainfoin
	LG14	Clover
	LG15	Mixed Crops–Group 1 Leguminous
	LG16	Mixed Crops–Group 2 Leguminous
	LG20	Winter Field beans
	LG21	Winter Peas
Energy Crop	SR01	Short Rotation Coppice
Grassland	FA01	Fallow Land
	HE02	Heathland and Bracken
	PG01	Grass
Non–Agricultural Land	NA01	Non-vegetated or sparsely vegetated Land
Water	WA00	Water
Trees	TC01	Perennial Crops and Isolated Trees
	NU01	Nursery Crops
	WO12	Trees and Scrubs, short Woody plants, hedgerows
Unknown Vegetation Or Mixed Vegetation	AC00	Unknown or Mixed Vegetation

Note that the variety of land cover codes may change during releases of CROME due to variability in ground truth however the LUCODE will remain same.

Acknowledgements

We would like to acknowledge support and guidance from the following sources and individuals:

- ESA's Scihub and Alaska Satellite Facility for download access to the Sentinel-1 Data;
- Guido Lemoine (Joint Research Centre, Italy) and Andrea Minchella (previously at Satellite Applications Catapult, UK) for pointers on how to batch process large number of Sentinel-1 images.
- SNAP Discussion forum members on very helpful tips on improving the performance of batch processing and issues surround Sentinel-1 radar data.
- Asger Petersen and Gregers Petersen (Septima.dk, Denmark) for advice on how to use GDAL and parallel processing of large amount of geospatial data.