Rural Payments Agency one agency

Crop Map of England

Product Specification- v.2017.2

Non-Sensitive Information

Contents

Abbreviations

CROME – Crop Map of England

ESA – European Space Agency

GRD – Ground Range Detected

LUCODE – Land Use Code

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.

1.1.2 Reviewers and Contributors

Name	Role	Team
Sanjay Rana	Lead Author	GI Tech Team
Andrew McClune	Co-Author	GI Tech Team
Liann Rafter	Contributor	GI Tech Team
Matauka Yeta	Reviewer	Land Technical Data Steward

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 plus 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 as soon as after.

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 to revise the land use types to homogenise land cover over an area and to 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 Design Directorate. The document is for use by RPA staff and third party suppliers. 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

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 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 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 ESA website.

2.2 Geometry

Unlike conventional land cover classification maps, the classification of crops and land cover in CROME is represented as hexagon 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). 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 is better than the agricultural/administrative polygonal representation because it doesn't rely on any third party vector product, which are often needed to create a good quality polygonal representation.

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 doesn't 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.

It was realised early on that it would be computationally non-trivial and theoretically problematic to classify the crop types for the entire country in a single process. The reasoning behind the latter logic was that horticulturally, crops and cropping patterns in England vary dependent upon the climatic zones in England (see Figure 2).

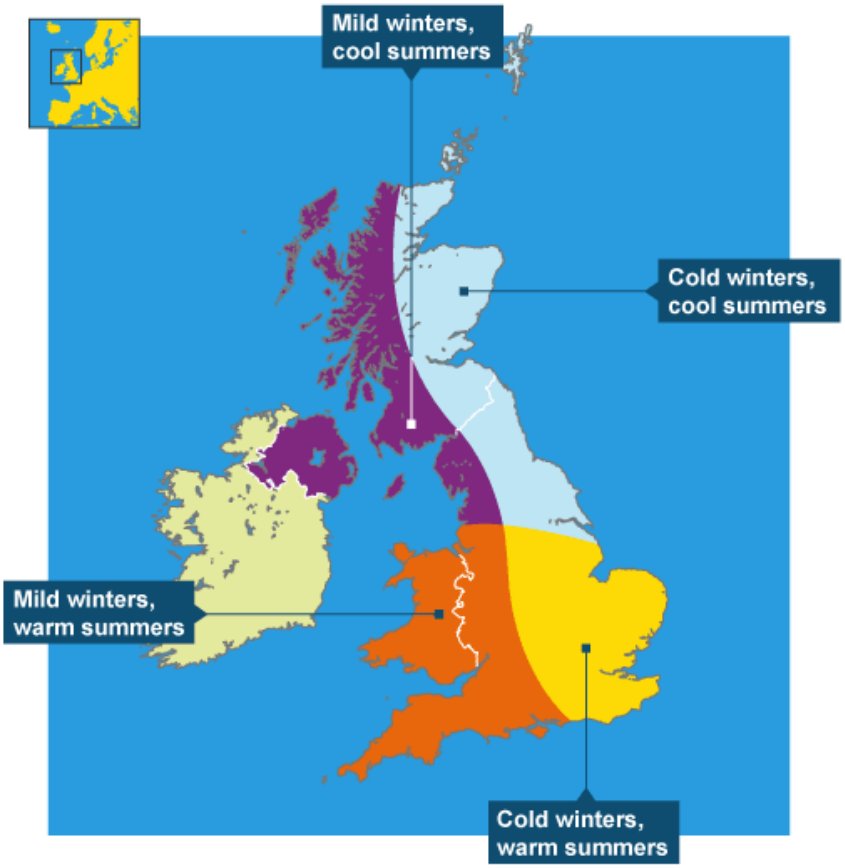

Figure 2. General climate zones in the UK (Source: BBC)

Therefore, CROME was produced by partitioning England into different "agricultural-climate" zones (see Table 1 and Figure 3) and processing multiple parts in parallel. For further optimisation, the individual parts were further subdivided in smaller parts to process multiple months in parallel. Major roads were used to ensure that a parcel can only belong to one of the zones. It is important to note that some of the zone names and their extents don't relate with other established administrative or census extents associated with zone names.

Table 1 lists the number of CROME cells associated with each CROME zone. The CROME zones have grouped together in the public distribution for easier handling.

Data Distribution Zone	Zone Name	Zone Code	Number of Cells
North	North East	NE	4623267
	North West	NW	1087961
	Farne Isles	FR	77
	Hauxley Isles	HX	25
Midlands	East Midlands	EM	4994319
	West Midlands	WM	3837793
South East	South East	SE	9349322
	Isle of Wight	IW	91798
South West	South West	SW	7445233
	Scilly Isles	SC	5042
	Lundy Isles	LN	1174
		Total	31436011

Table 1. CROME processing zones with their zone codes and cell counts respectively.
Figure 3. CROME classification Zones

4 CROME Classification

The classification of the topography was performed using Random Forest classification, 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) to known land covers, previously collected by field inspectors. The ground truth points and radar images were fed into the workflow, outlined in Figure 4, making up the four main work packages described below in further detail.

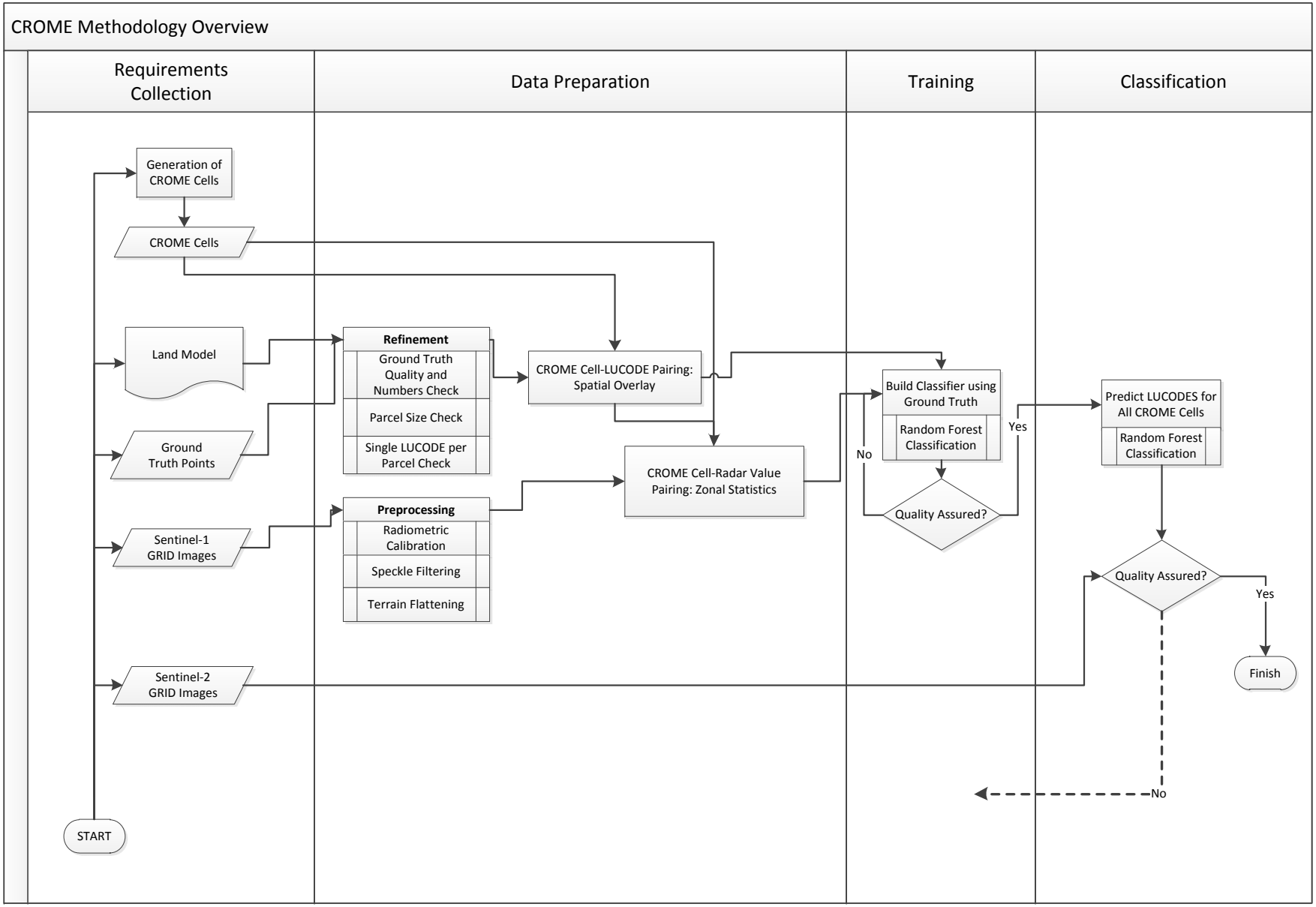

Figure 4. Workflow for CROME classification

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 four 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: These are a record of the spatial location and type of crop and land use found by the RPA field inspectors during late-spring and summer period of 2017. 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 crop/land use type.
    b. Parcel area was bigger than the 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 2017 very high resolution satellite and Ordnance Survey OpenMap-Local (but only those that were visible in 2017 Sentinel-2 Satellite Images). A total of 83388 ground truth samples (i.e. CROME cells with associated with ground truth labels used for Training) were used for building the Random Forest classifier.
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 August 2017, covering the main land masses of England. A total of 1466 scenes were used. These can be downloaded from the scihub (scihub.copernicus.eu) or Alaska Satellite Data Facility.
e) Sentinel-2 Level1C images: These are Sentinel-2 Optical images for the period of January till July 2017. These can be downloaded from the scihub (scihub.copernicus.eu).

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 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) “Zonal Statistics” analysis is performed to calculate the monthly mean and standard deviation 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, and VV, VH, VV/VH standard deviation. Statistics were calculated of each polarization combinations (VV, VH and VV/VH) from 1466 radar scenes captured from January to August 2017 (totaling 4398 images). Monthly 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.

d) 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 (d), and the monthly radar backscattering coefficient values derived in step (c) 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 monthly radar backscattering coefficient 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.

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.

4.4 Classification

Each CROME cell is then applied a LUCODE from the Random Forest classification ruleset. A random visual check using 2017 Sentinel-2 images was performed to detect obvious misclassifications. Simple cases of misclassifications (e.g. slightly rough areas of manmade surfaces conflicted with grass and vice versa) were corrected by calculating normalized difference vegetation index, Modified Soil Adjusted Vegetation Index, Visible Atmospherically Resistant Index, Greenness Index, and Blueness Index.

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.

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 or date range of the images referenced to assign the land use code to the CROME cell. The date format used is YYYYMMDD e.g. 20171116. 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.

5.2 Data Format and Naming

The dataset is supplied in the ESRI shapefile format, together with a valid 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:

<ZONECODE>_<YYYYMMDD>.shp

where:

<YYYYMMDD> is the release date of the dataset

<ZONECODE> is the mnemonic for each CROME 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 v2017.2 has an overall accuracy of 86% and a Kappa coefficient of 0.85, which confirms that a high quality of land use classification is achievable with Sentinel-1 radar data, complimented with Sentinel-2 for distinguishing grass/rough surfaces.

The Ground Truth data didn't 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 doesn't 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 an 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.

7.10 Spatial Completeness

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

7.11 Lineage

Crop types were derived by using Random Forest Classification on Sentinel-1 Radar data, and refined using vegetation indices 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**	**AC07**	**AC16**	**AC17**	**AC19**	**AC20**	**AC23**	**AC32**	**AC36**	**AC44**	**AC58**	**AC59**	**AC63**	**AC64**	**AC65**	**AC66**	**AC67**	**AC68**	**AC69**	**FA01**	**LG03**	**LG04**	**LG07**	**LG11**	**LG14**	**LG20**	**PG01**	**TC01**	**Total**	**User Accuracy**
CROME	AC01	194					3	1		9			7		4		2	2				2	2								226	85.8
	AC03		66			1			6		12	4										1	1	2							93	71.0
	AC07			8																											8	100.0
	AC16				16																										16	100.0
	AC17	2	2	1		156		2	2		1		4	4						1		11		4		1	3			4	198	78.8
	AC19	1		1			90										1	1					1						1		96	93.8
	AC20							6																							6	100.0
	AC23								12																						12	100.0
	AC32	3					7			106					1			2					1								120	88.3
	AC36										21																				21	100.0
	AC44		6						1		1	91										3		2							104	87.5
	AC58						1						37																		38	97.4
	AC59													5																	5	100.0
	AC63	1													221	1		2		3	1	1							3		233	94.8
	AC64															26		1													27	96.3
	AC65						3								1		124												1		129	96.1
	AC66						2	1		8						1	5	334			1		1					4	1		358	93.3
	AC67									1	4				1				181												187	96.8
	AC68																			15											15	100.0
	AC69																1				24										25	96.0
	FA01			1				5					1				2					84		1					5		99	84.8
	LG03	5				1				1	2		6	1								3	141		3			2		4	169	83.4
	LG04																							14							14	100.0
	LG07																								33						33	100.0
	LG11																									23					23	100.0
	LG14																					2					25		2	1	30	83.3
	LG20											1			1				1				1		1			99			104	95.2
	NA01	1	1			3	4	2		2	1	1	7	2			1					32	3	1		1	1		16	13	92	
	PG01	3		5			1	1		2		1	2		1	1						11	1			8	6		277	7	327	84.7
	TC01			1		1				1			6	1				1				8				3			5	83	110	75.5
Total		**210**	**75**	**17**	**16**	**162**	**111**	**18**	**21**	**130**	**42**	**98**	**70**	**13**	**230**	**29**	**136**	**343**	**182**	**19**	**26**	**158**	**152**	**24**	**37**	**36**	**35**	**105**	**311**	**112**	**2918**	
Producer Accuracy		92.4	88.0	47.1	100.0	96.3	81.1	33.3	57.1	81.5	50.0	92.9	52.9	38.5	96.1	89.7	91.2	97.4	99.5	78.9	92.3	53.2	92.8	58.3	89.2	63.9	71.4	94.3	89.1	74.1		

Table 3. Confusion Matrix of CROME v 2017.2

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	WA01	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.