JECAM Guidelines for cropland and crop type definition and field data collection

This document is the first version of the JECAM Guidelines for field data collection. It has been prepared based on 24 JECAM site reports for 2013, and the discussions from several working sessions in the framework of different international on-going efforts, including the ESA Sen2Agri and FP7-SIGMA projects which supports this work. The JECAM partners also provided feedback and suggestions during the JECAM science meeting (21-23 July 2014, Ottawa). This is a living document and will be revised and updated as required.
Contributors : Defourny P., Jarvis I., Blaes X., ...(all the people who will endorsed the document)

I. Objective

The JECAM network facilitates data sharing and collaborative satellite remote sensing research to develop crop assessment and agricultural monitoring methods for a large variety of agriculture systems. This JECAM *Guidelines* document aims to establish a common language for the agriculture monitoring research community regarding the ***definition*** and data collection ***protocols*** of both ***cropland*** and ***crop type*** information. It set the basis of a common approach for cropland and crop type map validation.

GEOGLAM was initiated in 2011, after JECAM inception in 2009. In response to GEOGLAM, JECAM has now become the foundation of the R&D component for GEOGLAM, consequently the JECAM objectives have evolved. The refocusing of the JECAM mandate has necessitated the need for a more coordinated approach to space-based EO observations as well as in-situ observations. This document outlines considerations and requirements for in-situ data, and a separate document outlines EO data considerations. It is expected the minimum data set concept will be applied to a subset of all JECAM sites. These sites will be selected based on their engagement level, representativeness, and desire to openly share EO and in-situ data within the JECAM community.

II. Cropland and crop types definition

The challenge is to share common cropland and crop type definitions that (a) are relevant to, and compatible for, in situ and satellite remote sensing observations, and (b) encompass as much of the diversity found within global agricultural systems as possible.

A. Annual cropland

In a context of crop monitoring systems, the restriction of croplands to "*land that is actually cultivated on an annual basis*" seems relevant, excluding the different stages of fallows and perennial crops. The annual cropland can thus be seen as a distinct land cover type next to other land cover classes (such as permanent cropland, grassland, shrub land, water bodies, etc.).

The general definition of annual cropland (including area affected by crop failure) could be as follows:

The **annual cropland** is a piece of arable land that is sowed or planted at least once within a 12 months period.

FAO has also started to develop an overall framework defining the cropland according to the LCML ISO standards to further document it (cf. preliminary overview in Annex I).

B. Annual cropland class from remote sensing perspective (legend definition)

In the context of global mapping, the identification and the validation of the cropland class is limited by the resolutions of the remote sensing imagery. A specific definition is then proposed:

The **annual cropland from a remote sensing perspective** is a piece of land of minimum 0.25 ha (min. width of 30 m) that is sowed/planted and harvestable at least once within the 12 months after the sowing/planting date. The annual cropland produces an herbaceous cover and is sometimes combined with some tree or woody vegetation* .**

*the herbaceous vegetation expressed as fcover (fraction of soil background covered by the living vegetation) is expected to reach at least 30 % while the tree or woody (height >2m) cover should typically not exceed a fcover of 20%.
**There are 3 known exceptions to this definition. The first concerns the sugarcane plantation and cassava crop which are included in the cropland class although they have a longer vegetation cycle and are not yearly planted. Second, taken individually, small plots such as legumes do not meet the minimum size criteria of the cropland definition. However, when considered as a continuous heterogeneous field, they should be included in the cropland. The third case is the greenhouse crops that cannot be monitored by remote sensing and are thus excluded from the definition.

An "annual cropland map or mask from remote sensing" can be updated on a seasonal or a yearly basis to take into account the extension/reduction of cultivated areas.

C. Crop types and land cover classes (legend definition)

The proposed general legend gathers the global diversity of crop types and follows **a hierarchical grouping of crops to share common aggregation levels**. This grouping is adapted from the Indicative Crop Classification (ICC) developed by FAO in the frame of agricultural censuses[1]. As the considered crop types occur only within the annual cropland area (as defined in section II.B), all the permanent crop type listed in ICC are not considered here. In the same way, all meadows and grasses are considered in separated land cover classes.

Land cover		Crop group		Crop class		Sub-class	Cropping practices
Annual cropland	1	Cereals	11	Wheat	111	winter crop	tillage vs no tillage
					112	spring crop	residues vs no residues
			12	Maize			grain vs fodder crop
			13	Rice			Irrigated vs rainfed
			14	Sorghum			pure vs mixed crops
			15	Barley			…
			16	Rye			
			17	Oats			
			18	Millets			
			19	Other cereals	191	Mixed cereals	
					192	Other	
	2	Vegetables and melons	21	Leafy or stem vegetables	211	Artichokes	
					212	Asparagus	
					213	Cabbages	
					214	Cauliflowers &	

[1]FAO Statistical Development Series, A system of integrated agricultural censuses and surveys, World Programme for the Census of Agriculture, 2010, ISBN 92-5-105375-8, http://www.fao.org/docrep/009/a0135e/A0135E00.htm#TOC

						broccoli	
					215	Lettuce	
					216	Spinach	
					217	Chicory	
					219	Other	
			22	Fruit-bearing vegetables	221	Cucumbers	
					222	Eggplants (aubergines)	
					223	Tomatoes	
					224	Watermelons	
					225	Cantaloupes and other melons	
					226	Pumpkin, squash and gourds	
					229	Other	
			23	Root, bulb, or tuberous vegetables	231	Carrots	
					232	Turnips	
					233	Garlic	
					234	Onions (incl. shallots)	
					235	Leeks & other alliaceous vegetables	
					239	Other	
			24	Mushrooms and truffles			
			29	Other			
	4	Oilseed crops	41	Soya beans			
			42	Groundnuts			
			43	Other	431	Castor bean	
					432	Linseed	
					433	Mustard	
					434	Niger seed	
					435	Rapeseed	
					436	Safflower	
					437	Sesame	
					438	Sunflower	
					439	Other	
	5	Root/tuber crops with high starch or inulin content	51	Potatoes			
			52	Sweet potatoes			
			53	Cassava			
			54	Yams			
			59	Other			
	6	Beverage and spice crops	62	Spice crops	6211	Chilies & pepers	
					6212	Anise, badian, and fennel	
					6219	Other	
	7	Leguminous crops	71	Beans			
			72	Broad beans			
			73	Chick peas			
			74	Cow peas			
			75	Lentils			
			76	Lupins			
			77	Peas			
			78	Pigeon peas			
			79	Other			
	8	Sugar crops	81	Sugar beet			
			82	Sugar cane			
			83	Sweet sorghum			
			89	Other			
	9	Other crops	91	Grasses and other fodder crops			
			92	Fibre crops	9211	Cotton	
					9212	Jute, kenaf, and other similar crops	

					9213	Flax, hemp, and other similar products	
					9219	Other	
			93	Medicinal, aromatic, pesticidal, or similar crops			
			95	Flower crops			
			99	Other			
Perennial crops							
Grassland & meadows							
Fallows							
Shrub land							
Forest							
Bare soil							
Build-up surface							
Water bodies							

This comprehensive legend aims to provide a common framework to translate a site-specific legend to a compatible one across sites using the same hierarchical levels.

An annual cropland area or any other crop type parcel can be labelled as irrigated when irrigation infrastructure is visible on the ground. This is a restrictive definition of irrigation practices but it will allow a consistent definition to be used across regions and seasons.

III. Field data collection for cropland and crop type

A. Objective

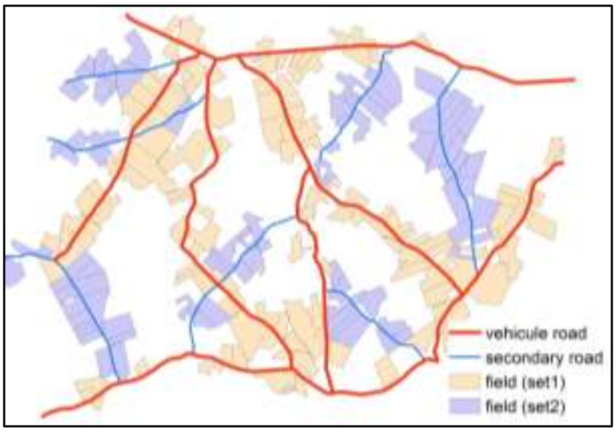
The quality of field observations is very important since they serve as independent reference data to assess remote sensing products (***validation***[2] of cropland and crop type maps). Other field observations can also be used for *calibration*/training purposes. The objective is thus to capture the diversity of crop types and land cover classes (as defined in the section II.C) observed in the JECAM site and collect a sufficient number of samples for each class. The number of observations and the sampling design define the quality of the accuracy estimate. Random sampling is theoretically most suitable to provide independent information and avoid spatial bias due to the location of field observations. Similarly, sampling design based on clusters of samples aligned on a systematic grid with a random start guarantees the selection of a representative and spatially well distributed sample. These statistically-sound sampling plans can't always be implemented on the ground but are most suitable when reference data can be collected by visual interpretation of very high resolution imagery.

Beyond simple validation based on target of opportunity, a more systematic approach for ground truth is proposed here to serve as minimum common protocol to collect reference in situ observation. Due to the diversity of the JECAM sites, the proposed sampling strategy has to be adapted according to the site conditions and expectations are that it will be implemented on a best effort basis.

B. Field observation for validation data set

The ***crop type*** must be identified by field observation during the corresponding growing season for each sampled parcel, field or piece of land larger than 0.25 ha with a minimum width of 30 m. This information can be collected using different tools, such as an application developed for a tablet/smartphone showing the position of the operator on a recent very high resolution image (for

[2] see Annex II – CEOS WGCV land product validation hierarchy



instance, from Bing or Google maps) and the direction of operator movement. Based on ground inspection, the operator adds points (or polygons) to the map by clicking on the screen and defining the corresponding crop types. Several applications have been tested and are described in a separate JECAM document (In-Situ Field Tools).

Collecting crop type points only partly supports cropland validation. It is also important to add a significant set of "non-cropland" points (i.e. other ***land cover*** classes) to complement the validation data set obtained through field observation.

The proposed sampling strategy consists in "windshield survey" along the roads from motorized vehicle. This approach allows a data collector to easily and rapidly capture the entire crop diversity from all visible fields (set 1 as illustrated in the figure). The main principle of this approach is to identify long transects across the JECAM site by selecting a set of appropriate roads. It is however recommended that this sampling strategy be complemented by regular additional transect (set 2) using secondary roads and tracks to reduce the spatial bias brought about by roadside sampling. Several transects running in various directions (NS and EW) ensure the whole coverage of the area of interest while secondary road sampling complements the validation set with less biased data. This secondary sampling has to be regularly spread over the entire site.

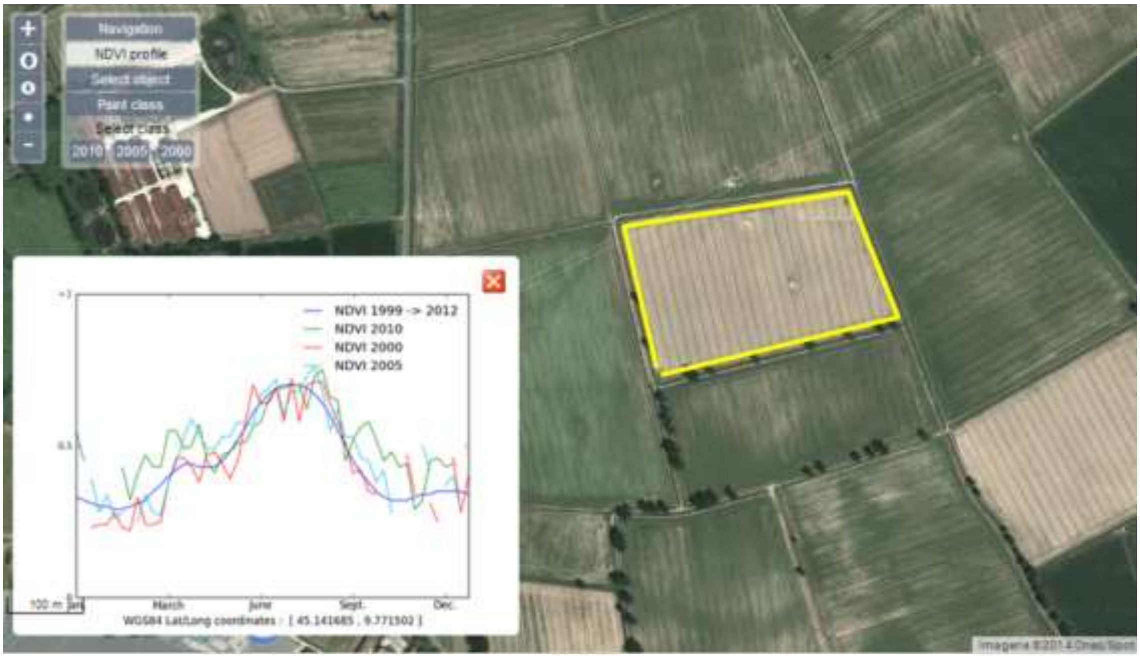
As a rule of thumb, the target sampling density for a cropland map of a typical JECAM site[3] could be around 1 observation/5 sq. km. For the main crop types[4] of a typical JECAM site, the sampling density could range from 1 observation/sq. km of annual cropland to 1 observation/10 sq. km for very simple agricultural landscape. If parcel boundaries are available in a vector format ex ante, the windshield survey can proceed by polygon labeling and a sampling density of 1 observation per sq. km can provide a sampling rate of 1 % of the JECAM site area. . .

C. Complementary data collection for cropland information

Land cover information can be also collected by experts through photo-interpretation using updated very high resolution imagery and associated NDVI profile in a web interface. Experts have the ability to label randomly drawn objects (see yellow boundaries in the figure, next page) based on (1) VHR images and (2) NDVI temporal profiles (10 years mean, previous year, and current year) computed from medium resolution images. Such an interface is already operational as illustrated below and will be made available on request for the different JECAM sites.

[3] According to GMFS experience, much larger area of interest (>50 000 sq. km) can be covered by a stratified sampling with a density of 1 observation/150 sq. km.

[4] The main crop type(s) can be defined either by the JECAM site managers or by the different crop types that correspond separately to at least 10 % of the annual cropland and cover together about 75% of the annual cropland surface.



The legend is language independent and use color code and symbology.

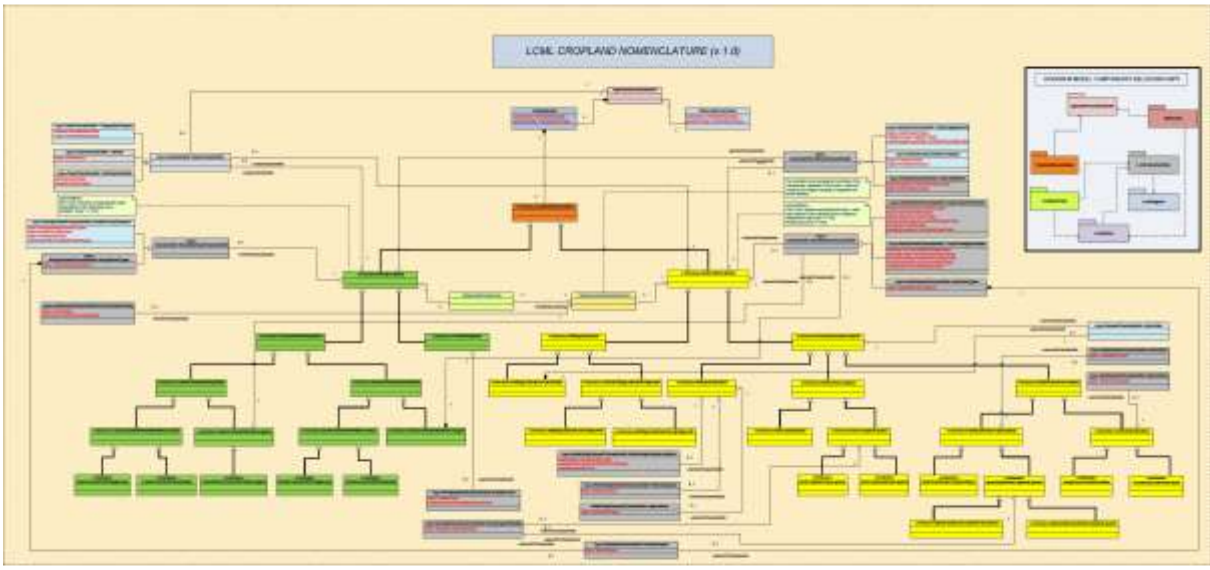
Symbology	Label
	Annual cropland
	Permanent cropland
G	Grassland (including meadows)
	Shrub land
	Build-up surface (sealed surface)
	Water bodies
	Forest
	Bare soil
?	“Data unclear” if the interpreter cannot decide
	Fallow land
	Other land cover

In order to assess the quality of visual interpretation, two to three operators should label samples in order to cross check the quality and quantify the consistency of the photointerpretation results.

	independent sample sets			cross-validated sample sets		
	set 1	set 2	set 3	set 4	set 5	set 6
operator A	■			■		■
operator B		■		■	■	
operator C			■		■	■

Protocol example for quality control of expert interpretation to build reference dataset.

Annex I: LCML Cropland Nomenclature (v1.0)



Annex II: The CEOS WGCV Land Product Validation Hierarchy

Stage 1 Validation	Product accuracy is assessed from a small (typically < 30) set of locations and time periods by comparison with in situ or other suitable reference data.
Stage 2 Validation	Product accuracy is estimated over a significant set of locations and time periods by comparison with reference in situ or other suitable reference data. Spatial and temporal consistency of the product and with similar products has been evaluated over globally representative locations and time periods. Results are published in the peer-reviewed literature.
Stage 3 Validation	Uncertainties in the product and its associated structure are well quantified from comparison with reference in situ or other suitable reference data. Uncertainties are characterised in a statistically robust way over multiple locations and time periods representing global conditions. Spatial and temporal consistency of the product and with similar products has been evaluated over globally representative locations and periods. Results are published in the peer-reviewed literature.
Stage 4 Validation	Validation results for stage 3 are systematically updated when new product versions are released and as the time-series expands.