European Commission

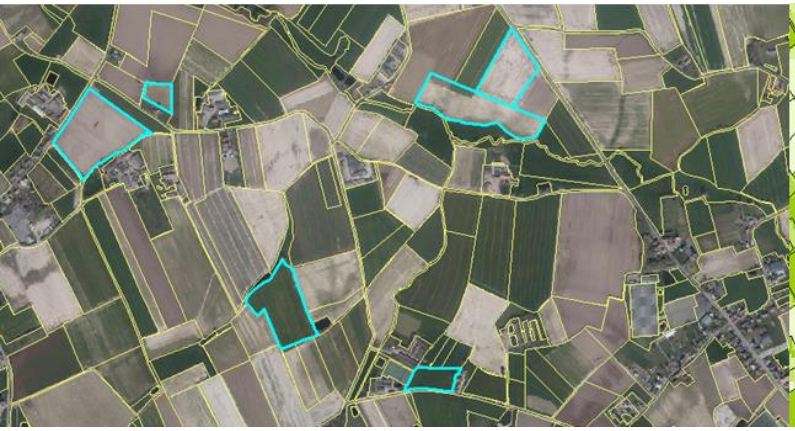
JRC TECHNICAL REPORTS

Assessing the Validity of the Feature of Interest in the context of the CAP Checks by Monitoring

Progress report

Pavel MILENOV, Blanka VAJSOVA, Gilbert-Madalin VOICAN, Laura ACQUAFRESCA, Konstantinos ANASTASAKIS and Guido LEMOINE

2021



Joint Research Centre

EU Science Hub
https://ec.europa.eu/jrc

JRC123711

How to cite this report: MILENOV Pavel *et al.*, Assessing the Validity of the Feature of Interest in the context of the CAP Checks by Monitoring, Ispra, 2021, JRC123711

Contents

The work presented in this report results from the joint efforts of the authors, with the indispensable support from the whole GTCAP team. Authors would like to specially thank to their colleagues, Dominique FASBENDER, Nataša LUKETIC, Csaba WIRNHARDT and Michele MERONI, for their valuable contribution to specific parts of the work. Special thanks to Philippe LOUDJANI and Aleksandra SIMA from GTCAP for the revision and proofreading of the report. Authors would like to thank also the MS Agricultural Administrations for the data provided and technical support offered.

Glossary of Terms

AG	Agro-Forestry
AL	Arable Land
CAP	Common Agricultural Policy
CbM	Check by Monitoring
CwRS	Control with Remote Sensing
DEM	Digital Elevation Model
DIAS	Copernicus Data and Information Access Services
DN	Digital Number
DSM	Digital Surface Model
ENL	Equivalent Number of Looks
GRD	Ground Range Detected
GLC	Global Land Cover
GSAA	Geospatial Aid Application
GTCAP	Geodata and Technologies for the Common Agricultural Policy
FOI	Feature of Interest
IACS	Integrated Administration and Control System
IQR	Interquartile Range
LQ	Limiting Quality
LPIS	Land Parcel Identification System
LULUCF	Land Use, Land Use Change and Forestry
MS	Member State
QA	Quality Assessment
PG	Permanent Grassland
PC	Permanent Crop
SAR	Synthetic Aperture Radar
SNR	Signal-To-Noise Ratio
TG	Technical Guidance document

Abstract

The objective of this document is to provide a rather detailed summary of the work done by GTCAP in the period October 2019 – April 2021, on the development of automated methods to assess the conditions on Feature of Interest (FOI), in the context of the Checks by Monitoring (CbM). The importance of the subject arises from the key role of the FOI in the CbM, as being the spatial reference for all signal extractions from Sentinel data and the item for all decisions made for the relevant scenario expected. Although focussed on the compliance aspects of the FOI, the document also covers the capturing of more generic information on FOI conditions, relevant in a broader environment and climate-related contexts (EU CAP conditionality, green infrastructure, LULUCF, provision of ecosystem services, etc.).

The work presented in this progress report does not stem from one single research project, but from several parallel research and technical activities, conducted by different groups dealing with specific topics (CbM, LPIS design, LPIS QA), where the FOI aspects play important role. Most of the development was done in close (multi-actor) collaboration with experts from the relevant EU MS administrations and other EC services, and went hand-in-hand with an incremental built-up of common knowledge about the addressed phenomenon and its conceptualization in the domain of geographic information (GI). Despite of the implementation-oriented character of the objectives, the work had to cover also a number of theoretical and exploratory aspects. The overall framework was very dynamic, with user requirement evolving and changing in time. The methods and approaches presented should be considered as prototypes, which could be further developed/extended in the future, on the basis of more mature and tailored requirements, providing the availability of sufficient resources. Authors hope that despite of the intermediate character of the results presented, this progress report provides useful and first-hand information to the EU CAP technical community on the FOI implementation aspects and helps to "streamline" the future work in this respect.

The given report is organized in chapters. Chapter 1 provides the concept of the Feature of Interest. Chapter 2 describes the different FOI assessment methods designed, as well as their theoretical basis. Chapter 3 explains the elaboration of the reference data used for testing and validation of the methods. Chapter 4 introduces the area dimension in the FOI assessment. Chapter 5 presents results and discuss the outcomes. Chapter 6 deals with the implementation aspects, new solutions and future prospects. Chapter 7 outlines some specific use cases, such as LPIS update and cross-compliance, where the FOI assessment methods are providing added-value.

1 The concept of the Feature of Interest (FOI)

Checks by Monitoring (CbM) introduced the concept of Feature of Interest (FOI) to deal with the spatial aspect of the bio-physical phenomenon present on Earth. FOI is in fact the "space" occupied by the observable physical object on the ground; its spatial "footprint". In the EU CAP context, it often coincides with the single unit of agricultural management (single crop field or particular land use, located on homogeneous agricultural land cover).

In the CbM system, the FOI has two spatial representations (features). The first is derived from the Geospatial Aid Application (GSAA) and uses a polygon as geometric primitive. The second is derived predominantly from Satellite data (mostly Sentinel 1 and 2) and can be expressed in different "value structures" - statistical metric, clusters of image pixels, image segments. Both representations (being abstractions of a real-world phenomenon, in the context of ISO 19101) serve to represent the very single true physical object, or FOI (Figure 1).

Figure 1: Examples of single unit of agricultural management in two different regions of Europe. (a) field prepared for potatoes in North-West Europe (b) field with barley in Mediterranean

The information collected from Sentinels on a given FOI could reveal the presence and also persistence of "objects" of different nature inside. It could further shed light on the characteristics of those individual bio-physical phenomena and their spatial distribution (Figure 2). It could also indicate whether the FOI representation from GSAA reflects the true physical object behind.

Spatial heterogeneity within the FOI representation detected by Sentinel signal, could relate to three cases:

- Inherent and expected variations within the physical entity: A typical example is the pro-rata grassland where the herbaceous and woody life forms co-exist and constitute a stable "intrinsic mix".
- Alien physical entities present in the same unit of management: A typical example is the presence of an object of non-agricultural nature (for example buildings) within the FOI representation.

- Several physical entities corresponding to different units of management: A typical example is the presence of two plots – one of arable land and one of permanent grassland - within the FOI representation.

The latter two require particular attention, since they reject the initial hypothesis that:

1. the FOI representation from the GSAA relates to one and only one physical object on the ground,
2. this object is of purely agricultural origin and belongs to a single agricultural land cover category (arable land, permanent grassland, permanent crop),
3. it is spatially congruent to both FOI representations (from GSAA and from Sentinel).

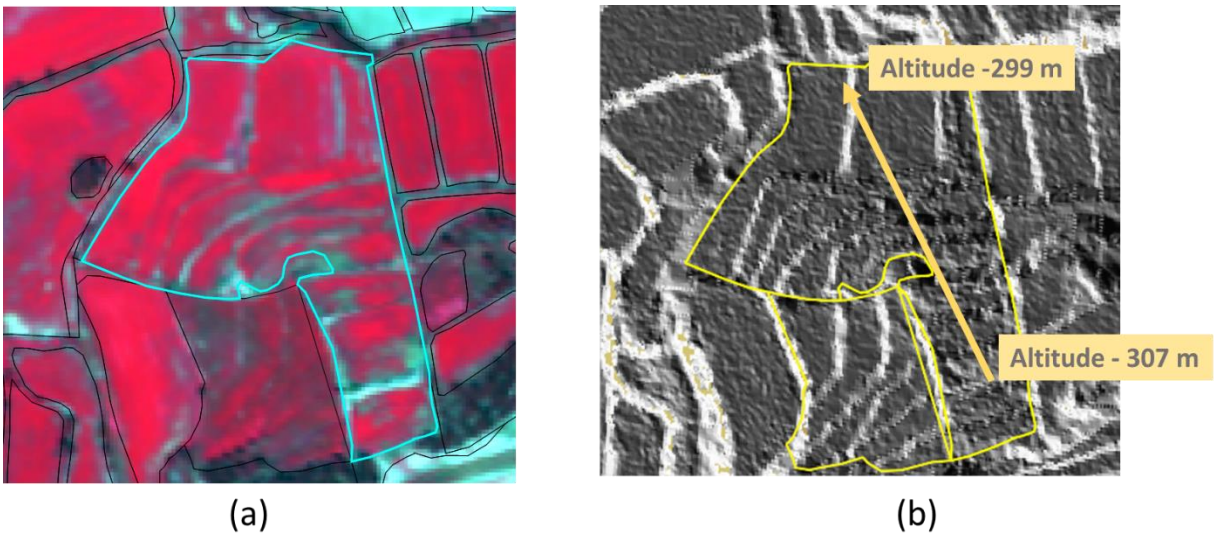


Figure 2: Example of barley field in Catalonia (ES), being single unit of management and at the same time heterogeneous. Heterogeneity is caused by the "intrinsic mix" between the cropped area and the terraces used to regulate the slope and make it suitable for arable cropping. (a) the parcel as appearing on very high resolution satellite imagery; (b) the parcel as appearing on local digital surface model. (VHR: GeoEye-1, May 2019, © 2018 European Space Imaging; DSM: Modelo digital del terreno en las áreas de riesgo potencial significativo de inundación (ARPSIs), MAPAMA

The validity of this initial hypothesis is an important "boundary" condition of the CbM's domain of discourse, as it ensures that:

1. the area component - officially known hectares of agricultural land cover - provided by the Integrated Administration and Control System (IACS), is correct and fixed upfront;
2. data derived from Sentinel is exclusively associated with the properties of the monitored physical object, which guarantees meaningful analysis and tangible CbM decisions.

The truthfulness of the FOI representation in relation to the actual FOI depends on the system design and quality of LPIS/GSAA datasets; the closer the representation to the actual FOI, the better the performance of any processing in CbM. For example, FOI representation containing issues, such as non-excluded ineligible features, multiple land cover classes, would most probably induce noise during the automatic processing of Sentinel signal (Figure 3).

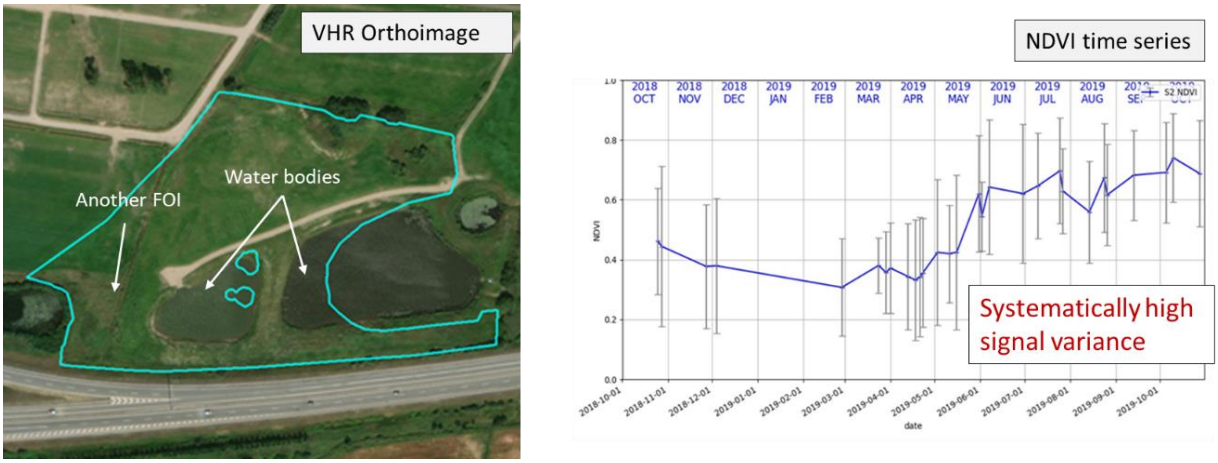


Figure 3: Issues with spatial cardinality propagating a noise in Sentinel signal used in CbM

The importance of the spatial heterogeneity in the FOI and the spatial congruency of the FOI representations to the physical object on the ground have triggered the design of a special set of methods to detect and eventually quantify the spatial heterogeneity and congruency, for CbM purposes. Since the spatial congruency of the FOI representation to the actual FOI is manifested through "one to one" cardinality between the representation and the true phenomenon, the term applied in the CbM is "spatial cardinality", by convention expressed with code G1. It is the first spatial type of information extraction in the CbM; a second, generic type of information extraction dealing with FOI heterogeneity, is called "spatial variability" and expressed with code G2 (JRC, Checks by Monitoring Quality Assurance Framework, ANNEX III, Identifying the CbM decisions, version 1.1 2020).

Once integrated within the CbM, both G1 and G2 are expected to provide key information along the decision process:

- A confirmation of the "one to one" spatial cardinality (validity of the initial hypothesis) will ensure suitability of the given GSAA associated with the payments schemes, to act as FOI representation for the CbM process;
- A rejection of the "one to one" spatial cardinality (Figure 4) will require either:

    1. An interaction with the farmer, when the GSAA-FOI cardinality is "one to many"; or
    2. A spatial aggregation process when GSAA-FOI cardinality is "many to one".

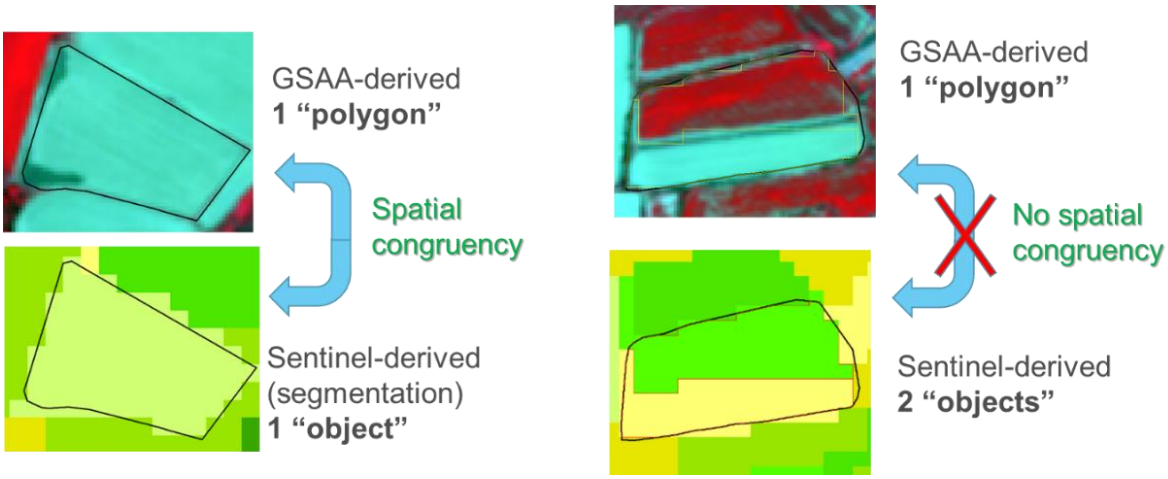


Figure 4: Illustration of the "spatial congruency/cardinality" in the context of Checks by Monitoring

The "spatial variability", G2, in other hand, could confirm the anticipated presence of persistent heterogeneous properties within the physical object (pro-rata grassland, presence of landscape features), as shown in Figure 2, or could detect the presence of multiple agricultural land uses within FOI, (a commonly occurring issue for the scenarios related to greening payment scheme).

2 Methods and Algorithms

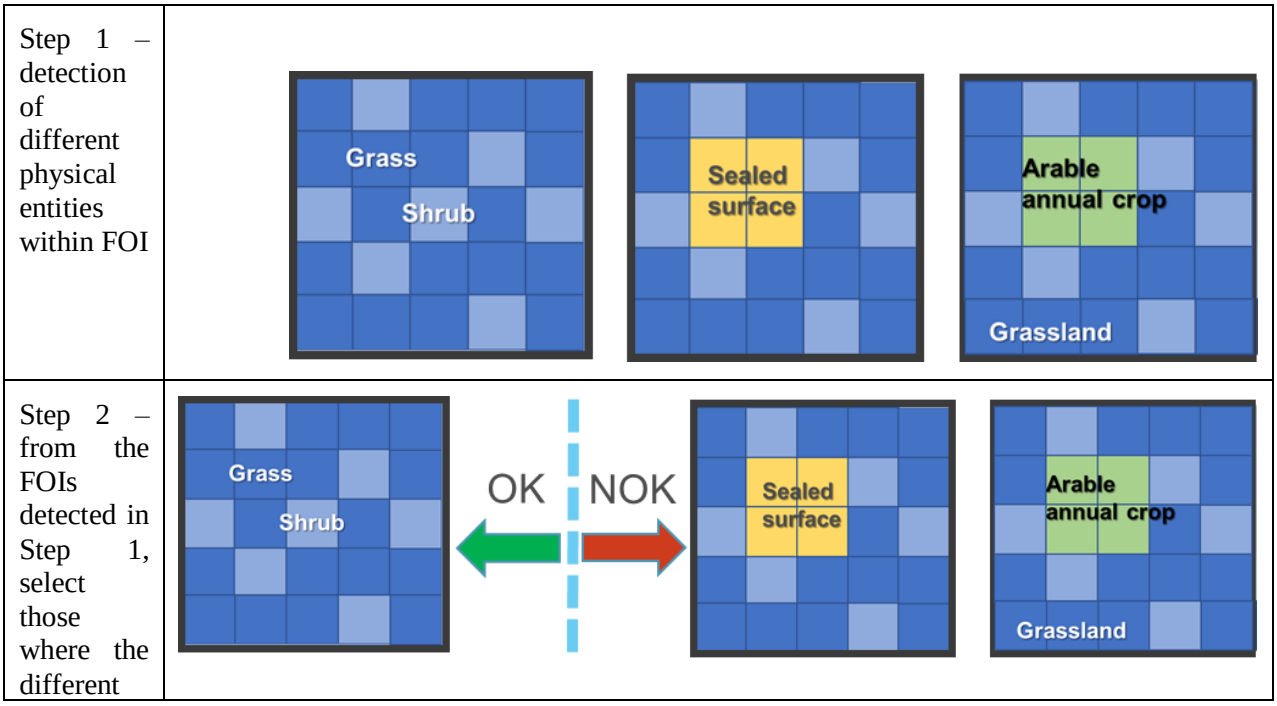
GTCAP launched, already in 2019, a series of activities to develop and test different methods for the detection of the spatial cardinality and spatial variability within the FOI representation. There were three aspects involved:

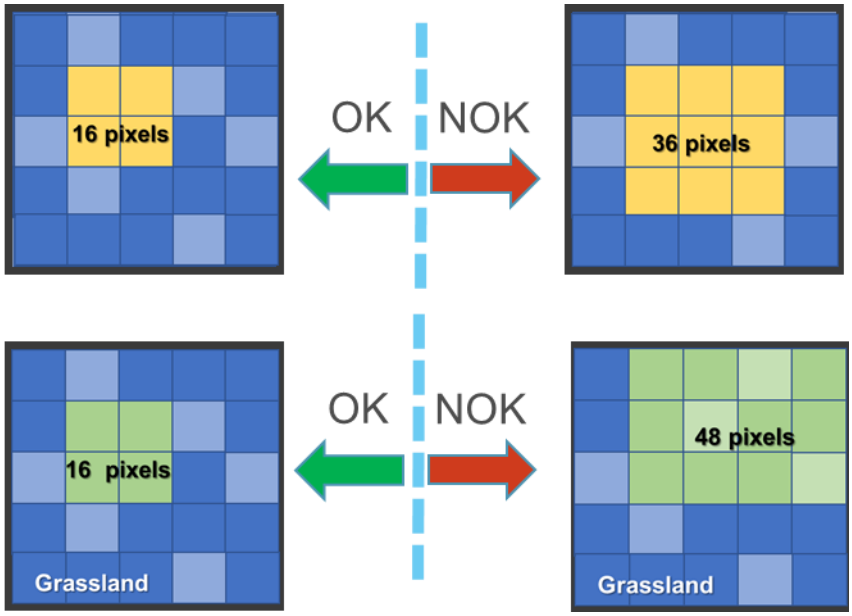
1. Elaborate a number of methods using different Sentinel signals, at the level of proof of concept;
2. Test the performance of the developed prototypes;
3. Select and compile the most successful methods in a Jupiter notebook (to be further deployed in JRC DIAS GitHub). More information of the subject is given in the following sections.

A tiered-approach for the development of the method was established, with the aim to cover both spatial cardinality (G1) and spatial variability (G2); with the former considered a special case of the latter. For spatial cardinality G1, there were two sub-cases identified depending on the severity of the spatial mismatch between the FOI representations. The approach comprised the following three steps (tiers):

1. Step 1: Search for "different objects" within the FOI representation from GSAA – spatial variability (G2);
2. Step 2: Search for "different physical entities" that could invalidate scenario – spatial variability (G2) to a degree that it jeopardizes the Sentinel signal with a given scenario. This is the first sub-case of severity;
3. Step 3: Search for "alien entities" big enough to invalidate the area – spatial cardinality (G1) not respected to a degree it jeopardizes the area component as provided by IACS. This is the second sub-case of severity.

Figure 5: Illustration of the 3 steps (tiers). Each shaded square represents an area occupied by a certain type of physical object (land cover)



physical entities, are not expected	**Inherent variations of given characteristics within physical entity (pro-rata grassland)** **Different physical entities present in the same unit of management** **Different physical entities each related to different units of management**
Step 3 – from the FOIs detected in Step 2, select those where the different physical entities, are big enough to jeopardize the area component	

The implementation of step 3 certainly required a minimum threshold for the size of the different physical entities within the FOI representation, beyond which the area component would be jeopardized. Based on technical research in the course of this project and following the logic applied in the LPIS update (JRC 2015), it was agreed that occurrence of contiguous clusters with more than 20 pixels (equal to 0.2 ha for Sentinel imagery) within the FOI representation, indicates the presence of physical objects big enough to challenge the area component provided by IACS. The pixel size and thus final area depends on sensor used. More information on the subject is provided in a separate chapter below.

For the G1 method, further assumptions have been made for the input conditions, related to the detection of "one to many "cardinality i.e.:

- persistence on the ground for sufficient time to invalidate the given scenario;
- manifestation through the presence of entities having different behaviour in time;
- distinct behaviour that can be captured by the Sentinel signal;
- different physical entities that are sufficiently large in image pixel cluster size to challenge the area component.

The focus of the work was exclusively on detecting cases where the FOI representation from GSAA relates to many objects/units (GSAA ⊃ FOI). The opposite case, when one object/unit relates to many GSAAs (GSAA ⊂ FOI) was not explored, since it was considered as not jeopardizing the "CbM boundary" conditions. Nevertheless, this latter case plays a role in the assessment and processing of the so-called "small parcels".

Although the G1 methods were designed to detect issues with the spatial cardinality, the ultimate outcome was a confirmation of the validity of the FOI with respect to:

1. the "officially known area" and the type of agricultural land cover recorded in IACS;
2. the correspondence of the graphical representation of the FOI (one-to-one spatial match with reality, or correct portion of larger unit).

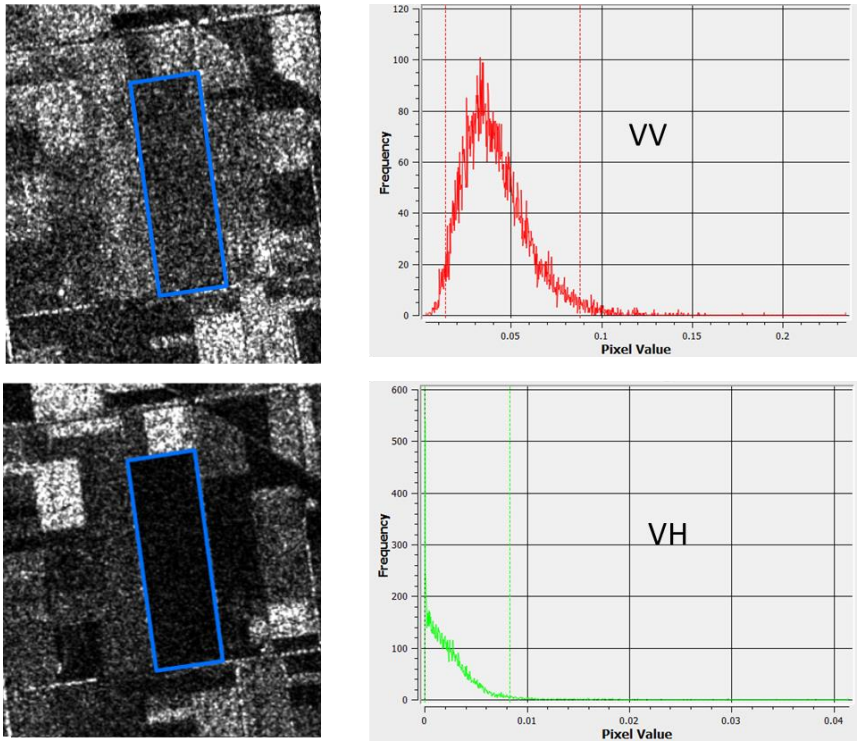
Four methods were initially developed and tested:

- Analysis of Sentinel 1 backscattering's speckle noise: Assumes that in homogeneous fields the SAR backscattering speckle is following a Gamma distribution.
- Threshold on Sentinel 2 signal-to-noise ratio: It uses the ratio between the observed NDVI average and observed NDVI standard deviation.
- Unsupervised clustering through S2 image segmentation: It looks for clusters of pixels (grouped in segments) with distinct behaviour in time.
- Multi-temporal S2 supervised classification: It assesses the land cover types found within the FOI representation derived from pixel-based supervised machine learning.

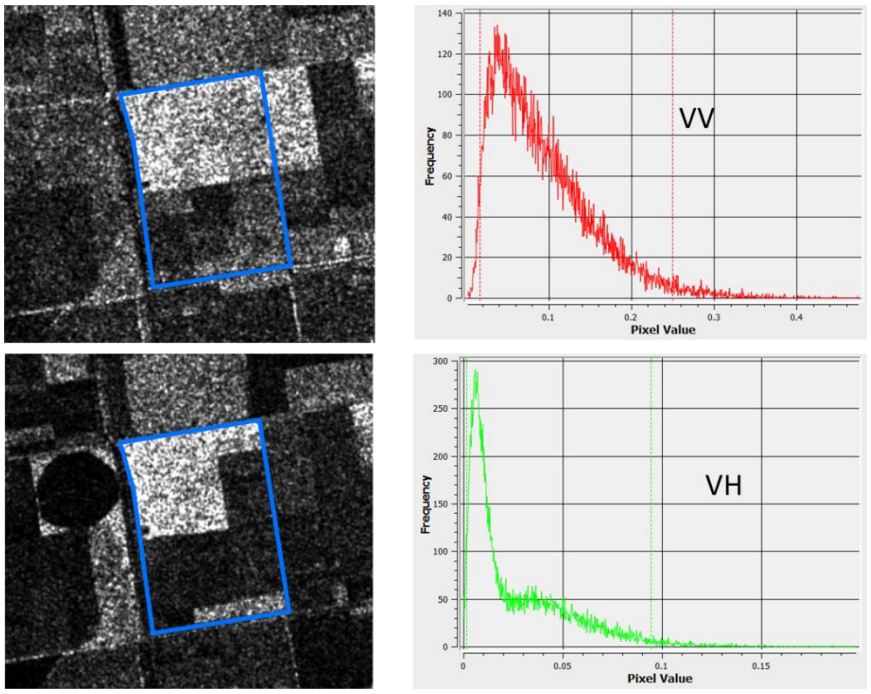
2.1 Analysis of S1 backscattering's speckle noise

2.1.1 Description of the method

The methodology of the test stems from the well accepted fact that over homogenous fields the variance in backscattering intensity (speckle statistics) of the SAR signal follows a Gamma distribution (Nezry 2014). The statistical test compares the observed variance computed over the field (FOI), with a theoretical variance, if the field would be homogeneous. The theoretical variance is modelled via the Gamma distribution using Equivalent Number of Looks (ENL) and the calculated mean backscatter coefficient for the observed field (FOI). The ENL is solely dependent on the processing parameters that are used to generate the geocoded CARD-BS product from the Level 1 GRD input. ENL ~ 4.4 for the 10m pixel spaced CARD-BS. The test simulates x number of random repetitions to calculate and compare the observed with the theoretical (expected for homogenous field) variance. If the positive difference between observed variance and the hypothetical variance for the homogenous field could be considered statistically significant (denoted by alpha, set usually between 0.05 and 0.01), and such difference is observed over a number of sequential observations (Sentinel 1 images), the homogeneity hypothesis for the given field is rejected and the result of the test reports that the given field (FOI) is heterogeneous.



(a) Spatial variations of the radar signal backscattered by homogeneous", or "textureless" target (as the blue parcel above) are only due to speckle, which statistical properties depend on the radar system and the image production system. The probability distribution function of the speckle in intensity can be approximated, by a Gamma distribution.



(b) In the case above (blue parcel containing sub-parcels) distribution of backscatter for VH has "two bumps". It is in fact, a superposition of two Gamma distributions with two different parameters. The shares of the superposed distribution are equal to the proportions of the sub-parcels. The superposed distribution has a larger variance than a "clean homogeneous Gamma distribution", as further explained in Annex I.

Figure 6: Histograms of the backscattering coefficient (Sigma_0) for two polarization modes – VV and VH- on (a) homogeneous and (b) heterogeneous fields.

The basic equations used are given below:

$$ENL = \frac{M^2}{\sigma^2} \Rightarrow \sigma^2 = \frac{M^2}{ENL}$$ theoretical variance for homogenous field

$$V = \frac{1}{n}\sum_{i=1}^{n}(sigma0_i - M)^2$$ …observed variance of tested field

Input parameters

M (mean)	average backscatter coefficient
σ	standard deviation
n	number of pixels within the tested field
ENL	equivalent number of looks (depends on processing; for S1, GRD, it is 4.4)
$sigma0_i$	backscatter coefficient for pixel i, (called Gamma_0, in the JRC DIAS hub)
V (var)	variance (= σ^2)

The statistical hypothesis test compares the observed sample variance (that follows a Chi-square distribution) with the expected variance under the Gamma distribution hypothesis. The test requires to compute a probability on both distributions. Unfortunately, there is no clean analytical expression for this probability. In order to cope with this issue, the algorithm generates simulated values from both distributions and computes a numerical estimations of the probability. In practice, if the simulations (n_{sim}=100 000) issued from the variance estimate are in majority (more than 100*(1-alpha) %) larger than the simulations issued from the Gamma distribution, then the field can be marked as heterogeneous. Otherwise, there is no statistical evidence that the field is heterogeneous and the field is marked as homogeneous.

For the calculations, the natural values of backscatter coefficient are used, instead of the decibel values. The method considers the use of the VH polarized signal only, as according to the scientific evidence it is more sensitive to the difference in volume scattering of a fully closed vegetation canopy and surface scattering of a bare soil compared to VV polarized signal (Chauhan 2016) and (Harfenmeister, Spengler and Weltzien 2019). Both ascending and descending modes of the Sentinel 1 are used.

The reliability of the test was performed on 10 consecutive repetitions over the same image. The outcome resulted stable and therefore reliable.

There are two periods defined for the sequential Sentinel 1 observations: (1) March-May and (2) March – July. They intend to reflect the time frame for the active vegetation on winter/spring arable crops and for permanent grassland, respectively. More details are given in the chapter on the preparation of the reference data for quality assessment.

Two decision scenarios for detecting and confirming heterogeneity have been tested:

- FOI is heterogeneous if detected on 3 and more S1 images
- FOI is heterogeneous if detected on more than 50% of the S1 images within a month

Both scenarios have been tested with and without applying negative buffer of 10 meters.

2.1.2 Meaning of the output produced

The raw values were extracted in from the Google Earth Engine platform, due to the availability of Sentinel 1 raw signal data over the tested area at that time (beginning of 2019). The method was implemented in the Jupyter notebook (same symbols as above, used in the script). It produces a table with the outcome of the comparison of the variances for each of the observations (Sentinel 1 images/ acquisitions). The main input parameters are: the FOI ID (OBJECTID_1); the calculated mean value of the observed backscatter (mean); the calculated variance of the observed backscatter (var); the number of pixels found within the FOI (n), which could be slightly different for each S1 observation (due to the georeferencing), and the number of simulations.

Table 1: Output of the S1 "Gamma Fit" method

OBJECTID_1	mean	var	n	nsim	obstime	ObsP	test
2282048	0.00428365	0.000009	177	100000	2019-05-02T17:54:58	1	1
2282048	0.00473049	0.000012	177	100000	2019-05-08T06:01:14	1	1
2282048	0.00503025	0.000010	177	100000	2019-05-14T17:54:59	0.99996	1
2282048	0.01085112	0.000042	178	100000	2019-05-20T06:01:15	0.99989	1
2282048	0.008556	0.000019	176	100000	2019-05-26T17:54:59	0.83784	0
2282048	0.00403037	0.000005	182	100000	2019-05-02T06:00:18	0.99619	1
2282048	0.0099532	0.000034	174	100000	2019-05-03T17:45:58	0.99932	1
2282048	0.00500312	0.000006	181	100000	2019-05-08T17:54:06	0.7655	0
2282048	0.00492752	0.000009	178	100000	2019-05-14T06:00:19	0.99999	1
2282048	0.00622181	0.000025	179	100000	2019-05-15T17:45:59	1	1
2282048	0.01223551	0.000061	178	100000	2019-05-26T06:00:20	0.99999	1

The main output parameters are as follows:

ObsP — number of cases (as fraction of the total number) when simulated information from observed variance is bigger than the one derived from expected variance for assumed to be homogeneous fields. If this percentage value is higher than the set threshold (1-alpha, e.g. 99% confidence level), there is an indication that the homogeneity hypothesis should be rejected (i.e. test=1)

obstime — observation time of the S1 image acquisition

test — decision on homogeneity/heterogeneity, i.e.0/1

As could be seen from the example below, to assess a given field (FOI) for the period of March-May, approximately 10/11 images (depending on the position of the FOI within the Sentinel 1 granule) were available for the analysis and for the decision on heterogeneity/homogeneity. For some dates, even two images could be acquired (for instance: S1A ascending and S1B descending) and therefore assessed.

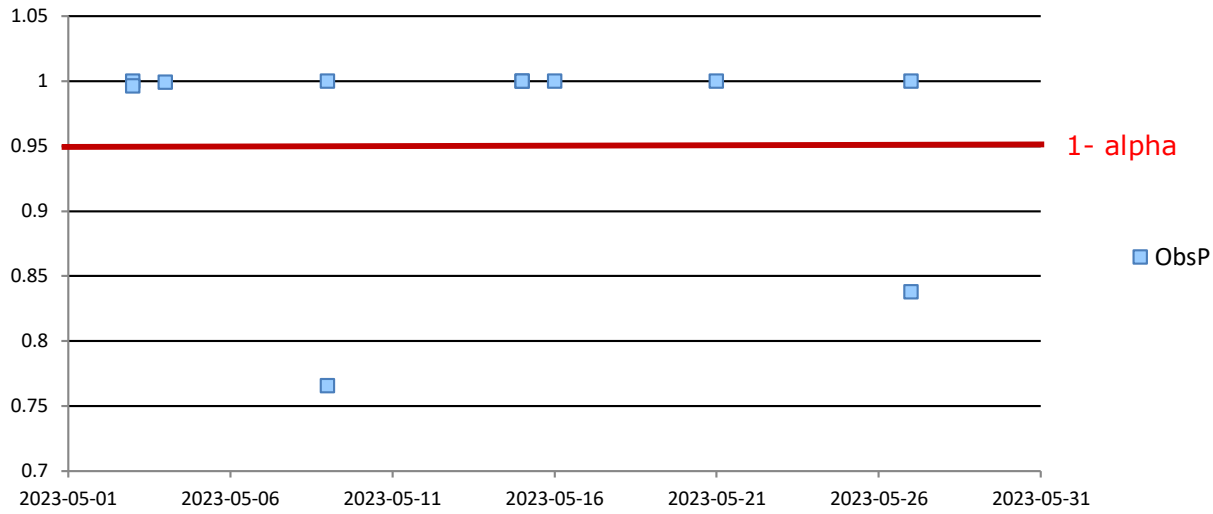


Figure 7: Visualization of the ObsP values over time for an FOI of 9.9 ha, flagged as being heterogeneous.

2.1.3 Calibration and validation

The performance of the method in the different scenarios was assessed regularly during the design periods in order to fine-tune the parameters to increase the detection accuracy. Two types of assessment have been applied:

(1) against the reference dataset of 390 FOIs, 50% of which were below 0.6 ha, interpreted using optical data (VHR alone and VHR + Sentinel 2);
(2) visual analysis of 100 random parcels from the total FOI population over the test area, all with area above 2ha, on the Sentinel-1 images themselves.

The validation started with the results of the S1 Gamma method on the same FOIs for one single date in May (close to the date of the VHR acquisition, used as reference), in order to understand how well the method performs on single observation. Then, the focus was put on how to best combine the results from different dates to improve the overall detection performance in relation to the FOI heterogeneity.

For each of the validations, confusion matrices have been prepared and kappa coefficients[1] have been calculated.

The resulted agreement found between the S1 single image approach and the reference optical data was very low (kappa from 0.008 to 0.052). There could be

[1] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/

several explanations for that. First, the overall parcels size of the test area (Catalonia, Spain) was probably too small (Tomas 2020) and not suitable for calibration and more rigorous studying of the behaviour of algorithm. Some fields were so small that there were no results from backscatter extraction. Small fields (e.g. < 1 ha) have insufficient number of samples to reliably estimate the statistics needed in this test. The rugged topography of the area also contributed to higher observed variance of the signal backscatter, and consequently, to the high number of commission errors. There were also numerous parcels with small ring and narrow corridors within. The choice of the VH polarization could be also a factor; however, further tests are needed to confirm such relation. Second, it was in general challenging to compare decisions from optical and SAR data, as both were based on concepts laid on different physical basis and thus revealing the reality from different perspective. The optical reference data has been revised several times in order to align both observation methods to address the same properties of the land phenomenon. The big difference between the spatial resolution of the VHR and the S1 (valid also for S2), and thus the diverse information content also played a role.

The resulted agreement found with the SAR data, used as reference, was higher (kappa 0.24), but overall remained rather low. Part of the problem in this case, was the more challenging photointerpretation of the SAR images.

In any case, accuracy seems to improve once results from at least two images in the given period are used. The outcome from a single S1 observation might be influenced by accidental excessive precipitation; thus the use of more observations would counterbalance such sporadic effect. Overall, averaging several consecutive S1 images over a time interval (e.g. a month) would reduce speckle (include the effective ENL) and result in a better detection of permanent variation within a parcel. This led to the assumption that the results derived from one S1 image will not be reliable, unless if information is available about the weather condition (precipitation) over the area in this specific period. Further comparisons with more dates were made at later stage, with the more refined versions of the reference dataset. Another element that contributed to the improvement of the performance (on sufficiently large parcels) was the application of negative buffer on the FOI geometry, to reduce the effect of bordering pixels. More information on the subject is given in the chapter on results and in Annex I.

2.2 Threshold on S2 Signal-to-Noise Ratio (SNR)

2.2.1 Description of the method

The method is based on somehow similar assumption as the previous one that the presence within the FOI of different physical features or variability of a given bio-physical property will be manifested through the spatial variability of the Sentinel 2 signal. In this case, the used metric is the so-called signal-to-noise ratio (SNR), expressed as the ratio between the observed average and observed standard deviation of the Normalize Difference Vegetation Index (NDVI), and calculated from the S2 pixels within the FOI. If smaller than 5, it indicates that the FOI representation might not be homogeneous.

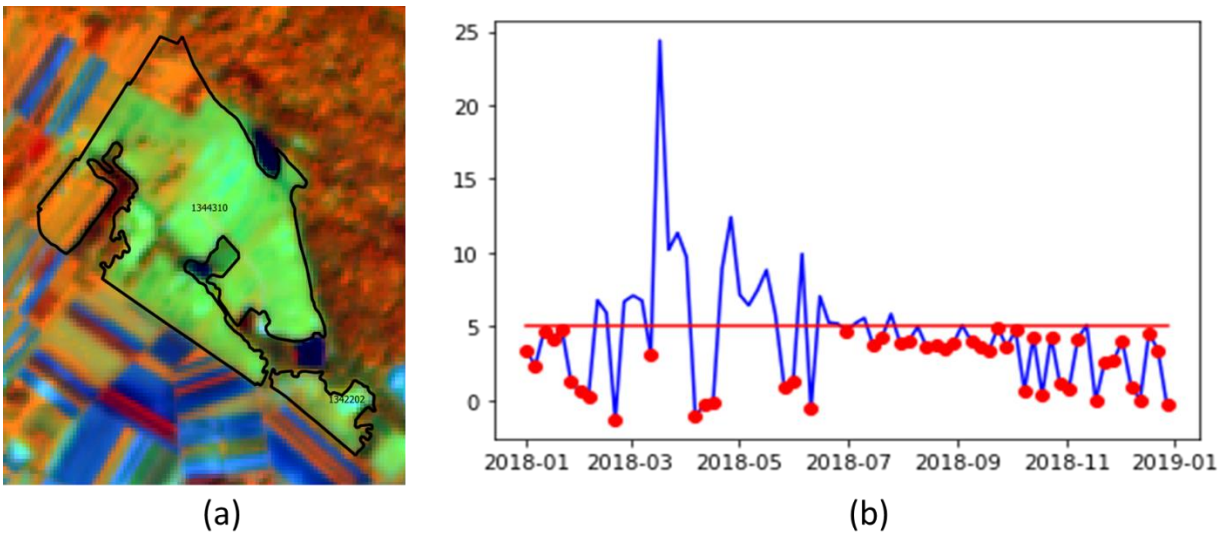


Figure 8: (a) Example of FOI with notable heterogeneity in land use, well-visible on optical data, (b) Example of evolution of the Signal-to-Noise Ratio values over time

The basic equation used is given below:

$SNR = \frac{M}{\sigma}$ Single-to-Noise Ratio

Input parameters

M (mean) average value of the NDVI

σ standard deviation of the NDVI

Both the mean and the standard deviation are calculated from the Level 2A product of Sentinel-2, which is atmospherically corrected. For each S2 acquisition, the FOI is flagged as potentially heterogeneous, if SNR < 5. This threshold, although arbitrary, was chosen after a number of trials, as a good compromise between the commission and omission errors.

As with the previous method, in order to confirm heterogeneity, which is persistent, the SNR should be constantly below the value 5, for a number of S2 acquisitions.

The periods defined for the Sentinel 2 observations were the same as previous: (1) March-May and (2) March – July.

Two decision scenarios for detecting and confirming heterogeneity have been tested:

- FOI is heterogeneous if detected on 3 and more S2 images
- FOI is heterogeneous if detected on more than 50% of the S2 images within a month

Both scenarios have been tested with and without applying negative buffer of 10 meters around the FOI perimeter.

2.2.2 Meaning of the output produced

The method was implemented as stand-alone python script (accessing Sentinel-2 data on the JRC DIAS GitHub. It produces a table with the SNR outcomes for each individual observation (Sentinel 2 images/ acquisitions). The main input parameters are: the FOI ID (OBJECTID_1); the calculated mean value of the observed NDVI (Mean); the calculated variance of the observed backscatter (StDev); and the number of pixels found within the FOI (n).

Table 2: Output of the S2 SNR method

OBJECTID_1	Mean	StDev	n	ObsDate	SNR	Flag
54478401	0.2381	0.143009	315	3/19/2019	1.664933	1
54478401	1.046815	0.19792	315	3/29/2019	5.289084	0
54478401	1.144773	0.222681	315	3/31/2019	5.140857	0
54478401	1.023469	0.362445	315	4/5/2019	2.82379	1
54478401	1.193018	0.279646	315	4/13/2019	4.266167	1
54478401	1.177289	0.397448	315	4/15/2019	2.962119	1
54478401	1.125523	0.352868	315	4/18/2019	3.189641	1
54478401	1.094556	0.257311	315	4/20/2019	4.253827	1
54478401	1.101215	0.257511	315	4/23/2019	4.276385	1
54478401	1.60382	0.710691	315	5/25/2019	2.256706	1
54478401	0.2381	0.143009	315	3/19/2019	1.664933	1

The main output parameters are as follows:

ObsDate	Date of the S2 observation (acquisition)
SNR	The value of the Single-to-Noise Ratio
Flag	Decision on homogeneity/heterogeneity, i.e.0/1

2.2.3 Calibration and validation

The performance of the method in the different scenarios was assessed regularly during the design periods in order to fine-tune the parameters and increase of the detection accuracy. The data for validation was a reference dataset of 390 FOIs over a test area in Catalonia.

The validation started with the results of the S2 SNR method on the same FOIs for one single date, in order to understand first how well the method performs on a single observation. The results from one S2 date are compared visually with the relevant image and the agreement found was more significant comparing to the

S1 Gamma method. The SNR approach seemed less conservative than the S1 Gamma, producing much lower number of commission errors when ran on single image. Results obtained showed also that the SNR method had good performance with respect to FOI cardinality, and rather doubtful with respect to heterogeneity. This required further investigation; however, it should be noted that the comparison was done on FOIs with permanent crops, where SNR based on NDVI might not be sufficiently sensitive. The critical parameter to properly set was the minimum number of images where the given phenomenon (observed heterogeneity) is present.

Some attempts for fine-tuning, by adjusting the threshold (set at 5 by default), have been made. By increasing the threshold, the FOI heterogeneity detection performance seems to improve for all land use categories, the most notable increase obtained was for arable land (AL). As for the FOI cardinality detection, the threshold of 5 seemed to give the best result. For AL, the optimal performance related to cardinality check, was obtained with 4 as a threshold. Still the number of commission and omission errors remained rather high, with kappa= 0.4 for AL and kappa = 0.2 for PG. Part of the problem was the difference in the perspective on heterogeneity and cardinality applied when creating the reference data, being much more holistic and context-related than the rather simple statistical approach used in the detection. The interpretation of the FOI characteristics on the reference data evolved during the project implementation in order to converge with both observation methods (SNR and visual assessment for the reference data) and address the properties of land phenomenon in the same way.

Aside comparison between the S1 Gamma and the S2 SNR results has been made. It showed higher agreement between both methods, than the agreement between each of the method to the reference data (Figure 9 and Figure 10). This was a further indication that, while both detection methods are based on the physics of the phenomenon (signal backscatter/light reflectance capabilities of the vegetation canopy and soil), the visual interpretation of the reference data uses other contextual information which cannot be easily reflected by the selected metrics of the designed methods.

	status	SNR	
		homogeneous	heterogeneous
S1	homogeneous	**433**	**40**
	heterogeneous	**38**	**23**

Figure 9: Comparison between the S1 Gamma method (alpha =0.01) and the S2 SNR method (SNR<5), based on 543 S1-S2 image pairs assessed on 27 fields.

	status	S2 visual (cardinality)	
		no cardinality	cardinality
SNR	no cardinality	72	11
	cardinality	4	10

7 fields the same

	status	S2 visual (cardinality)	
		no cardinality	cardinality
S1 test	no cardinality	23	12
	cardinality	53	9

Figure 10: Comparison between the S1 Gamma method (alpha =0.01) and the S2 SNR method (SNR<5), based on. one S1-S2 image pair assessed on 100 fields. Seven of the FOIs found as omission errors, were the same for both methods

2.3 Unsupervised clustering through S2 image segmentation

2.3.1 Description of the method

This method is conceptually different from the previous ones. Instead of assessing the heterogeneity and cardinality using alphanumerically based statistical metrics at the level of the FOI, the unsupervised clustering method tries to assess the FOI conditions through an object-oriented analysis of contiguous segments (clusters), detected within the FOI and extracted upfront from the image pixels. There are two versions of the method:

(1) fully object-oriented, where both the extracted image segments and the FOI geometries are treated as objects with topological relationship between them, allowing the conduction of complex contextual analysis;

(2) simplified, where the extracted images segments are treated as thematic raster, and where the subsequent analysis is based on the zonal statistics derived at the level of the FOI vectors, used as regions.

While the former method provides more sophisticated instruments for the analysis, the latter is considered more suitable for implementation, as being less complex and eventually more portable.

The underlying idea of the method is that every bio-physical object/phenomenon has particular behaviour in time, which depends on the material (biotic or abiotic) it is made of, and its properties. Objects of biotic nature interact with their environment (soil, air) and exchange material and energy through their biological cycle, which is rather short (could be few months within the year). Abiotic features, in other hand, exchange material and energy through their geological cycle (if natural) or its "functional" cycle (if anthropogenic). These cycles are rather long and could span from decades to eons. If one assumes that, each specific type of matter has a particular temporal behaviour associated with its life cycle, one ought

to be able to depict certain type of land phenomena by studying the dynamics of a particular characteristics captured through the given observation method, without necessarily tackling its structural-physiognomic aspect (Figure 11).

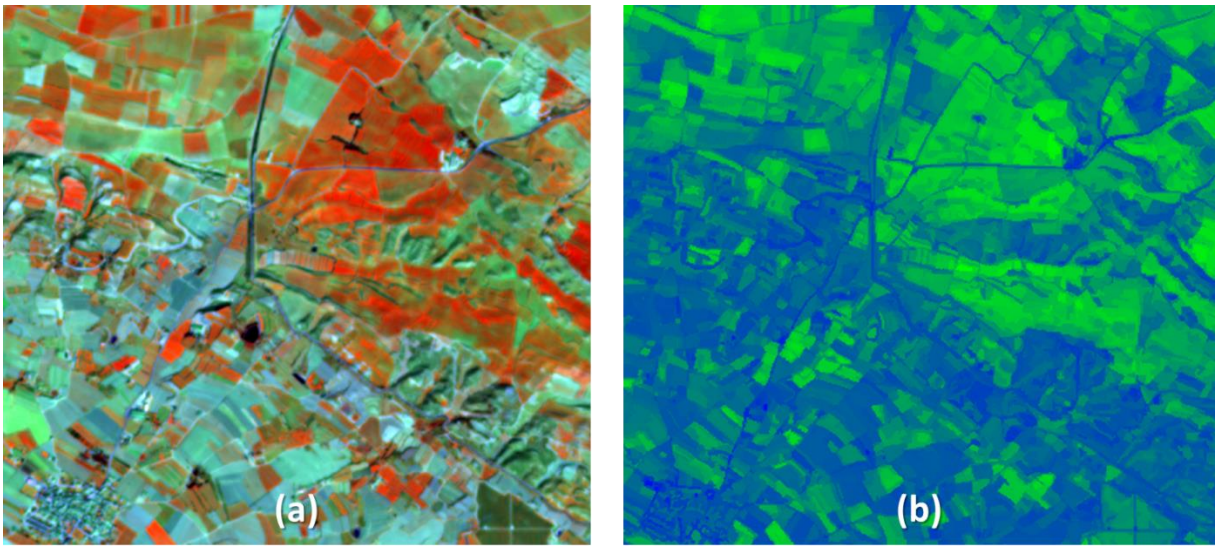


Figure 11: (a) Part of rural area in Spain, as seen from Sentinel-2 image. The artificial surfaces and build-up features are clearly visible, due to their specific appearance in the given spectral combination (bright colour, rectangular shape); (b) Image segmentation over the same area, generated from 20 Sentinel-2 images acquired all over the year. The colour shade of each segment shows the maximum difference of mean intensity from all images (from dark blue to green). Artificial surfaces and build-up areas are in darkest shades of blue, since they have the same reflectance through the year; thus, the maximum difference in brightness intensity is minimal.

The initial step involves a multi-resolution segmentation of all cloud free S2 L2A images within a pre-defined period. The Sentinel 2 bands used are: BLUE (B2), GREEN (B3), RED (B4), RED EGDE B6), NIR (B8) and SWIR-1 (B11). The spectral bands with spatial resolution of 20 meters are re-sampled to 10 meters. Each individual image band is defined as separate "image layer" in the process workflow. The segmentation uses region grow method, based on relative homogeneity criteria, which takes into account the standard deviation of the "spectral colours" and the deviation of the resulting segments from a compact (or smooth) shape. The granularity of the segmentation is controlled by scale factor, which is set in a way to provide sufficient detail, without excessive over-segmentation. All Sentinel 2 bands for all acquisitions have the same weight in the segmentation process. No specific geometric adjustment between S2A and S2B are applied upfront, under the assumption that the predominance of the geometrically correct images will compensate at certain extent the adverse influence of these occasional shifts. The FOI boundaries are introduced as separate thematic layer to separate the segments falling in neighbouring FOI representations. There were two periods defined for the multi-temporal stack used for the segmentation: (1) March-May and (2) March – July.

For every resulted segment, the maximum difference in the mean intensity of each segment, in all "image layers" in the time series, is calculated with the formula given below:

$$Maxdiff = \frac{\max_{i,j \varepsilon Kb} |\acute{c}i(v) - \acute{c}j(v)|}{\acute{c}(v)}$$, with feature value range $\left[0, \frac{1}{Kb} c_k^{max}\right]$

Input parameters

- *i, j* are image layers
- $\acute{c}(v)$ is the brightness of image object v
- $\acute{c}_i(v)$ is the mean intensity of image layer i of image object v
- $\acute{c}_j(v)$ is the mean intensity of image layer j of image object v
- c_K^{max} is the brightest possible intensity value of image layer k
- K_B are K_b is the number of image layers with positive brightness weight with $K_B K_b = \{k \in K: w_k = 1\}$, where w_k is the image layer weight.
-

$\acute{c}(v) = \frac{1}{w^B}\sum_{k=1}^{K} w_k^B \acute{c}_k(v)$, with feature range $[c_k^{min}, c_k^{max}]$

- w_k^B is the brightness weight of image layer k with $w_k^B = \begin{cases} 0 \\ 1 \end{cases}$
- K is the number of image layers k used for calculation
- w^B is the sum of brightness weights of all image layers k used for calculation with $w^B = \sum_{k=1}^{K} w_k^B$
- $\acute{c}_k(v)$ is mean intensity of image layer k of image object v
- c_k^{min} is the darkest possible intensity value of image layer k
- c_k^{max} is the brightest possible intensity value of image layer k.

The parameter Maxdiff is used as sufficiently reliable proxy for the dynamics of the reflectance of the bio-physical objects in the different spectral bands (Roshan Pande-Chhetri 2017). This reflectance-based metric is then considered as indicator for the material-behavior of the biotic/abiotic object/phenomenon. The second step involves the calculation of the decile points of the whole population of segments, ranked by the values of the maxdiff. They are further used as thresholds to assign (classify) each of the segments to its correspondent decile range (no fuzzy rules for class assignment are defined). Each decile range becomes a class of a particular range of reflectance (behavior) dynamics.

The neighboring classified segments (10 class nomenclature) having the same class label are further merged in an upper topologically – linked object layer. Small segments are subsequently eliminated by their merging with the common object with the largest common boundary and with similar spectral characteristics.

Although the method was design to address all three steps (tiers) in the FOI assessment with respect to cardinality and variability, the focus was given to the detection of those physical entities within FOI that either invalidate the scenario or compromise the area component. FOIs having segments (with notable size) labeled with different classes and with particular class distance were flagged as potentially "invalid" with respect to 1-1 cardinality.

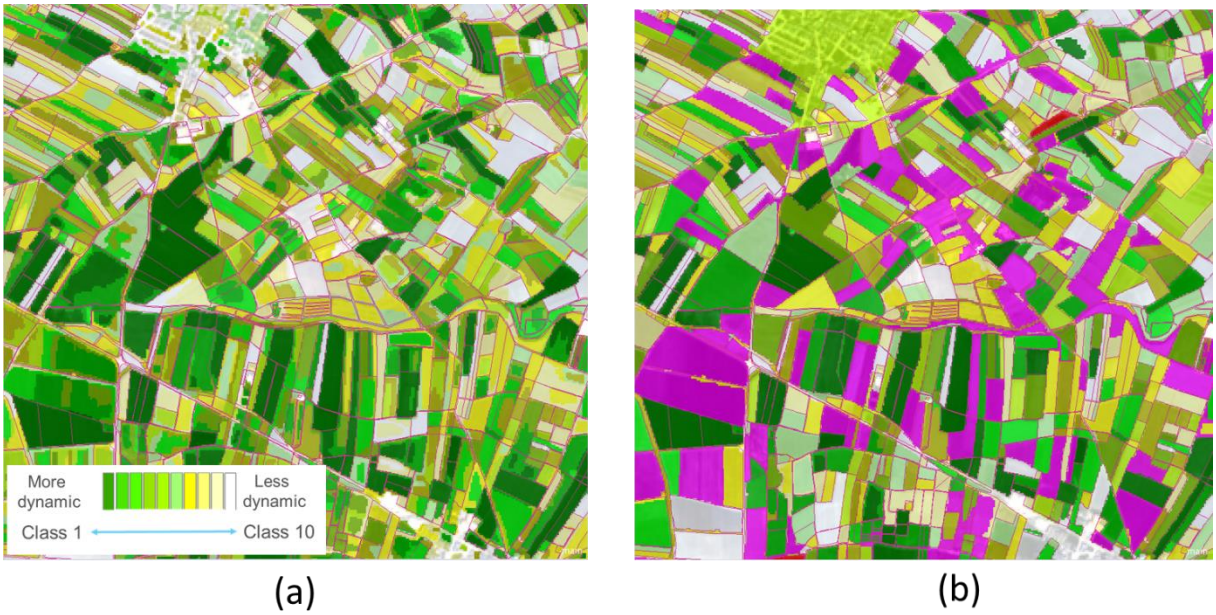


Figure 12: (a) FOI boundaries (in red) with the segments classified according to the quantile range of the maxdiff they belong to; (b) FOIs found with potential problem with respect to cardinality are highlighted in magenta.

Three different rulesets for flagging an FOI as having cardinality problem, have been tested (all, with and without the application of negative buffer of 5 meters):

Ruleset 1:

- Two or three segments with class difference >= 2, within the FOI

AND

- Four or more segments with class difference >= 1, within the FOI

Ruleset 2:

- Two or three segments with class difference >= 2, within the FOI

AND

- Four or more segments with class difference >= 1, within the FOI

AND

- Different segments occupy at least 10% of the FOI area

Ruleset 3:

- More than one segment with at least 10/20 pixels

AND

- Class range difference between all segments >= 1

2.3.2 Meaning of the output produced

The method was developed as proof-of-concept in eCognition Developer, where the above-mentioned metrics are pre-defined and available. It uses as input, the cloud free Sentinel image scenes over the area of interest, extracted upfront from the JRC DIAS GitHub, as well as the FOI geometries with the relevant GSAA attributes, as ESRI shape file. The designed ruleset provides two outputs – table with statistics for each FOI assessed and a thematic raster file (classified segments). Some of the information in the exported table, is taken directly from the GSAA attributes, such as: FOI ID (OBJECT_ID), FOI area in hectares (Area_ha), and declared land use (LU_ type)

Table 3: Output of the S2 unsupervised clustering method

OBJECTID_1	Area_ha	LU_type	Num_Segm	Class_Diff	Num_Layers	Flag
2474394	0.5085	Barley	1	0	20	0
2474395	0.1763	Barley	1	0	20	0
2474451	0.0941	Non-productive	0	0	20	0
2474452	1.2399	Non-productive	1	0	20	0
2474472	0.7003	Agro-forestry	1	0	20	0
2620982	3.1764	Non EFA alfalfa	3	2	20	1
2622683	1.3894	Vineyards	3	2	20	1
2693167	27.3019	Flax	13	4	20	1
2771319	5.9152	Barley	7	3	20	1
2782920	7.3308	Non-productive	14	5	20	1
2782921	7.3308	Non-productive	14	5	20	1

The main output parameters are as follows:

Num_Segm	Number of segmented within the FOI
Class_Diff	Maximum difference in the class values of the FOI segments
Num_layers	Number of Sentinel images (layers) used
Flag	Decision on cardinality issue (not present/present, i.e.0/1)

To allow independent statistical analysis also outside the eCognition environment an alternative method for assessing the FOI conditions was designed.. It uses the thematic raster file alone, overlays it with the FOI geometry, and applies spatial analysis, based on the rules defined above, in a separate GIS environment.

2.3.3 Calibration and validation

The performance of the method in the different scenarios was assessed regularly during the design periods in order to fine-tune the parameters and increase of the detection accuracy. Different settings of the segmentation (scale, shape, compactness), as well as different weighting of the image bands, have been tested.

Although a comprehensive analysis of the outcomes from the different setups is not yet performed, the preliminary observations showed no significant difference in their performance. The standard (default) segmentation setting (compactness = 0.8, shape = 0.1) seems to work the best. Also excluding certain image bands images bands from the segmentation process and the maxdiff analysis did not change substantially the results. One reason for this could the possible correlation between some of the spectral bands with respect to their behavior in time. Although the amplitude of each spectral band is different, thus allowing a more precise quantitative assessment of bio-physical parameters, as green LAI and chlorophyll content, their temporal evolution is rather similar and could be grouped in two principle "vectors" – the one related to the part of light spectrum, which is reflected by vegetation, and the other related to the part of light spectrum absorbed by vegetation (Figure 13). Previous studies on the use of object-oriented approach for heterogeneity analysis, show that the use of NIR and RED bands alone in the segmentation process, provide sufficiently accurate clusters, while increasing the computation time (Tasdemir K 2012)

Another aspect was the sufficiency of the observation sampling rate to "reconstruct" adequately the behavior of the reflectance. In order to have significant population of samples to construct the Maxdiff histogram and to allow a contextual assessment outside with FOI, the method uses entire Sentinel images. In order to reduce the impact of clouds and shadow, it excludes all S2 images reported with any clouds in the scene metadata. This reduces significantly the number of useful observations; thus the sampling rate. However, the analysis done during the preparation of the reference data for testing (explained in the following chapter), as well as the empirical knowledge of the life cycle of the agricultural phenomenon shows that in order to capture consistent presence of physical features that could invalidate the FOI integrity (1-1 cardinality), the minimum required frequency of observations is 1 image per month (a fairly achievable target even for the Nordic countries).

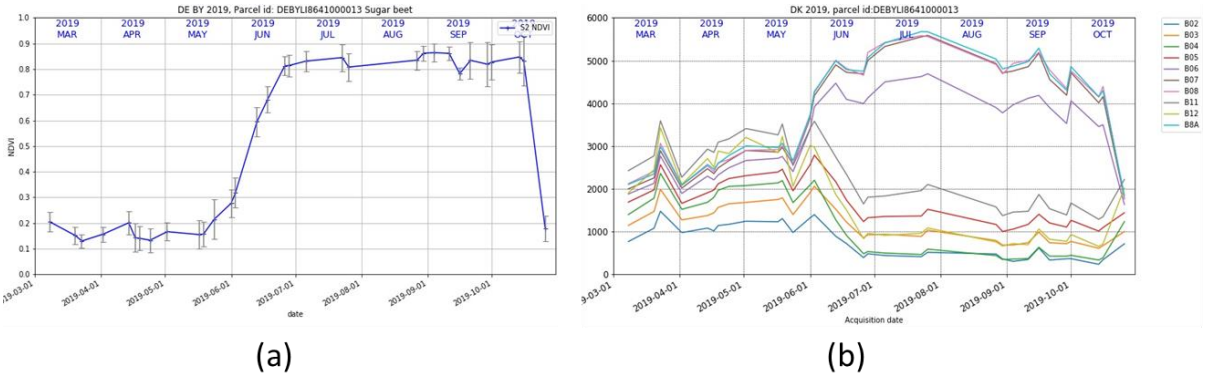


Figure 13: (a) NDVI time series of an parcel cropped with sugar beet; (b) Time series of the full set of Sentinel-2 bands over the same parcel. The two behavior trends (high reflectance – low reflectance) are clearly visible

The more challenging element to define was the distribution and density of the Sentinel-2 observations along the timeline. There will be always some months with more could free observations than others. So far, the method uses all of them, which might create an unbalance (bias) in the segmentation and maxdiff feature space toward the months with denser sampling; thus affecting adversely the representation of the behavior. This aspect should be studied further.

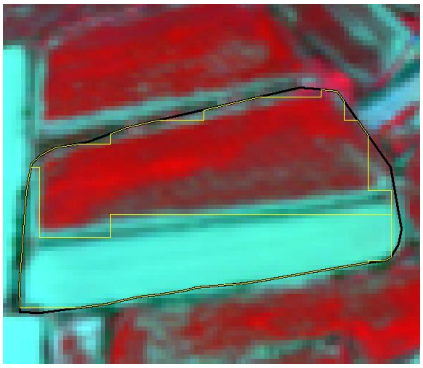
One possible solution could be to decide on the minimum period of occurrence (at least several consecutive dates) of a given unexpected phenomenon in order to confirm a presence of an issue with cardinality. This period of occurrence probably depends on the type of land cover/land use defined on the given FOI. The two periods defined for the segmentation - March till May and March till July – where meant to address different crop and land use types.

Last but not least, the method was found sensitive to the noise introduced by the quantization effect of the bordering features, creating meaningless segments at the edge of the FOI, propagating in the analysis. The adoption of negative buffer of 5 meters (more conservative value applied, due to the post-processing of the initial segments already performed) partially resolved the problem; however, with the risk to omit important information on FOI with complex shape.

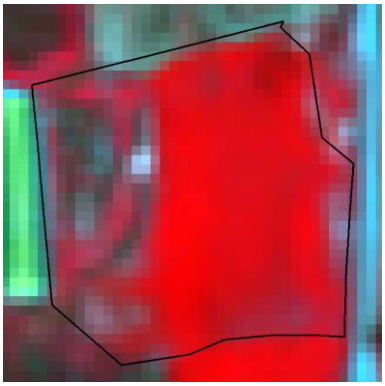
The unsupervised clustering method was tested regularly though its development against the reference dataset of 390 FOIs prepared over the test area of Catalonia. The assessment revealed that while the multi-temporal segmentation and unsupervised classification (based on 10 classes depicting 10 behaviors) seems accurate in depicting clusters with different behavior, the subsequent decision rules for the FOI validity (number of segments, max class difference) should be revised. In general, the method was sufficiently accurate to capture the FOI with cardinality issue, at the expense of relative large amount of false positive (FOI detected being with problem, while they were not). The overall kappa was in the range of 0.2 and 0.3.

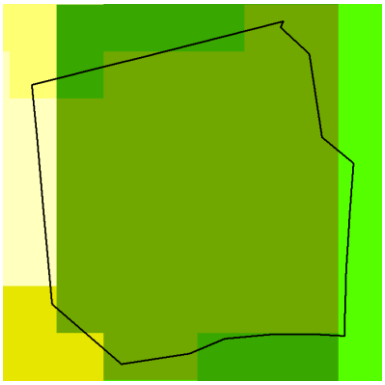
The main problem was that the decision rules for the FOI assessment were generic and applicable for all types of land cover land use phenomena, while they should be tailored to the specific context. The rules (class difference >=1) were designed to detect presence of different objects within the FOI, regardless of their type – presence of different crops or presence of crop and a building. Such high sensitivity, however, accidently picks up also intra-parcel variability, typical for the agricultural landscape (soil inundation, impoundments, landscape elements, terraces), as the one of Catalonia (Figure 14). There were certainly also some technical issues related to the cluster extraction method itself, such as: (1) adverse impact on segmentation of objects with high contrast present on the FOI border (for example, paved roads), which results in neighboring segments propagating within the FOI entity; (2) adverse impact on segmentation of the occasional geometric shift in some of the Sentinel 2B images.

Correct detection of the presence of two agricultural parcels within the FOI (black contour). The two parcels visible on the VHR imagery (left picture), resulted in two segments with notable class difference of the (right picture)

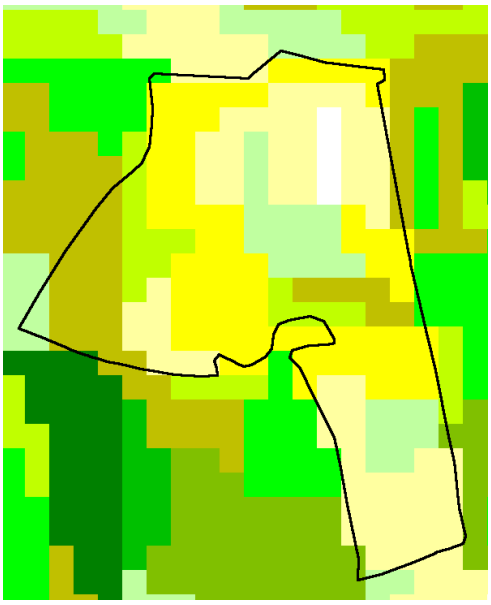


Omission of the presence of two agricultural land cover within the FOI (black contour). The two parts visible on the VHR imagery (left picture), were merged in one segment (right picture)

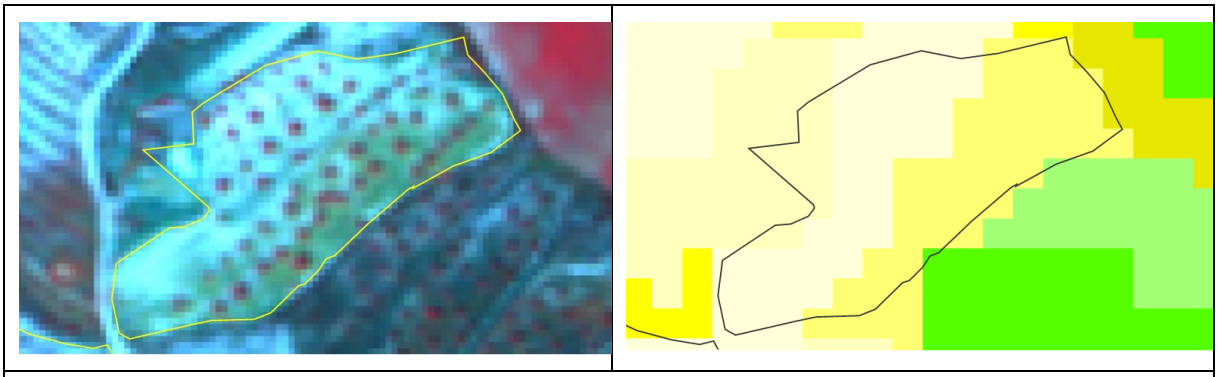




False detection of presence of different land cover within the FOI (black contour). The terraces that are inherently part of the eligible areal of the agricultural parcel visible on VHR imagery (left picture), were captured as different segments with notable class difference (right picture)



Commission error due to inherent variability typical for permanent crop (vineyard)



Commission error due to "Noise" from parcel border

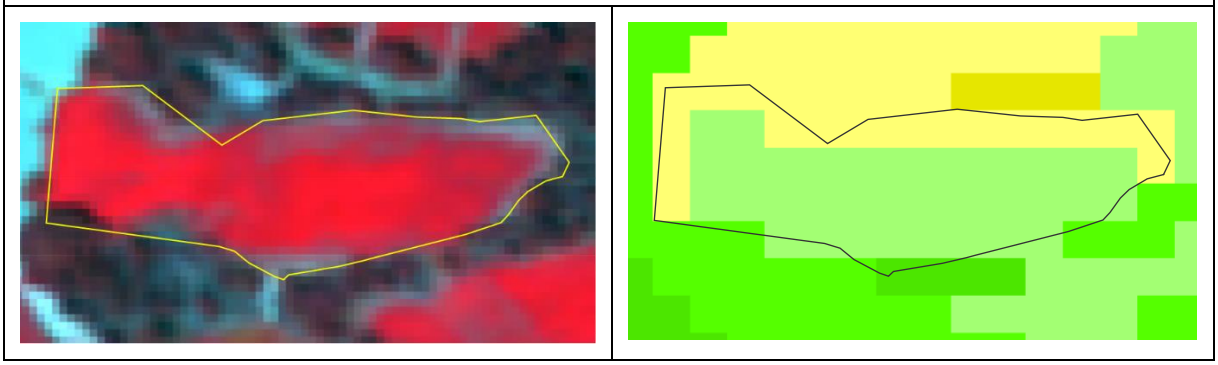


Figure 14: Example of correct and incorrect detection of FOI cardinality problem,

2.4 Multi-temporal S2 supervised classification

2.4.1 Description of the method

This is another clustering method, which in contrast with the previous one, uses supervised machine-learning to assess the FOI conditions. The underlying idea is to check whether there is more than one type of land cover/land use found within the FOI. This is achieved through an automatic pixel-based supervised classification of multi-temporal Sentinel-2 data, generating for each FOI, the types of land cover/land use and the percentage of each of type encountered inside. The quality of the method depends on the accuracy of the supervised classification; notably, the method used for classification and the quality of the training data used. For the purpose of the design and testing, the in-situ data was the Land Parcel Identification System from 2019 (Cultius) over the area of interest, combined with the annual land use data from 2019 GSAA, provided by the Catalonian Paying Agency).

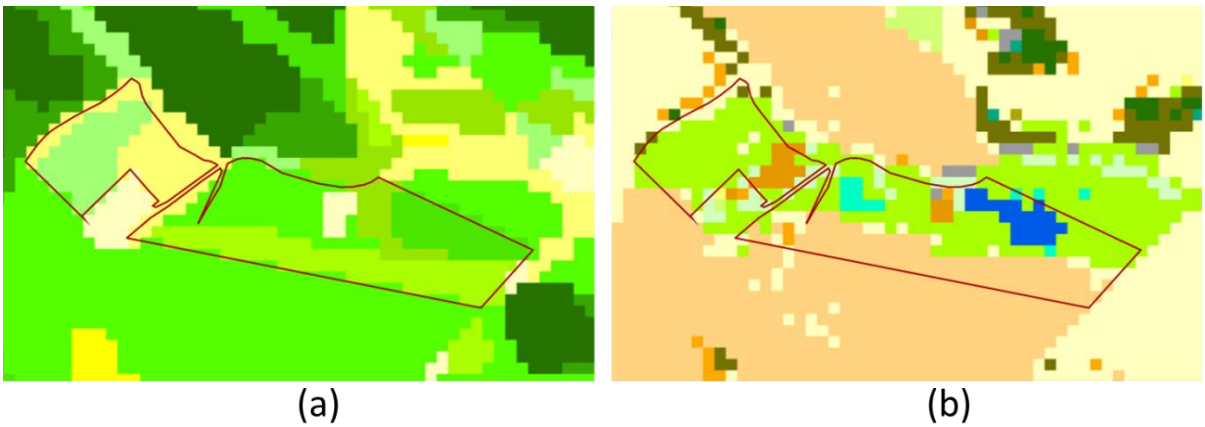


Figure 15: Comparison of the outcomes of the unsupervised and supervised clustering method over the same FOI.(a) unsupervised segmentation: colours show the time-related dynamics in the brightness, (b) supervised classification: colours show the class assigned, based on the pre-defined nomenclature (for ex. green colour stands for set-aside)

The supervised classification method is based on the open source EO-Learn library, developed by Synergise and available on the Sentinel Hub. The workflow implemented is based on the EO-Learn official examples:

- https://eo-learn.readthedocs.io/en/latest/examples/land-cover-map/SI_LULC_pipeline.html#Part-1

- https://eo-learn.readthedocs.io/en/latest/examples/land-cover-map/SI_LULC_pipeline.html#Part-2

The first step is the extraction of the Sentinel-2 images (entire scenes). The initial version of the method extracted the S2 data from Sentinel Hub. Later, the "image harvest" was established through the JRC DIAS Hub. The EO-Learn library uses "EO-Patches", where the Sentinel data is transformed into numpy arrays, stored on the local storage. All the calculations are then made using the data from the EO-Patches. Cloud masking is applied on the Sentinel data using Sentinel Hub's cloud detector. The method required the calculation of additional metrics, using

the existing Sentinel data, such as NDVI, NDWI, Euclidian Normalization, which were included in the EO-Patches.

The vector polygons representing the FOI data from the GSAA were converted to a raster mask, which was added to the EO-Patch. The GSAA data has been pre-processed: correcting some topological errors (parcel overlaps) from the initial data and grouping some of the land use categories (for LPIS data as well), according to pre-defined rules, in order to obtain more consistent training data for the machine learning algorithm and also to be in line with the scope of the steps/tiers defined for the FOI heterogeneity analysis. The Sentinel 2 data for the supervised classification was Level 2A and bands B02, B03, B04, B08, B11, B12 were used. The 20 meter bands were resampled to 10 meters, using nearest neighbourhood method.

UP_SP	NOME_SIGPAC	LandCover_SIGPAC	PROD_NOM_GSAA	Land_Use_GSAA	Land_Cover GSAA
AG	CORRIENTES Y SUPERFICIES DE AGUA	WATER	IMPRODUCTIUS	NON-PRODUCTIVE	NON-PRODUCTIVE
CA	VIALES	ROADS	IMPRODUCTIUS	NON-PRODUCTIVE	NON-PRODUCTIVE
FL	FRUTOS SECOS Y OLIVAR	PERMANENT TREE CROP	OLIVERES	OLIVE TREES	PERMANENT CROP
FO	FORESTAL	FOREST AREAS	ALTRES SUP. FORESTALS	AGRO_FORESTRY	AGRO_FORESTRY
FS	FRUTOS SECOS	PERMANENT TREE CROP	NOGUERES	WALNUT TREES	PERMANENT CROP
FY	FRUTALES	PERMANENT TREE CROP	PERERES	PEAR TREES	PERMANENT CROP
IM	IMPRODUCTIVOS	NON-PRODUCTIVE	IMPRODUCTIUS	NON-PRODUCTIVE	NON-PRODUCTIVE
OF	OLIVAR - FRUTAL	PERMANENT TREE CROP	OLIVERES	OLIVE TREES	PERMANENT CROP
OV	OLIVAR	PERMANENT TREE CROP	OLIVERES	OLIVE TREES	PERMANENT CROP
PA	PASTO CON ARBOLADO	WOODY GRASSLAND	ALTRES SUP. FORESTALS	AGRO_FORESTRY	AGRO_FORESTRY
PR	PASTO ARBUSTIVO	GRASSLAND WITH SCRUB	ALTRES SUP. FORESTALS	AGRO_FORESTRY	AGRO_FORESTRY
PS	PASTIZAL	NATURAL GRASSLAND	ALTRES SUP. FORESTALS	AGRO_FORESTRY	AGRO_FORESTRY
TA	TIERRAS ARABLES	ARABLE LAND	IMPRODUCTIUS	NON-PRODUCTIVE	NON-PRODUCTIVE
TH	HUERTA	ARABLE LAND	ORDI	BARLEY	ARABLE LAND
VF	VIÑEDO - FRUTAL	PERMANENT SHRUB CROP	VINYES	VINEYARDS	PERMANENT CROP
VI	VIÑEDO	PERMANENT SHRUB CROP	VINYES	VINEYARDS	PERMANENT CROP
ZU	ZONA URBANA	ARTICIAL SURFACES AND ASSOT	IMPRODUCTIUS	NON-PRODUCTIVE	NON-PRODUCTIVE

Figure 16: Grouping of some of the land cover/land use categories for the purpose of classification

The next steps to follow, were related to some additional data preparation required for the machine learning process: (1) removal of cloudy scenes (keeping only the ones with >80% valid coverage); (2) concatenation of the S2 bands, NDVI, NDWI, NORM in a single table; (3) conduction of temporal interpolation; (4) creation of task for linear interpolation in the temporal dimension; (5) provisions of the cloud mask for the interpolation functions, (6) raster data cleaning (removing the artefacts with a width of 1 pixel and removing the edges between polygons of different classes); (7), splitting input data patches in training/validation sets.

The compiled and cleaned dataset was used to construct and train the selected machine-learning model, which was LightGMB (Light Gradient Boosting Machine). The accuracy of the predicted classed was checked using the validating EO-Patch selected. The predicted classes were exported as TIFF files and further combined in a mosaic.

Using the raster file containing the predicted classes, the method performs a count of the pixels inside each polygon from the initial GSAA data (basis for the FOI), grouped by class. It flags those FOIs having pixels from different classes, which comply with certain pre-defined rules on the pixel percentage. There are two versions: with and without application of negative buffer of 10 meters.

2.4.2 Meaning of the output produced

The method has two distinct parts:

(1) the supervised classification, based on EO-learn openly available notebook and

(2) the pixel and cluster assessment part

The latter has two versions. In the first version, the pixels and clusters of pixels are transformed into vectors (polygons). The resulting vectors are loaded into a database and further processing and analysis is performed through stored procedures developed in this purpose. The analysis is made based on the area of the vectors derived from the clusters of pixels.

In the second version the determination of the percentage of pixels from each class and their grouping in clusters is made through analysis performed at the level of the thematic raster (using rasterstats and opencv). All the pixels whose centroids are located inside the given FOI are taken into account. The analysis is done directly at the level of pixels and correspondent clusters and not at the level of the area represented by these pixels or clusters. There are no significant differences between these two methods. In the first method, the pixels located at the FOI limit are also taken into account (with their respective overlapping area). After clipping the vector resulted from polygonising the pixels at the edge of the FOI, the area located inside the FOI is used to determine the area of the cluster. In the second version, if the centroid of a pixel is not located inside the FOI, this pixel will not be used to determine the cluster size for the analysed FOI. The period defined for the multi-temporal stack was the month of May.

The FOI assessment part produces two types of outcomes – one for FOI spatial variability (heterogeneity) and one for FOI spatial cardinality.

Once a FOI is found to contain pixels belonging to more than one class, and one of the classes accounts for a percentage between given thresholds (the current values are between 30% and 70%), this FOI is flagged as heterogeneous. This means that there is no single class present in the given FOI polygon. All pixels are counted, regardless their position and spatial arrangement inside the FOI.

The 30-70 percent threshold was chosen in order to limit the influence of the neighbouring elements on the FOI, such as: roads, buildings, artificial sealed areas, forest and others. The selected percentage is arbitrary and could be revised in the light of future studies.

Table 4: Output of the S2 supervised classification method (FOI heterogeneity)

OBJECTID_	Area (ha)	LU_type	Class1	P_Class1	Class2	P_Class2	Class3	P_Class3	Class4	P_Class4	foi_h
1289290	1.639	SOFT WHEAT	112	67.47	54	32.53	0	0	0	0	1
1299239	0.706	SOFT WHEAT	47	64.38	25	34.25	0	0	0	0	1
1323312	0.977	BARLEY	0	0	96	97.96	0	0	0	0	0
1780682	1.206	SOFT WHEAT	68	53.97	0	0	43	34.13	0	0	1
1792890	0.810	SPINACH	63	78.75	13	16.25	0	0	1	1.25	0
1833824	0.442	BARLEY	0	0	0	0	0	0	41	97.62	0
1926450	0.454	SOFT WHEAT	0	0	0	0	46	100	0	0	0

2145658	1.665	CORN	0	0	141	83.93	0	0	0	0	0
2236861	0.597	EFA ALFALFA	33	56.9	0	0	0	0	14	24.14	1
2351360	0.441	CORN	0	0	2	4.76	0	0	40	95.24	0

Note: not all classes are shown.

The main output parameters are as follows:

Class [i]	i-th class in the applied nomenclature
P_class [i]	Percentage of pixels from class[i] in the FOI
Foi_h	Decision on homogeneity/heterogeneity, i.e.0/1

Once a FOI is found to contain clusters of contiguous pixels belonging to more than one class, and at least two of these clusters belonging to different classes are more than 20 pixels each (one Sentinel-2 pixel being 10mx10m), it is flagged as being with cardinality problem. The "20-pixel rule" is explained in a separate chapter below.

Table 5: Output of the S2 supervised classification method (FOI cardinality)

OBJECTID_	Area (ha)	LU_type	Clusters	foi_c
1289290	1.639	SOFT WHEAT	(1, 106), (2, 66)	1
1299239	0.706	SOFT WHEAT	(1, 49)	0
1323312	0.977	BARLEY	(2, 112)	0
1780682	1.206	SOFT WHEAT	(1, 68), (3, 56)	1
1792890	0.810	SPINACH	(1, 61)	0
1833824	0.442	BARLEY	(4, 48)	0
1926450	0.454	SOFT WHEAT	(3, 41)	0
2145658	1.665	CORN	(2, 128), (6, 28), (7, 24)	1
2236861	0.597	EFA ALFALFA	(1, 33)	0
2351360	0.441	CORN	(4, 43)	0

The main output parameters are as follows:

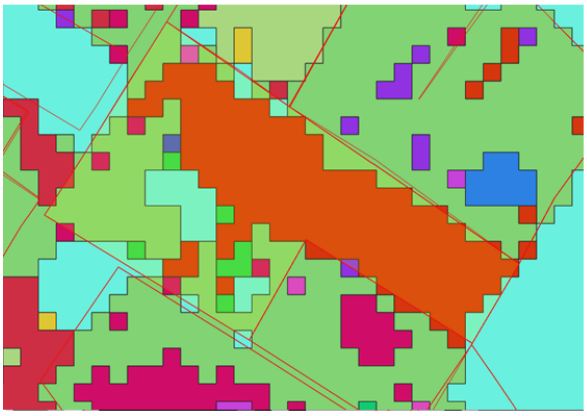
Clusters	(class [i], number of pixels of class [i]), (class [j], number of pixels of class [j]),….
Foi_c	Decision on cardinality issue (not present/present, i.e.0/1)

2.4.3 Calibration and validation

Several trials were made with different training data involved. A problem found with the "standard" training dataset, based on a combination of the 2019 LPIS and GSAA, was that it covered extensively the area and typology related to agricultural land cover/land use, but provided scarce information on location and typology of non-agricultural areas. This resulted in an over-representation of the agricultural classes in the nomenclature as well. As a consequence, the objects of non-agricultural nature ended as being classified as agricultural, with the closest available class, although with very low degree of probability. It was difficult to know whether the given object is actually of non-agricultural nature, or if it is simply a specific agricultural type, with low representativeness in the training data.

To overcome this problem, the 2019 training LPIS/GSAA was complemented with those non-agricultural land cover types from the 2017 Catalonian land cover dataset (UsosCobertes_2017), considered stable since still observable in 2019. This resulted in two parallel approaches for assessing FOI though supervised machine learning based on EO-learn library: (1) using the 2019 GSSA only as training data, and (2) using 2019 GSAA together with non-agricultural areas from the 2017 land cover dataset of Catalonia. The 2017 non-agricultural land cover classes were grouped in three generic categories: water, artificial areas and roads. FOIs where the main land cover type found occupies between 30-70% of the pixels enclosed, are flagged as being heterogeneous and having a potential cardinality problem. The initial results from the supervised classification were in favour of the approached based on the combination of 2017 and 2019 training data. There was, however, high number of cases where agricultural areas were wrongly classified, as non-agricultural. Around 1000 FOIs were found, as having more than 10% of non-agricultural land cover within, which was unrealistic, considering the existing knowledge on the LPIS/GSAA quality in Catalonia. An assessment was made against the reference data (with its four FOI cardinality types defined), as well. The "30%-70%"class abundance threshold defined seemed technically sound and correct; it is however the output of the classification that needed further improvement. The poor accuracy in some cases could be explained by: (1) lack of immediate post-processing of the produced thematic raster file; (2) the relatively small size of training area, not sufficient for the learning process; and (3) the geometric misalignments between the S2A and S2B acquisitions.

The complexity of the Catalonian landscape and the typology of land cover/ use played a role as well. Some of the land cover types, such as permanent crop, requires the consideration of specific context-related factors, which are easy to account with human visual interpretation, but difficult to formalize in machine readable manner. For example, for the fruit trees FOIs, the classifier is generating some incorrect results due to high variations of the intra-row space - a pattern that is easily recognisable on the individual Sentinel 2 images (as being inherently part of the permanent crop class.)



(a)

Identify Results

Feature	Value
pixel_count_be_fl_17classes	[3]
fid	17407
(Derived)	
(Actions)	
fid	17407
CODE_OBJ	0000280A67538B05
GWSCOD_H	8546
GWSNAM_H	Savooikool - industrie
IN_GEBR_01	N
IN_GEBR_P	J
IN_GEBR_PH	N
area	22422.101345380040584
Class1	4
Class2	1
Class3	1
Class4	0
Class5	0
Class6	0
Class7	23
Class8	0
Class9	7
Class10	121
Class11	67
Class12	0
Class13	0
Class14	0
Class15	0
Class16	0
Class17	0

(b)

Figure 17: (a) Outcomes of the supervised classification; (b) pixel-based analysis within the FOI

In summary, the accuracy of the supervised classification was found as the key determining element for the success of this method of FOI assessment. It depends on many factors, such as: the in-situ data, Sentinel 2 scenes quality (cloud cover), methods used for data post-processing, classification algorithm used, etc. With all components set in optimal manner, this approach has the potential to provide the most comprehensive information on FOI conditions – not only the presence and location of different clusters, but also their exact nature in land cover/land use terms. Still, the critical success factor is the quality and completeness of the training data, still suffering from the lack of consistency of land cover and land use nomenclatures, which often prevent the correct definition of the correspondent training data labels (A Comber 2005). There were attempts made to improve the classification accuracy and to disambiguate the "arable land" category into more detailed individual crop groups, using also some ancillary crop-related datasets, as communicated by Spain during the IACS data sharing project with JRC.

Finally, there were some trials performed with the recently published Land Cover Map of Europe 2017 - a product resulting from the Phase 2 of the ESA S2GLC project. The initial verifications with the ESA S2GLS showed that the thematic accuracy seems comparable with the results from the other setups. However, the land cover nomenclature is different. At the end, it was decided that ESA S2GLS product would not bring substantial added value.

3 Building the reference

3.1 General considerations

As with any EO-based detection method, the most pertinent question in the FOI assessment project was how performant the detection methods are. The answer to that question was found rather challenging due to the "ill-defined" concepts of land cover and land use – concepts, which were at the hearth of the notions of spatial variability and spatial cardinality. The clarity of the definitions was highly dependent on the clarity of the semantics used. Terms as "bio-physical object" or "unit of management" needed further conceptualization and explanation. Although, the project introduced from the beginning some formal and rigorous definitions of FOI heterogeneity and cardinality, it became clear in the course of the project implementation that there are many context-specific variations in the perception of these phenomena. This required special care in the preparation of the quality assessment methodology, and associated reference data.

In any case, the work was done with the assumption that a given detection method will be considered as performant, if it could replicate in automated way, the counterpart visual photointerpretation process.

In this respect, the preparation of the reference data had two objectives:

(1) to analyse the feasibility of using Sentinel-2 itself, as reference data (in visual photointerpretation mode) for validation of the FOI assessment methods;

(2) to prepare a reference dataset using Sentinel-2 imagery, over a selected area of interest. The derived information should represent as close as possible the situation on the field. Whenever the information content from Sentinel-2 is found not sufficient to depict the field conditions, ancillary data with higher spatial resolution (VHR imagery, national orthophoto) will be used.

Concerning the field conditions the photo-interpretation had to depict, the general agreement was that they should reflect two cases:

- FOI heterogeneity: (1) Inherent variations of given characteristics of the physical entity, represented by the GSAA parcel or aggregate or (2) Different physical entities present in the same unit of management, represented by the GSAA parcel or aggregate. This was done to reflect mostly Step 1 and Step 2 from the tiered approach.
- FOI cardinality: (3) Different physical entities that are sufficiently big to challenge the area component present in the same unit of management, represented by the GSAA parcel or aggregate. Often these entities relate to different units of management. This was done to reflect Step 3.

3.2 Design of the reference data

The area of interest selected to testing the methods, was the 2019 CwRS zone NOUR, located in Catalonia, Spain (Figure 19). The reason for such selection was the interest for collaboration expressed by the Catalonian administration on the subject and the peculiarity of the landscape, offering the opportunity to test the methods in various conditions. The dataset provided consisted of 365 randomly selected FOIs from the 2019 GSAA. Parcels with area below 0.2 ha were not included, as assumed not always monitorable with Sentinel. The dataset was topologically corrected, while keeping track on the number of parcels affected (few

parcels with permanent crops). For the assessment of the FOI conditions, a negative buffer of 5 m (i.e. half of the S2 images pixel size) was introduced, in order to avoid photo-interpretation errors due to influence of neighbouring pixels at FOI boundaries.

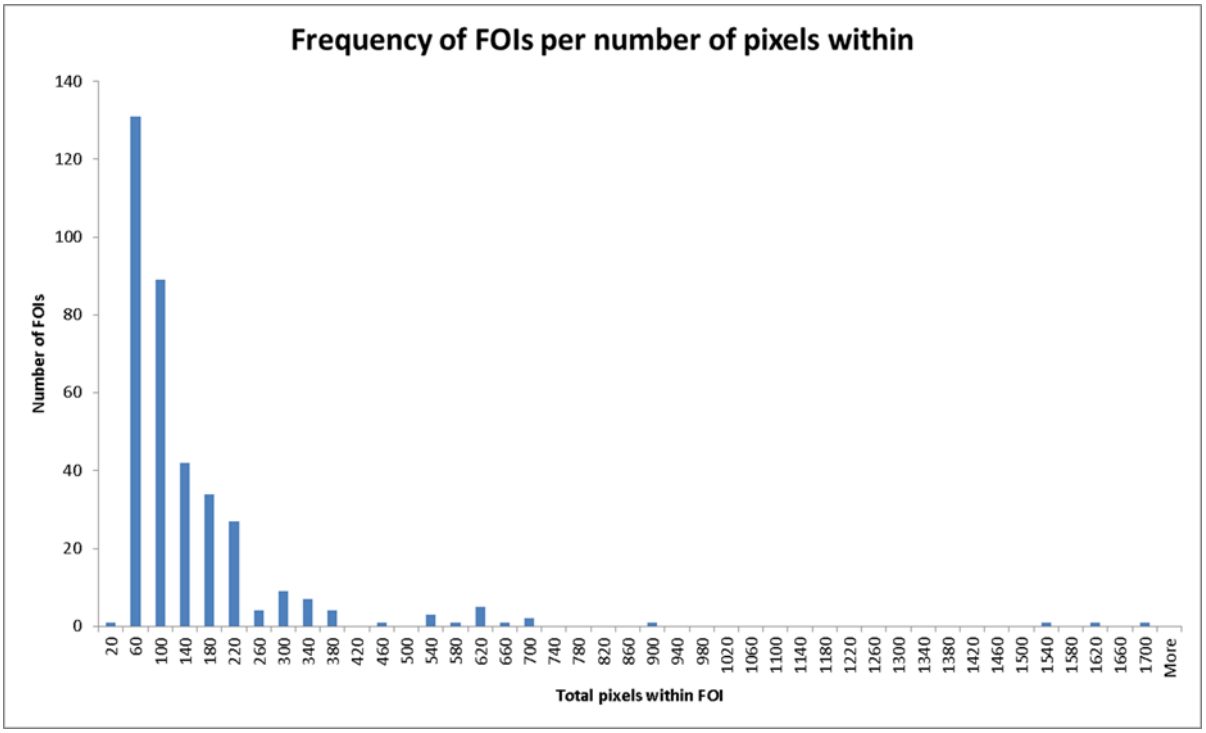


Figure 18: Histogram of the distribution of the FOIs in the reference dataset, according to the number of Sentinel pixels inside

The visual assessment of the FOI conditions (for cardinality and heterogeneity) had to take into account the temporal behaviour of key characteristics of the vegetation, such as plant phenology, being specific for each type of crop or land use. The degree of uniformity of vegetation cover was evaluated at the time of the maximum development of green leaf coverage. The presence of persistent spatial variability was confirmed when detected on a number of consecutive S2 observations; a number depended on the time span of the given biological (phenological) cycle and the dynamics of the temporal variability. Following the Nyquist–Shannon sampling theorem, the minimum density (frequency) of the S2 observations to come to a conclusion for presence of spatial variability was set to be at least twice higher than the expected temporal frequency of the phenomenon; 3-4 images a month (arable crops) and 4-5 images per season (permanent crops and permanent grasslands). Variations of the pixel colours observed over a small period of time (i.e. for 1 or 2 consecutive images) were not considered as indicative for the presence of persistent spatial variability.

When the Sentinel-2 data was not sufficient to make a decision (data gaps present or information content is poor), the photo-interpreter could use the provided VHR image from the CwRS campaign. They also turned to be very useful especially when the heterogeneity was caused by the presence of unaccounted permanent non-agricultural features (roads, houses, forest).

The photointerpretation had to account not only for the presence of different physical entities within the FOI, but also for their spatial arrangement. For example, permanent crop fields are constituted of a number of rows with woody

crops, alternated with rows of grass or bare soil, or even arable crop. Although, being "different objects", they create an intrinsic mix and pattern that is regular and uniform. This spatial pattern also reflects a particular type of agronomic activity, occurring on single unit of management and in line with particular land use scenario. In this case, the "heterogeneity" is considered as inherent characteristic of the permanent crops.

The elaboration of the reference dataset was rather time consuming (10-15 minutes per parcel), due to the need to report in great detail the S2 data used as reference and the associated observations (details on the structure of the reference data and the inspection protocol is provided in an Annex II). However, the collected supplementary information provides important and indispensable insight on the nature of the FOI conditions and a valuable input for the technical settings of the detection methods assessed.

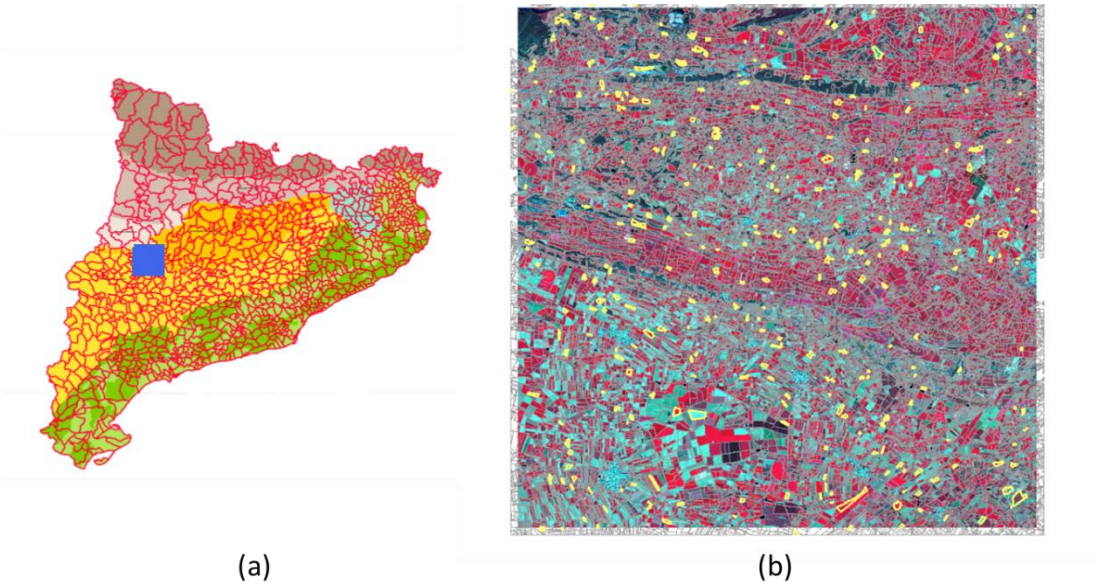


Figure 19: (a) Map with the administrative units in Catalonia with the area of interest shown in blue; (b) The area of interest (2019 CwRS of NOUR) with the sampled reference FOIs from the GSAA, shown in yellow

The photointerpretation process allowed identifying several main types of FOI cardinality-related observations:

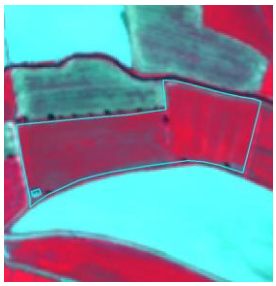
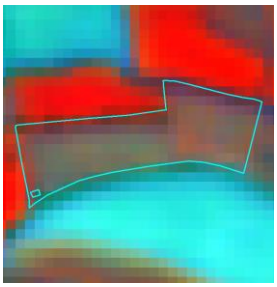
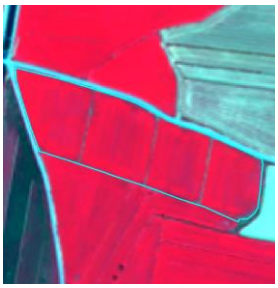
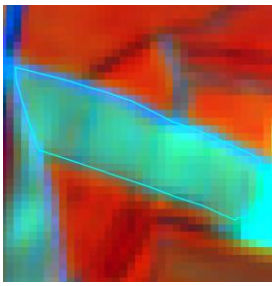
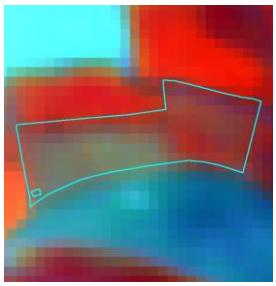
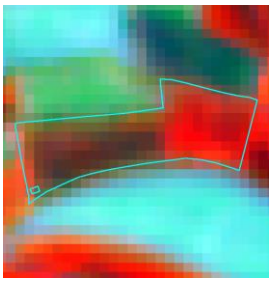
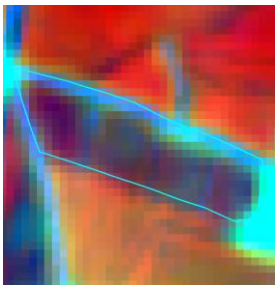
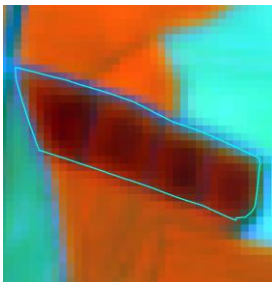
(1) FOI representation from GSAA contains more than one physical entity representing distinct units of agricultural management; each of them behaving differently in time. This is a critical case, where the area component is compromised (at least with respect to certain schemes) and CbM would not be able to yield meaningful results for the majority of scenarios. This case requires an immediate interaction with the farmer to ask for an updating of the GSAA.

(2) FOI representation from GSAA contains in principle more than one physical entity, but they all act as one single units of agricultural management, since they behave the same way (would be common case for permanent grassland or for permanent crop with internal roads crossing the entire FOI, often too small to be accounted in the GSAA/LPIS). Although, the CbM process would probably retrieve meaningful information, the area component might not be entirely correct (or at least "contaminated"). Interaction with farmer might be needed. There might be issues with LPIS.

(3) The geometry of FOI representation from GSAA alone indicates potential presence of more than one physical entity within. In the majority of the cases, these geometries are complex to a degree that it would be difficult for the Sentinel signal to retrieve meaningful information. Interaction with farmer might be needed. There might be issues with LPIS.

(4) FOI representation from GSAA contains several physical entities, but only one unit of agricultural management; this is a critical case when significant part of the FOI representation from GSAA is occupied with non-agricultural area or non-eligible area for the given scheme (thus, the initial established area in the GSAA is probably not correct). Interaction with farmer is required. There is an issue with LPIS.

(5) There are no cardinality issues between GSAA/CbM derived FOI representation, but very apparent heterogeneity within the FOI (anisotropic or isotropic) caused by numerous linear features (margins, terraces, hedges) or specific soil conditions that are inherently part of the physical entity.

<table>
<tr><td colspan="2"><u>Case 1:</u> Arable land, barley
more than one physical entity
more than one unit of management (different crops/ land use)</td><td colspan="2"><u>Case 2</u>: Arable land, barley
more than one physical entity one unit of management, subdivided internally (same crops/ land use)</td></tr>
<tr><td> 
VHR MS / 04/05/2019</td><td> 
S2 / 22 March 2019</td><td> 
VHR MS / 04/05/2019</td><td> 
S2 / 22 March 2019</td></tr>
<tr><td> 
S2 / 14 April 2019</td><td> 
S2 / 1 May 2019</td><td> 
S2 / 14 April 2019</td><td> 
S2 / 1 May 2019</td></tr>
<tr><td colspan="2"><u>Case 3:</u> Arable land, barley
more than one physical entity
one unit of management with complex geometry (same crops/ land use)</td><td colspan="2"><u>Case 4:</u> Arable land, EFA Land Laying Fallow
more than one physical entity
one of more entities of non-agricultural nature</td></tr>
</table>

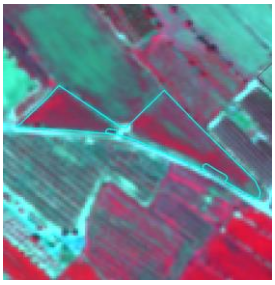
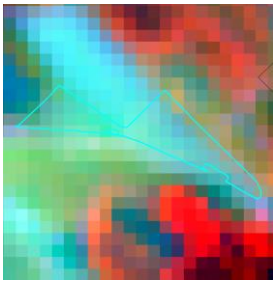
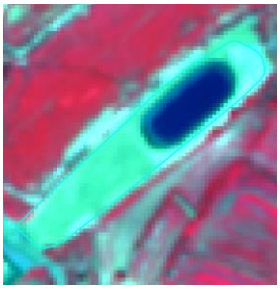
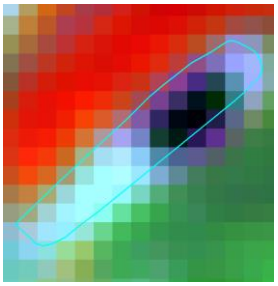
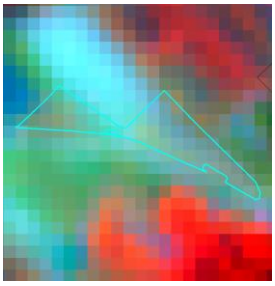
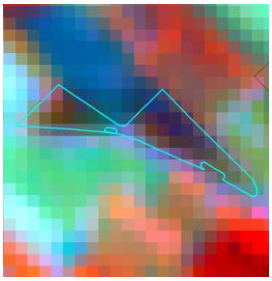
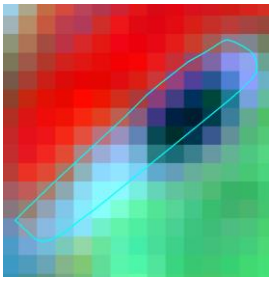
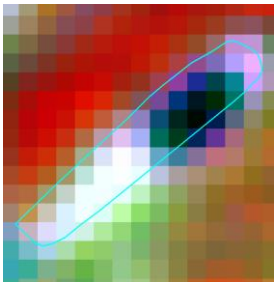
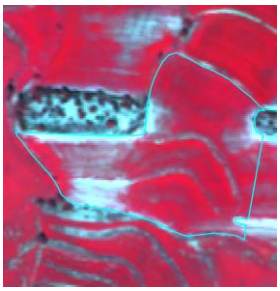
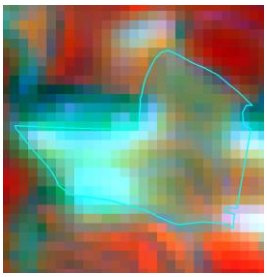
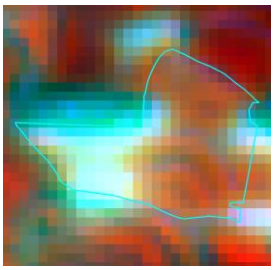
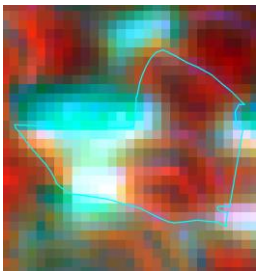
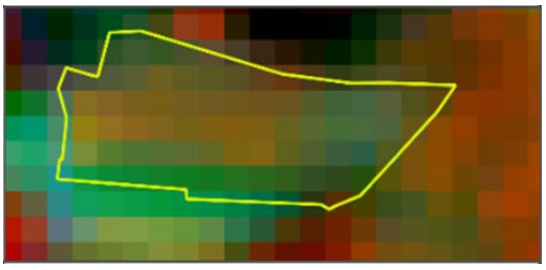
<table>
<tr><td> 
VHR MS / 04/05/2019</td><td> 
S2 / 22 March 2019</td><td> 
VHR MS / 04/05/2019</td><td> 
S2 / 22 March 2019</td></tr>
<tr><td> 
S2 / 14 April 2019</td><td> 
S2 / 1 May 2019</td><td> 
S2 / 14 April 2019</td><td> 
S2 / 1 May 2019</td></tr>
<tr><td colspan="2"><u>Case 5-1:</u> Arable land, barley
more than one physical entity
one unit of management on complex landscape and/or specific soil conditions (same crops/ land use)</td><td colspan="2"><u>Case 5-2:</u> Arable land, EFA Land Laying Fallow
more than one physical entity
one unit of management on complex landscape and/or specific soil conditions (same crops/ land use)</td></tr>
<tr><td> 
VHR MS / 04/05/2019</td><td> 
S2 / 22 March 2019</td><td colspan="2"> 
VHR MS / 04/05/2019</td></tr>
<tr><td> 
S2 / 14 April 2019</td><td> 
S2 / 1 May 2019</td><td colspan="2"> 
S2 / 14 April 2019</td></tr>
</table>

Figure 20: Typology of the FOI "integrity problem" found

3.3 Methodological and conceptual shift

The observed complexity of the spatial variability and cardinality called for a breakdown of the research problem into smaller parts that are easier to address in isolation. As a consequence, the project team decided to focus the work on those aspects of the FOI assessment methods that could address the validity of the FOI used in the CbM, by checking whether the correspondence and 1-1 cardinality between the GSAA and the unit of management is respected. This was considered the most crucial and generic check, applicable irrespective of the scenario defined. It logically implied a shift of the FOI cardinality assessment methods toward effective detection of cardinality cases 1 and 4 only, since they are related to problems with the correctness of eligible area, and with the validity of the CbM boundary conditions. Also, such cases might trigger the need for area measurement later in the CbM process. Case 2 was considered not actually a FOI cardinality problem in the given year, although its development could be monitored anyhow. Case 3 could be handled during the preparation and submission of the GSAA or through a better design of the LPIS reference parcels. Case 5 is not a cardinality issue, since it refers to an inherent and intrinsic land cover mix, expected for a given scenario.

In any case, the developed reference dataset reflected and reported thoroughly the FOI conditions with respect to heterogeneity and all cases of spatial cardinality, to allow for more in-depth analysis, in later stage.

4 Dealing with the area

The discussion with AGRI (D3 and H3 Units) and the outcomes of the GREXE meeting held on 16.06.2020, re-confirmed the need to focus the FOI assessment methods in detecting those cases where the area component (provided by the GSAA/LPIS) seems compromised and, as consequence, the CbM boundary conditions are not respected. Timely assessment of area dimension is also of key importance for some CbM scenarios, such as the ones related to greening payment scheme. For example, FOIs checked for crop diversification and flagged as yellow during CbM, could be found as having area problem during the follow-up late in the season. This would often lead to a measurement on the spot which would be against the CbM logic. So, it is important to have an indication from the FOI validations tests (G1 type) whether the problem detection is linked to the area component in order to allow timely communication to the farmer to confirm and rectify the problem.

It became clear that whichever FOI assessment method, or combination of them, is used, it should provide information on the size of the different clusters found within the FOI. In other words, what is the minimum size (in number of pixels) of the detected "alien" objects within FOI, which should trigger further analysis on the FOI area? It should be pointed out that this further analysis does not necessarily imply area measurement.

The image clustering tests, conducted in-house, showed that the information content of Sentinel -2 allows a detection of individual physical entities in systematic manner, when these entities are represented by clusters of at least 20 contiguous pixels (20pixels x 10m x10 m = 2000 m2). Occasionally, Sentinel data could be able detect smaller objects represented by 10 pixels in a cluster (1000 m2); however, 75% of these detected clusters would still be attributes to noise (for example, border slivers). The nature of the physical object and its shape further determines the numbers of pixels required in both directions. Compact and rectangular objects would be represented by roughly equal amount of pixels in both directions (i.e. 4x5, 5x6 pixels), while disperse and elongated objects would be represented by twice more pixels in one of the directions (i.e. 3x10, 2x14 pixels).

The results obtained were found in line with the general experience on Computer-Assisted Photointerpretation (CAPI), collected over the years. As reported in the JRC Technical Guidance for Management of Layer in LPIS (JRC 2015), to detect the presence of an object during CAPI, it must be at least 3 times the pixel size in both directions (Figure 21).

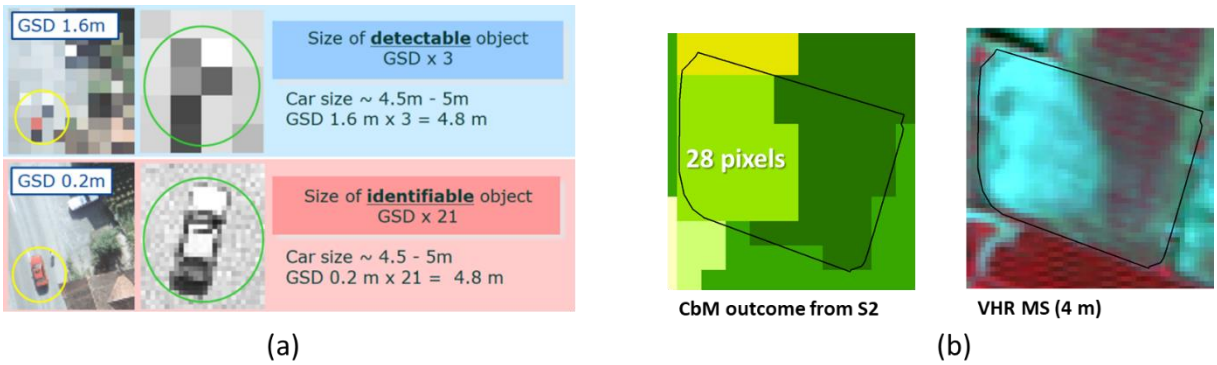


Figure 21: (a) sizes of detectable and identifiable objects, though CAPI; (b) Example of outcome from the in-house clustering tests

Considering the spatial resolution of Sentinel-2, the minimum size of an object to be detected would be 3x3x10x10 = 900 m2. When multiplied by 2 to account for the shape, the minimum size will become 2 x 900 = 1800 m2. This value is close to 2000 m2, which is the equivalent size of 20 contiguous Sentinel-2 pixels.

Thus, the "20-pixel rule" was accepted in the CbM for practical purposes to indicate the size of the detectable object above which we can be almost certain that there is need to have a further follow-up on the given FOI with respect to its validity (visual check, contacting famer).

It is obvious that the minimum size of depicted cluster will be sensor dependent. While it will be 2000 m2 for Sentinel, it will be less for sensors with higher spatial resolution. For example, for PlanetScope data, the threshold will be 2x(3x3x3mx3m) = 200 m2. Further tests on HHR (for example, Planet) data could be performed, to confirm the 200 m2 minimum object size derived from the "20 pixels in cluster" rule, for sensors operating at 5 meter and higher spatial resolution.

5 Assessment and results

5.1 Quality assessment setup

The setup for the evaluation the results from the FOI assessment methods was based on conceptual elements, some of which were later incorporated in the Technical Guidance on Quality assessment of the CbM (v.1.0)(Luketić N. 2020). The final version of the reference dataset consisted of 365 randomly selected FOI representations from 2019 GSAA, visually checked (with optical Sentinel data) for presence of spatial variability and issues with spatial cardinality (with special focus on cases 1 and 4) over a specific time period, which was different depending on the land cover/land use defined in the GSAA. The analysis of each FOI was based on visual interpretation of Sentinel-2 data, which comprised a graphical representation of the NDVI time series and a calendar view (structured array) of imagettes, over the defined period. Sentinel-2 signal was extracted from the JRC DIAS Hub and was displayed through a specially designed portrayal (Figure 22). In some cases, the VHR imagery from the 2019 CwRS zone was used.

The visual photointerpretation was made without taking into account the results of the methods and could be considered completely "blind" (not biased by their outcomes). However, during a 4-eye control phase, some subsequent plausibility checks in line with FOI detection outcomes were made. They proved to be useful for improvement of the methods and fine-tuning the quality assessment (QA) inspection process. Negative buffer was introduced to account for frequent bordering effects induced by neighbouring pixels. Photointerpretation was performed only on basis of the information provided by the Sentinel-2/VHR observations; no local expert knowledge was used.

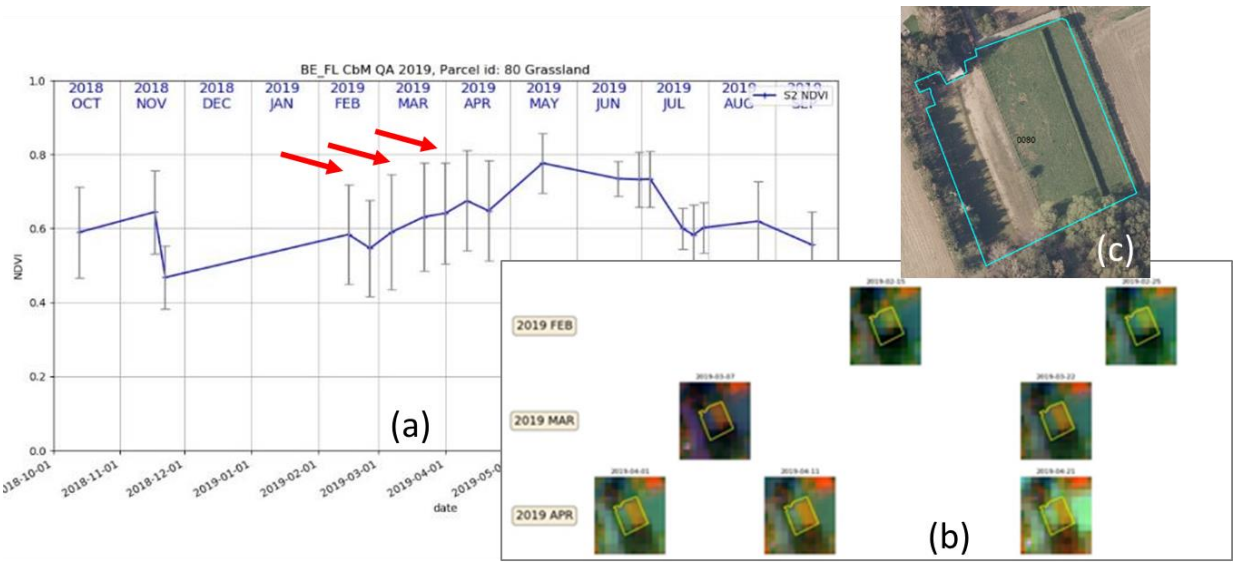


Figure 22: Example of portrayal of S2 signal for a given FOI. (a) Graph of the mean NDVI values of the FOI, with error bar (vertical lines) showing the standard deviation being persistently high (red arrows); (b) Calendar view of S2 imagettes showing the persistent presence of two management units within the FOI (confirmed also on the VHR image, image subset c)

To qualify an FOI as having cardinality problem, the operator is searching for persistent presence of different and well visible physical entities/objects within FOI, on the calendar view of imagettes. These entities should form clusters with

at least 20 pixels. He/she further checks the NDVI variance within FOI to confirm/reject the observations. The operators used an interactive environment, provided by a specially designed notebook in JRC DIAS Hib, to explore the object clusters when occurring within small-size FOIs. There were few cases of small FOIs (bellow 0.2ha), which were skipped, as not being relevant for the study, and replaced with others. A great variety of cases of spatial heterogeneity was observed over the Catalonian test site.

There were some revisions made with respect to the role of the VHR imagery in the visual photo-interpretation. In the initial version of the reference dataset, the FOI is flagged as having cardinality problem (FOI_Cardin=1, see the Annex II at the end), when such problem is observed either on S2 or on VHR independently. Actually, the VHR imagery should be used only to confirm a problem already observed on S2 (if a problem is observed only on VHR, then the detection failure would be due to the limitation of Sentinel, not to the deficiencies of the detection method). Thus, FOI should be flagged as having cardinality problem (FOI_Cardin=1), only if such problem is either observed and confirmed on S2 alone, or observed on S2 and further confirmed on VHR image. A new version of the reference data was made, using the later approach resulting in 20 FOIs being changed from FOI_Cardin=1 to FOI_Cardin=0, and one FOI changed in the opposite direction.

With respect to the FOI integrity issues (spatial cardinality cases 1 and 4), there were 3 different versions of the reference dataset produced:

- Version 1: FOI potential integrity problem found on S2 or on VHR independently (S2 OR VHR). Different objects found, regardless the size of the corresponding clusters - 92 out of 365 cases found
- Version 2: FOI potential integrity problem found on S2 alone and further confirmed on VHR (S2 AND VHR). Different objects found, regardless the size of the corresponding clusters - 72 out of 365 cases found
- Version 3: FOI potential integrity problem found on S2 alone and further confirmed on VHR. Different objects found with correspondent clusters of at least 20 S2 pixels - 13 out of 365 cases found

It is worth noting this reference dataset was meant to assess the detection capability of the FOI assessment methods, and not to assess the quality of the GSAA in Catalonia. From the point of view of LPIS/GSAA eligibility, only few of all the potential cases might have a real impact on the area component. The outcomes in the reference data were Boolean (1 – problem present; 2 – problem not present) in order to align them with the outcomes of the FOI assessment methods, which were also Boolean.

The assessment of the detection performance of the methods against reference dataset was made using the classical confusion matrix approach and kappa coefficient. The project team introduced a quality metric for checking the user accuracy, based on limiting quality (LQ) levels following ISO 2859/2 (used also in quality assessment of the CbM). The LQ level for commission errors was set to 10%, while for omission errors – 5%.

The final ranking of the FOI methods tried to account also for their deployment complexity and early–warning capabilities, in the sense of how early in the season

decision can be taken (Figure 23). There was a striving for balance between the omission and commissions errors, but with preferences towards reducing omission errors (truly present FOI problem not detected). The findings of the performance assessment are expected to contribute to the fine-tuning of the CbM QA method itself.

FOI methods	Suitable for		
	Capturing different things (heterogeneity)	Capturing different physical objects (cardinality)	Capturing different physical objects above given area (area congruency)
S1 Gamma	X	?	-
S2 S/N	X	?	-
S2 percentile	X	X	?
S2 segmentation	?	X	X
S2 supervised ML	X	X	X

Figure 23: Initial assumptions on expectations from the different FOI assessment methods

5.2 First results

5.2.1 Per method

Results of the methods were tested with two versions of the reference data. Version 2 was used to verify the outcomes both for all 4 FOI cardinality cases and for cases 1 and 4 only. Version 3 was used to verify the outcomes solely for cases 1 and 4.

5.2.1.1 Analysis of S1 backscattering's speckle noise

As described in point 2.1, the method indicates potential FOI integrity problem when the difference between observed variance and theoretical is statistically significant (alpha < 0.01). The scenario under test confirmed a potential integrity problem when found on more than 3 images within May. Negative buffer of 10 meters was applied. For some fields no results for backscatter extraction were derived, as they were too small. As explained above, such fields were excluded from the reference data.

Table 6: Results from the detection of S1 backscattering's speckle noise method

Ref Data: S2 AND VHR (V2)
All cardinality Types

	Detected	Not detected
True presence	31	41
True absence	94	194

Ref Data: S2 AND VHR (V2)
Types 1 and 4 only

	Detected	Not detected
True presence	16	17
True absence	109	218

Reference Data: 10 pix clusters (V3
Types 1 and 4 only

	Detected	Not detected
True presence	5	8
True absence	120	227

Results were not found promising, and there was not much room for improvement through the parameter settings, especially if the choice of the VH polarization as prime signal remains unchanged. Both commission and omission errors were above the defined LQ levels. The underlying reason could be that the method is based on simple descriptive statistical value (mean, StDev), which is then

compared to a threshold. However, these values assume a unimodal distribution of the pixel DN values within the FOI; an initial hypothesis, not valid for the problematic FOIs the method is trying to detect[2]. Alternatively, the assessment could focus on checking the modality of distribution of pixel values, such as the DIP test (Hartigan March, 1985). FOIs with bimodality or multimodality have high probability of having a problem with 1-1 cardinality.

In any case, the low accuracy of the S1 Gamma method (kappa around 0.06 for cases 1 and 4 on AL) has been significantly rectified when applying a negative buffer to the FOI perimeter. Tests with a negative buffer of 10 meters show notable reduction of commission (alpha) errors. However, the 10 meters seems too much for a number of FOIs (mostly Case 3 types), where the original single geometry is practically "broken down" in small separate polygons, which could not serve as a basis for meaningful data extraction. This obviously leads to an increase of the omission (beta) errors. Trials with smaller negative buffer value (for example, 5 meters) should to be performed in future.

5.2.1.2 Threshold on S2 signal-to-noise ratio (SNR)

As described in point 2.2, the method indicates potential FOI integrity problem when the ratio between the average NDVI and StDev NDVI is below 5. The scenario under test confirmed a potential integrity problem when found on at least 50% of the suitable images within May. A negative buffer of 10 meters was applied. For some fields no results for SNR extraction were derived, as they were too small. As explained above, such fields were excluded from the reference data.

Table 7: Results from detection of S2 signal-to-noise ratio method

Ref Data: S2 AND VHR (V2)
All cardinality Types

	Detected	Not detected
True presence	18	53
True absence	28	260

Ref Data: S2 AND VHR (V2)
Types 1 and 4 only

	Detected	Not detected
True presence	12	21
True absence	34	292

Reference Data: 10 pix clusters (V3)
Types 1 and 4 only

	Detected	Not detected
True presence	5	8
True absence	41	305

The results for the Sentinel-2 S/N ratio method were better than for the S1 Gamma method, for arable land (kappa= 0.4 for AL) and permanent grassland (kappa = 0.2 for PG), providing that only case 1 and case 4 of FOI cardinality are selected out of the 4 ones. Commission error seems close to the LQ of 10%. The FOIs detected with cardinality problem seem evenly distributed between case 1 and 4. There is no improvement observed on permanent crops (PC) and agro-forestry (kappa<<0.1). For permanent crops, only those FOIs where very notable difference in the distribution of the surface reflection is present could be detected as having potential problem. By increasing the threshold, the heterogeneity detection performance seems to improve for all land use categories; the most notable increase is for AL. As for the cardinality detection the threshold of 5 seems to give the best result; for AL the best result being with 4 as a threshold.

As said before, the SNR method assumes unimodal distribution of the pixel DN values within the FOI. An approach using percentile points (for example the 25/75

[2] Same logic is applicable for the SNR method

quartile) as the histogram analysis method does, could be better option that the simple mean/stdev ratio. It better accounts for the distribution and possible presence of clusters on the far sides of the histogram.

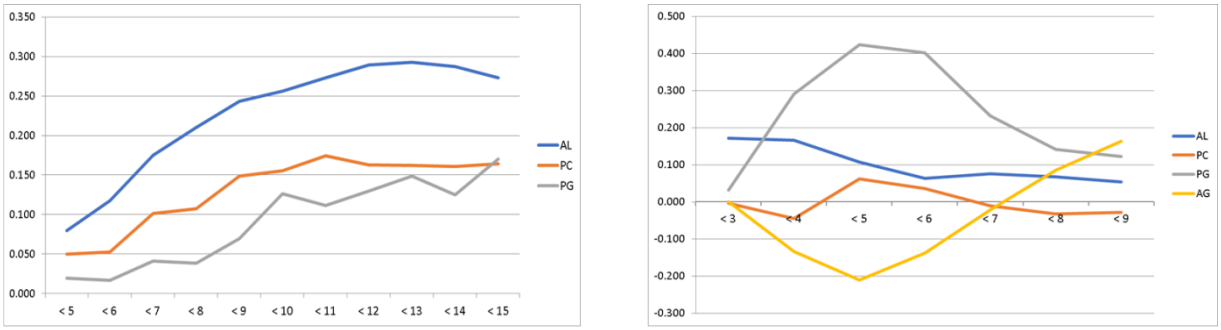


Figure 24: S2 SNR Method: Graphs showing the level of agreement, expressed through Kappa (Y-axis), for different threshold values of the SNR (X-axis) and agricultural land cover: (a) detection of FOI heterogeneity; (b) detection of issues with FOI cardinality

5.2.1.3 Unsupervised clustering through S2 image segmentation

As described in point 2.4, the method uses unsupervised clustering of S2 multi-temporal data in the period from March to May. The scenario under test confirmed integrity problem if there is a presence of: (1) two or three segments > 20 pixels, with class difference >= 2 OR (2) four or more segments > 20 pixels, with class difference >= 1 within the FOI. A negative buffer 5 meters was applied. Clusters of less than 3 pixels were not accounted. Since, the method runs over the entire Sentinel-2 scene, it generated objects/clusters within all FOIs; however, for those FOIs that are too small, the resulted objects were probably not detailed enough to provide an accurate assessment of the FOI conditions. As explained above, such fields were excluded from the reference data.

Table 8: Results from detection of the unsupervised clustering method

Ref Data: S2 AND VHR (V2)
All cardinality Types

	Detected	Not Detected
True presence	51	21
True absence	83	210

Ref Data: S2 AND VHR (V2)
Types 1 and 4 only

	Detected	Not Detected
True presence	25	7
True absence	109	224

Reference Data: 20 pix clusters (V3)
Types 1 and 4 only

	Detected	Not Detected
True presence	11	2
True absence	74	278

The results from the unsupervised clustering method were promising, although still far from the possible accepted thresholds, especially for the commission error. The performance with respect to the omission (beta) error seems better, especially with the third versions of the reference data, where omission error seems close to LQ of 10%; however, the sample is too small for more definitive answer. Kappa is between 0.2 and 0.3 when only cases 1 and 4 are considered (reaching kappa of 0.4 on arable land). Commission's errors were caused mainly by FOI of type 2 cardinality, which although detected by the method, should not be considered critical in the CbM context.

Commission errors are mostly due to: (1) adverse impact on segmentation of objects with high contrast present on the FOI border (for example, paved roads), which results in the propagation of neighbouring segments within the FOI entity. (2) adverse impact on segmentation of the occasional geometric in some of the Sentinel 2B images; (3) thresholds adapted to detect presence of two different crops from the same seasonal group, which being less conservative, accidently pick up also intra-parcel variability, typical for the Catalonian landscape (soil inundation, impoundments, landscape elements, terraces); (4) deficiency in the method for assessing the clusters within the FOI, initially relying on zonal statistics on FOI with 5 meter negative buffer. Results could be improved by applying a more targeted land cover-specific analysis, taking into account the spatial context (proximity to feature, sub-and super objects relation) of the given FOI, and substituting the zonal statistics with an object-oriented approach (using the topology relation between segments and between object layers) for the FOI assessment. Trials with other segmentation methods (ex. Watershed) are also an option.

5.2.1.4 Multi-temporal S2 supervised classification

As described in point 2.5, the method uses supervised machine-learning of S2 multi-temporal data from May (30 LC classes). The scenario under test confirmed heterogeneity, if the main land cover class found constitutes [30%-70%] of FOI area and confirmed integrity problem, if there are at least two clusters with different class label with at least 20 pixels present. The applied clustering is based on 8 connection kernel.

Table 9: Results from detection of the supervised classification method

Ref Data: S2 AND VHR (V2)
All cardinality Types

	Detected	Not Detected
True presence	11	62
True absence	24	268

Ref Data: S2 AND VHR (V2)
Types 1 and 4 only

	Detected	Not Detected
True presence	7	26
True absence	28	304

Reference Data: 20 pix clusters (V3)
Types 1 and 4 only

	Detected	Not Detected
True presence	3	10
True absence	20	332

Results from the S2 supervised machine-learning calculated for the FOI cardinality seemed promising as well (kappa = 0.51 for cases 1 and 4 on AL). This method had lower amount of commission errors. Omission errors could be further reduced by improving the input data for the learning process – increasing the geographic extent of the processed area, and using more accurate training data for the non-agricultural land cover types. The thematic raster obtained from the classification was not post-processed or filtered in order to preserve any information on linear features. Thus, the abundance of the different types of land cover within the FOI calculated with zonal statistics could refer either to compact clusters present or individual pixels completely scattered over the FOI. Also, the assessment hasn't included results with the application of negative buffer. Another important factor was the quality of the training data. For the agricultural areas, the method relied on the GSAA as input; however as with any farmer declaration dataset, it contains certain amount of error. The Catalonian administration kindly provided also the OTSC results over NOUR in 2019. The spatial overlay found that 180 FOIs from the reference dataset have been verified on the spot for their land use and area.

This data, although not used as training set in the current test, should be taken into account in further studies.

5.2.2 Overall

The comparison of the results from the methods with the collected ground truth provided an ambivalent picture, with no standalone method scoring sufficiently well for both commission and omission errors. Methods seem to be complementary rather than competing, each addressing and revealing different aspects of the observed phenomena.

The particular LPIS design (based on cadastral parcel) and the rugged landscape in Catalonia was an important factor that challenged the performance of the methods. Detection performance varied also depending on the type of agricultural land cover. Yet, the methods based on clustering seemed to score better in comparison with the two statistical methods assessed, especially for what concerns spatial cardinality detection. The methods based on analysis of pixel /cluster typology (class assignment) and their distribution within the FOI also provide better insight into the FOI conditions, than those based on metrics aggregated at FOI level (means, StDev). The statistical methods look promising for the detection of a more generic heterogeneity (related to G2 type of information extraction). With respect to the FOI typology, S1 Gamma method seems to pick better the cases 2 and 3 comparing to the S2 S/N, which performs better on cases 1 and 4.

Figure 25 provides an overview of the methods tested, together with the estimated strong and weak points, their complexity and deployment capabilities.

FOI assessment methods	Performance and Implementation				
	Strong points	Weak points	Possible improvements	Complexity of method	Portability to JRC DIAS notebooks
S1 Gamma	Consistent sampling, cloud independent, relation to land physics	Sensitive to terrain ruggedness, no area dimension, needs info on precipitation	SAR signal filtering, noise removal, thresholds adjustment	Simple	Easy
S2 S/N	Based on known metrics, intuitive	Threshold is arbitrary, assumes normal distribution, no area dimension, clouds / shadows	DIP test for normal distribution, more elaborate metrics	Simple	Easy
S2 interquartile	Reveal clusters through pixel distribution, no need for normal distribution	Sensitive to border effects and cloud shadows	Scene filtering, negative buffer	Simple	Easy
S2 segmentation	Derive clusters reflecting land behavior, spatial context, no training data need	Segmentation artifacts, information loss	Tuning of segmentation, spatial context rules	Moderate	Difficult
S2 supervised ML	Derive clusters related to land cover features, comprehensive statistics	Dependent on abundant training data with high quality	Improving the learning algorithm, better training data	Complex	Moderate

Statistical methods

Clustering methods

Figure 25: Overview table with the methods tested

The results also provided useful insights into certain methodological and technical challenges and caveats. Some of them are related to the cloud cover filtering and associated data gaps, other to the appropriateness of the statistical methods, the parameters of the segmentation process, the completeness of the training data or even to the need for introduction of ancillary information, such as digital terrain or digital surface models (Figure 2). Also, the predefined period for the Sentinel image time series was not always optimal for the type of land cover/use expected (as declared in GSAA).

The project also underlined the complex nature of the feature of interest, as representation of the real land phenomenon. The distribution of the bio-physical material within the FOI and its temporal behaviour has multi-faceted aspect, which could be captured in full, only with the implementation of different observation procedures (methods) that complement each other (Figure 26). In addition, the observed aspects are very much land cover and land use specific. The future work would require a close collaboration with the local EU MS experts in order to process the method outcomes in their proper context. The assessment of the FOI conditions will call for the adoption of tiered approach, addressing at each tier, a particular heterogeneity/cardinality type in specific land cover/ land use, or even scenario, context. Same philosophy should be adopted when assessing the performance of the methods.

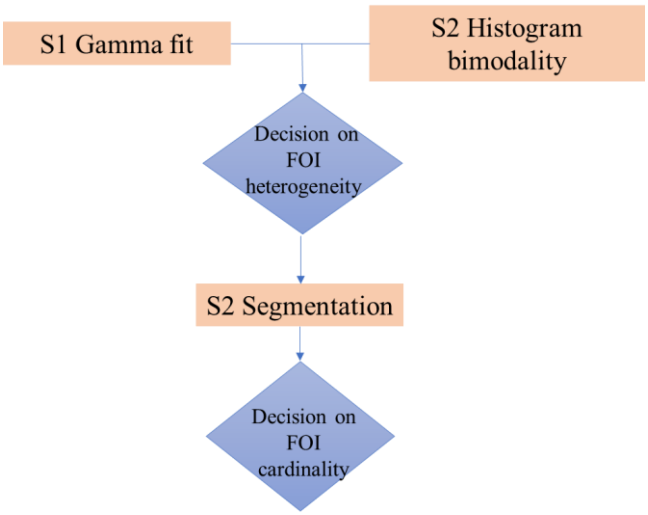


Figure 26: Methods as bundle product in CbM processing chain

The setting up of the quality assessment also revealed some important characteristics of the "FOI behaviour". A promising discovery from the built-up of the reference data was the fact that the majority of the observed "FOI inconsistencies" were visible already by the month of May, which is still within the GSAA application period (Figure 27). This suggests that both FOI spatial variability (G2) and FOI spatial cardinality (G1) checks could and probably should be incorporated as supporting tools during the declaration process to help farmers improve their GSAAs. It could be also of interest to alert farmers on FOIs with G1/G2 problems in the previous campaign, especially when these farmers are leasing/renting the related agricultural parcels for the first year. This could be beneficial also in case the future the GSAA declaration period is substantially anticipated i.e. Nov. Dec.

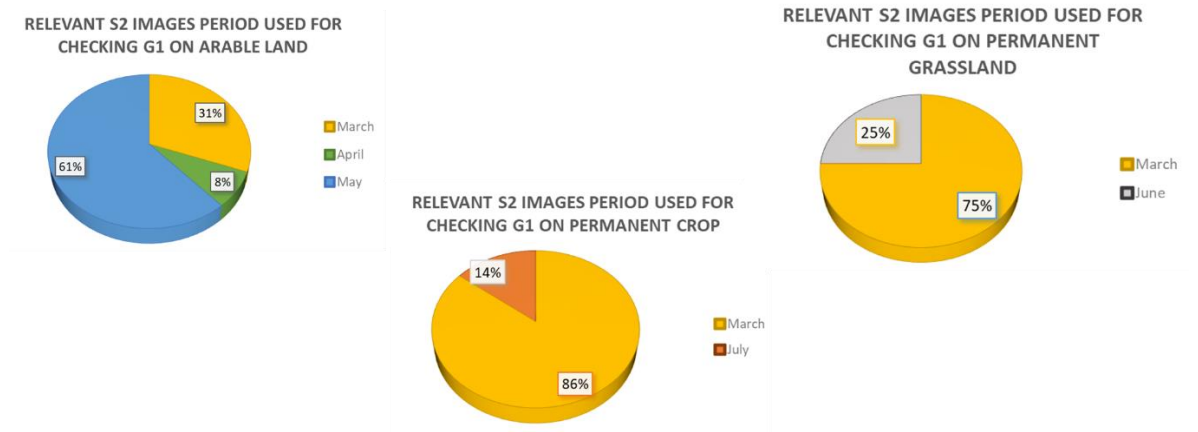


Figure 27: Relevant period for the detection of issues related to spatial cardinality (G1), as collected for the validation dataset in Catalonia (Spain)

Another outcome from the established inspection environment was the opportunity to enrich the portrayal options for the visualization photointerpretation. A possible addition could be the NIR-RED scattered plot, which could be useful to detect bare soil. A calendar view of the cumulative scatter plots (time evolution of clusters in the spectral space) has certain potential to spot easily trends within the FOI that

indicate or reject the formation of multiple physical entities (Figure 28). It could be useful to help defining the proper thresholds for the S2 SNR and S2 histogram analysis methods.

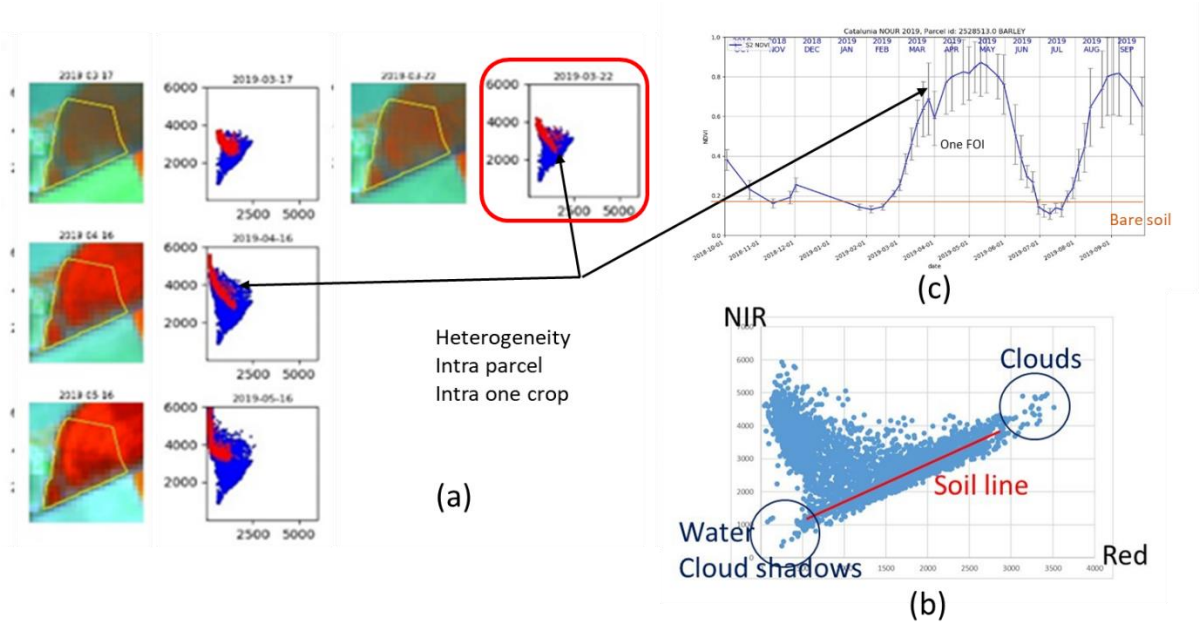


Figure 28: Use of NIR-RED scattered plot in FOI analysis (subset b). Red part: scatter plot of current date; Blue part: scatter plot of all previous dates. The high values of the StDev of the NDVI encountered of the given FOI in April-May (subset c) could be an indication of the presence of multiple units. However, the evolution of the scattered plots (subset a) shows that these high values are related to the variability of the vegetation development within a single unit.

5.2.3 FOI assessment methods in the context of CbM

The elaboration of the reference dataset for testing of the methods, gave the project team the opportunity to understand the capabilities and possible limitations of the information content of Sentinels (mostly the optical sensors) and to have a first glimpse on how the issues with the FOI integrity could be handled in the CbM process.

When creating the reference dataset for checking FOI integrity (heterogeneity, linked with spatial cardinality issue), the project team worked with three different set of assumptions: (1) there is a FOI integrity issue, detectable with Sentinel; (2) there is a FOI integrity issue, detectable with Sentinel and ancillary (VHR) data; (3) there is FOI integrity issue found in (1) or (2), which is big enough to challenge the area component. These assumptions yielded three different cases of FOIs, observable on the Sentinel data using visual methods. They could be somehow related to different components of the CbM workflow, as well as to specific aspects of the QA of the CbM, in general (Milenov P. 2020)

1. FOIs found on Sentinel alone, with conditions indicating an "integrity" problem - multiple physical objects within, internal features dissecting the FOI, features that should not be considered being part of the FOI. These are the cases the automated part of

the CbM (step 1 of QA CbM) should be able to detect, since the visual interpretation assumes there is sufficient data in the Sentinel imagery for the automated method to depict the issue. When detected in the CbM, these would require communication with the farmer or LPIS custodian, and eventually might lead to scenario non-compliance, if the issue cannot be addressed on time.

2. FOIs found as "having something" in Sentinel, but Sentinel data alone was not sufficient to confirm it, unless a further consultation with the available VHR data is made. This is where the automated method based of Sentinel or equivalent data is presumably not good enough. These are the cases, which when detected in the CbM, would end up as doubtful or inconclusive. They will require either a communication with the farmer/LPIS custodian or a subsequent verification in a follow-up procedure. In the latter case, the confirmation or disconfirmation of an "integrity problem" will come through expert judgement by visually consulting ancillary data with superior information content (VHR imagery, national orthophoto). Alternatively, and if technically feasible, such ancillary data could be integrated in the automated process, improving the detection accuracy and thus reducing the number of inconclusive cases upfront.
3. FOIs with an "integrity" problem, impacting the area component (in this case, clusters of at least 20 Sentinel pixels). These FOIs will be only a fraction of the FOIs found with an integrity issue. These are the cases, which when detected in the CbM, and not addressed by the farmer (if GSAA-related) or administration (if LPIS-related), would have potentially an impact on payment.

There are two research questions that come out from the above findings:

To what extent is possible to account in the automated workflow of CbM for all context-related information, easily extractable by manual photointerpretation in order to achieve the same level of performance between CbM and the expert judgement?

Here, the objective is to understand the extent to which we can replicate in the automated system, the cognitive reasoning applied by humans when assessing the FOI conditions. The evidence collected so far, from the internal studies and the EU MS experience, indicates the need for integration in the processing workflow of more data and information than the Sentinel signal alone, as well as to built-up more complex queries for feature extraction and analysis.

What is the true abundance of the FOIs with integrity problem in the entire FOI population?

It is known that the probability of a given detection outcome of being correct would depend on the prior distribution of the specific phenomenon in the whole FOI dataset (Schnuerch 2020). We also know, from the annual LPIS Quality Assessment (LPIS QA) that the issue with the FOI integrity, leading to a problem

with the area component (types 1 and 4), is rare. The abundance of this phenomenon is manifested through certain quality measures of the LPIS Quality Assurance Framework (area non-conformity and critical defects) and we could fairly assume that those EU countries, opting for operational implementation of the CbM, are having such FOIs roughly within 5% of their population (requirement of the CbM boundary conditions). This means that if a method works with 95% of accuracy both for detecting FOI spatial cardinality problem and for confirming the FOI integrity, the probability for an FOI detected as being with an issue, while it's not, could be even 50%. To increase the detection probability, the FOI assessment could comprise several independent tests applying different methods. This will reduce the number of cases for follow-up, by taking only those that seem more probable of having a problem, since they failed several tests and not just one.

These questions should be tackled and addressed in the future developments of the FOI assessment project.

The other important aspect is that there might be no need to follow-up all "inconclusive" FOIs. Most of them could be part of a specific category of phenomena; thus checking sample of them, would allow the automatic resolution of the others. In the particular case of Catalonia, the project team found that the "integrity problem" of some FOIs was related to presence of internal and stable linear elements, well visible on the digital surface model (DSM) of Catalonia publicly available through INSPIRE. Presumably, if this DSM is integrated in the automated processing workflow of CbM, these inconclusive FOI could possibly be handled correctly already in Step 1. Consequently, DSM data was found useful to "filter out" those FOIs found with "integrity problem" not related to spatial cardinality.

Another point to consider is that, whatever the spatial cardinality method is (statistical, cluster- based, object-oriented), the FOIs on which it is applied, should be of certain minimal size. For the method to detect "more than two sub-objects", the FOI needs to be big enough to allow the depiction of all individual parts. For clusters of minimal size of 20 pixels, the FOI needs to be at least 2 x 20 pixels. For S1 and S2, it means that FOIs with size of roughly 0.4 ha.

The purpose of the FOI assessment method currently developed is to provide the CbM forerunners with a prototype (or series of prototypes) for detecting and monitoring the status of the FOI, which can be embedded in their operational CbM process. It seems that most of the Paying Agencies engaged in the CbM so far do not have explicit checks of FOI cardinality or heterogeneity. One reason could be that they consider the presence of FOIs with integrity issues as rare in their system. However, many of them use the T4 type of information extraction (crop identification), which seems to partly be serving the purpose. The outcome produced by this T4 test could also tell something about the integrity and validity of the FOI. One can presume that G1 (and probably partially G2 as well) is implied in the "black-box" types of solution. Those FOIs resulting with the crop type/land use declared as confirmed with high degree of probability, are most probably those where the FOI representation from GSAA and Sentinels match and the cardinality is 1 to 1. Contrary, those FOIs resulting with the crop type/land use declared as not confirmed, could contain cases where the low probability for detecting the crop/land use declared is due to an issue with 1-1 cardinality. This assumption was confirmed by the 2019 QA CbM outcomes. The problem is that it would not be straightforward to identify those cases, from the outcomes of the "black-box" (mostly based on machine learning) only. Nevertheless, we can assume that for

those cases where another crop/land use is found with high probability, there is no issue with cardinality; So, it would be rather in the population of those where no particular crop/land use is detected with sufficient probability that these FOIs may present cardinality problem.

So far, there was an underlying assumption that the outcome of any decision related to FOI cardinality issue is Boolean (problem not found/problem found). In fact, and especially when T4 is used, there could be a third outcome, which is inconclusive. The way how such outcomes will be processed in the QA of the CbM is still to be defined.

Project team also agreed that rugged terrain with numerous small impoundments and eroded slopes, as was the case with of the Catalonian test site, influences the uniformity of the crop cover and adversely affect the assessment of cardinality. Situation could be further complicated by the systematic presence of internal landscape elements (mostly terraces) designed to regulate the land on slope and thus, making it suitable for agriculture. However, these elements, considered as obstacles along the path towards successful FOI cardinality assessment, are valuable features in the context of the CAP green infrastructure. Thus, whenever a FOI cardinality assessment method fails to perform due to excessive commission errors caused by the influence of terrain conditions/feature, it might be right tool to capture soil and landscape aspects, relevant for checking and monitoring implementation of conditionality requirements and farming practices beneficial for environment and climate.

The preliminary analysis of the finding evidenced that certain outcomes of the FOI assessment methods could provide useful information for checking certain GAEC commitments, such as retention of landscape features and preservation of terraces, defined in GAEC 7. For example, despite having relatively high commission errors for the spatial cardinality, the segmentation method proved to add value by detecting variabilities within the FOI, such as terraces, landscape features and shrub encroachment within natural grasslands – all very relevant in the context of compliance scenarios that cover CAP greening aspects (Figure 29).

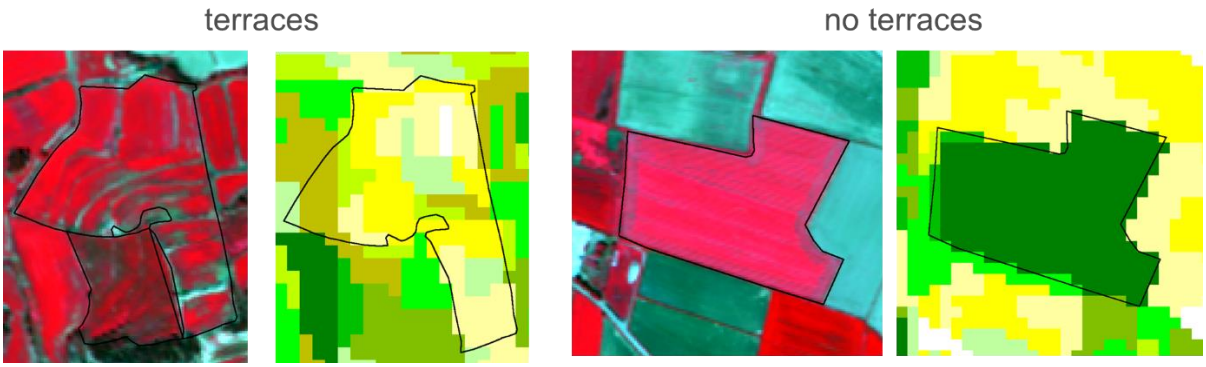


Figure 29: Parcels with terraces results with more images segments than those without (JRC Study case in Catalonia, ES)

6 Operationalization of the methods, new solutions, way forward

6.1 Deployment on JRC DIAS

FOI assessment methods were developed using different approaches, algorithms, libraries and tools. For that reason, they are in different stage of "readiness" to be deployed in the JRC DIAS Hub, as Jupiter notebooks. The S2 S/N and S2 interquartile analysis methods are easily deployable as notebooks in DIAS, since they are based on the minimum set of tools and infrastructure provided. Same applies to the S1 Gamma method, which although initially designed in GEE, was made portable also in JRC DIAS Hub, without significant efforts. The supervised machine learning method is available as notebook, and it uses the EO Learn library, which is open-source and can be installed just like any other ML method (as tensorflow, for example). The problem is that the classification is using Sentinel images and the API from Sentinel Hub, where the project team does not have full control. The most difficult to operationalize in JRC DIAS environment, would be the segmentation method, which relies partially on some proprietary algorithms from off-the-shelf software (e.g. eCognition) to develop the proof of concepts. Nevertheless, it would be possible to replicate it in the JRC DIAS notebook.

6.2 Towards a DIAS notebook on spatial cardinality

The main conclusion from the work inin 2020 was that clustering methods were more promising and, in fact, essential for the detection of issues in relation to FOI spatial cardinality. However, both the segmentation and supervised classification needed additional tuning. GTCAP decided to invest in the development of generic "best practice" notebook for the detection of clusters within FOI representation, provided by any thematic raster. A prototype of such notebook already existed as part of the supervised machine learning method.

This resulted in the so-called "FOI assessment notebook" that combined these methodological elements:

- Cluster assessment within FOI relies completely on information provided by thematic raster.
- Assumption that the classification of the clusters was done mainly using Sentinel (or equivalent signal).
- Could apply negative buffer prior to cluster assessment.
- Customizable to address both G1 and G2 aspects.
- For the cardinality assessment only: Deals with GSAA FOIs having cluster representation of at least 40 pixels.
- For the cardinality assessment only: Focussed exclusively on detecting, whether GSAA $\supset$ FOI (GSAA split case).

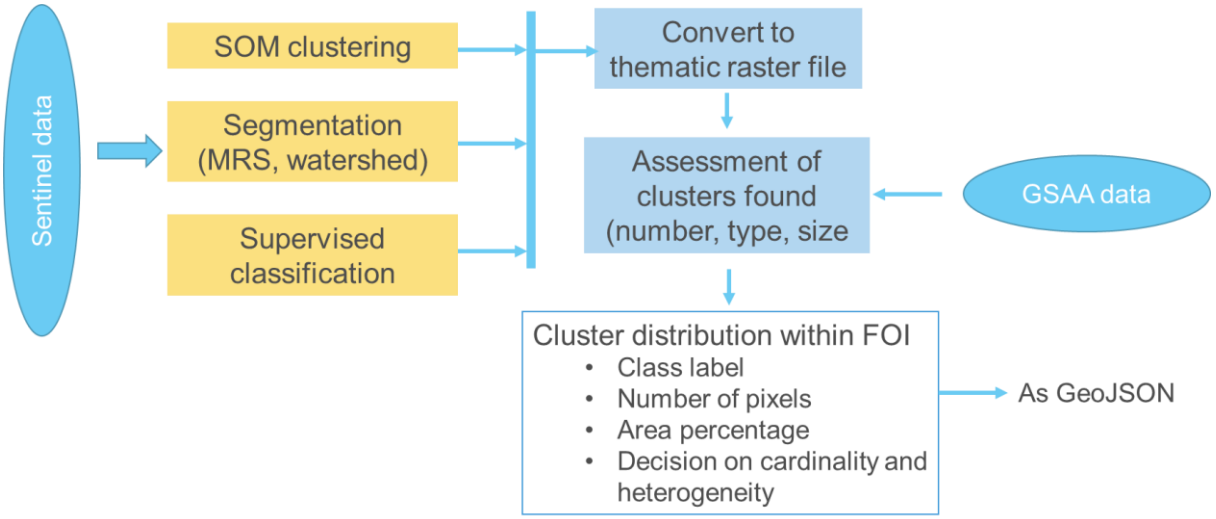


Figure 30: Workflow of the FOI assessment notebook. The thematic raster file could be derived from different methods, such clustering based on self-organized maps (SOM), image segmentation of pixel-based classification.

The FOI assessment is based on spatial analysis of a "thematic" raster file produced in advance. The thematic raster file can be the result of any image/raster processing method yielding a class label for each pixel using Sentinel or equivalent data, complemented with ancillary datasets- crop classification, behavioural analysis of land phenomenon, gridded data on soil, slope, humidity, etc. The starting hypothesis is that inside of a homogeneous FOI representation there should be only pixels of one type (label). For example, if the thematic raster is the result of a behaviour analysis, all the pixels inside the FOI should hold label associated only with one type of behaviour, for the observable a period of time. In case the thematic raster is the result of a crop classification, all pixels inside the FOI we should be labelled with one class only. The FOI assessment uses the analysis made on the presence and distribution of pixels with different label found inside the FOI. It can identify and visualize clusters within the FOI, and further analyse their size and spatial arrangement (Figure 31). Thus, the outputs provided are fully covering the spatial variability and spatial cardinality issues and are in line with the format explained in point 2.5.

For the FOI heterogeneity, there are two methods for pixel assessment – from the pre-filled database or directly in the notebook, using Python. The same applies for the FOI cardinality assessment, where the clustering can be:

- based on vector and area calculation (calculation made in a database).
- made directly at pixel level (analysis made in Python, directly in the notebook).

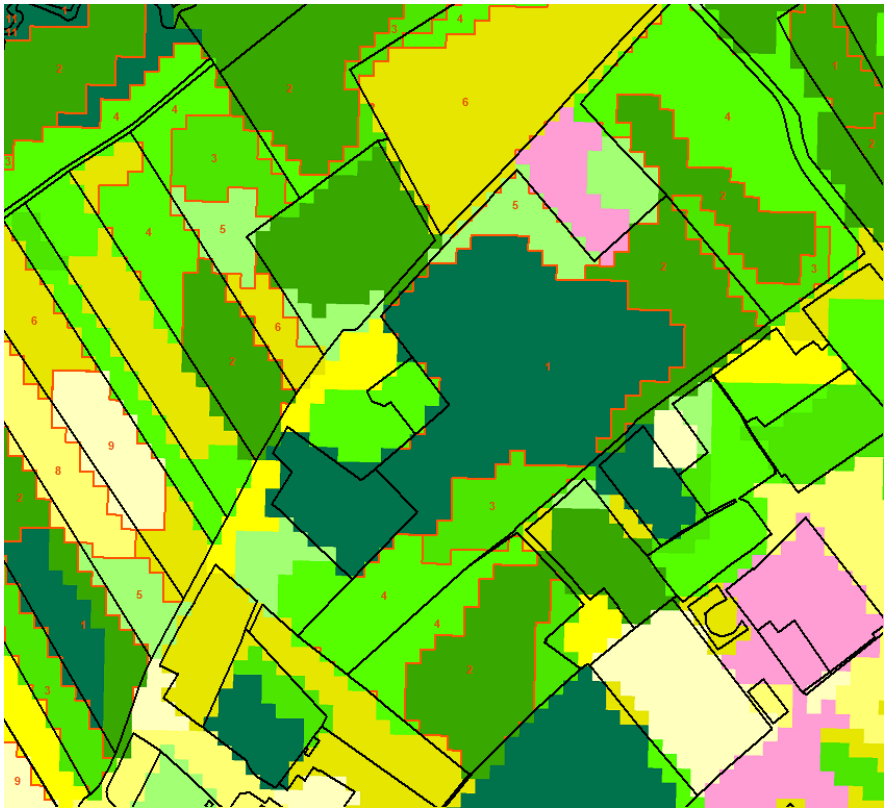


Figure 31: FOI cardinality notebook: Generated clusters exported as vectors. As seen, clusters are not generated over FOI below certain size.

There are some minimum required inputs and parameters for the notebook to work properly and provide tangible outcomes. They are listed below:

- reference data (users could load their own FOI data or select an existing table from the database).
- "thematic layer" - raster file (can be a raster resulted from previous. process - which is already on the server, or a new file uploaded by the user).
- raster file classes (not really "required" - can be generated by reading the raster file).
- heterogeneity thresholds (can be eventually hardcoded into some default values, but not recommended).
- type of pixel connectivity - 4 connected or 8 connected.
- cluster area parameter (the current default value is 2000 m2).

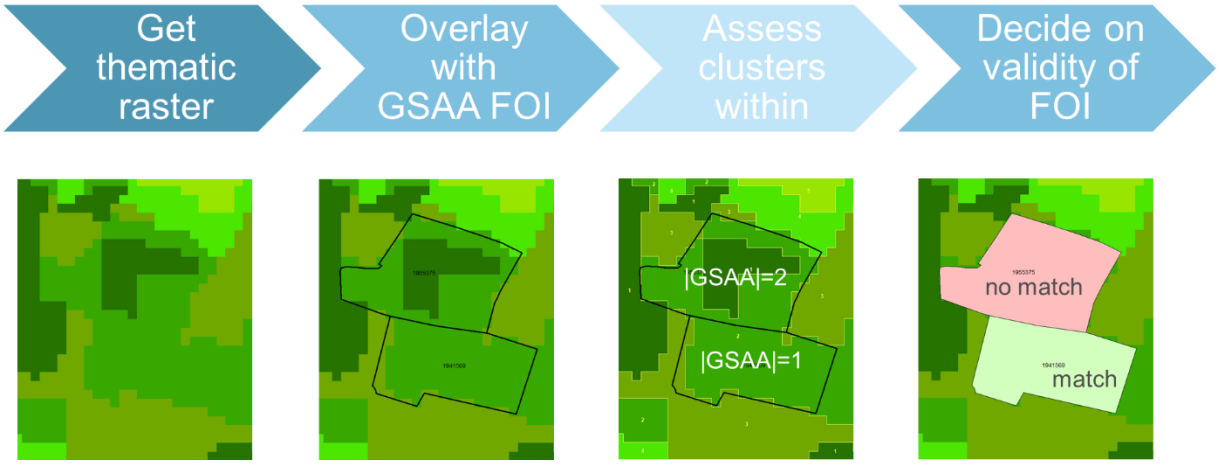


Figure 32: FOI assessment workflow. |GSAA| = n - number of clusters bigger than 20pixels found within GSAA-derived FOI

The cluster assessment notebook is being tested with thematic raster data generated from the unsupervised image segmentation method (in different bio-geographic regions). Full documentation is available on https://jrc-cbm.readthedocs.io/en/latest/cbm_foi.html. The complete workflow is given in Annex III.

6.3 New solutions: Interquartile (IQR) analysis of S2 signal

6.3.1 Description of the method

The notebook also supports a novel method, based on "interquartile analysis".. It monitors the pixel value distribution of the NIR and RED bands of Sentinel-2 and compares them to a uniform or normal distribution. It extracts the relevant statistics (min, p25, p50, p75, max, mean, stdev and count) from the Sentinel (Amazon S3), for each S2 B4, B8 and SCL bands, and stores in a database (roiYYY_s2_signatures). Table may contain 100s of millions of records, but is indexed and clustered, so retrieval is extremely fast. Although not yet formally evaluated, some initial trials evidence its ability to detect heterogeneous FOI representations. The algorithm is fully automated, easy to deploy, and generates both indices and image chip highlights (similar to the calendar view of imagettes, presented above).

As with the previous approaches, the IQR analysis method starts with the underlying assumption that any notable heterogeneity within the observed FOI will be manifested in the distribution of the Digital Number (DN) values of the images pixels within the FOI. It, however, analyses this distribution through the interquartile range (IQR) as statistical metric, instead of the spread of the DN values from the mean, expressed by the variance or the standard deviation. IQR accounts for the skewness in the histogram of pixel values and, in such way, it enumerates pixel subpopulation. A perfectly homogeneous field would have a near-uniform distribution of pixel values and, thus, no major differences in the

distribution across the bins of the histogram. However, a strong non-uniform distribution of pixel values will cause the histogram to be skewed towards the histogram bins that are populated more numerously than the others. IQR provides better indication for the presence of clusters with values being significantly different. IQR is usually derived from the histogram percentile, usually the n-th and (100 – n) th percentile. In our case, we compare p25 and p75. IQR is not affected as much by outliers that determine the minimum or maximum of the pixel distribution and can represent both randomly distributed values as well as discrete localised distributions. Thus, contrary to the SNR, it can tell more on whether the observed variability is random or segmented, with the latter being a strong indicator for the presence of multiple objects or units of management.

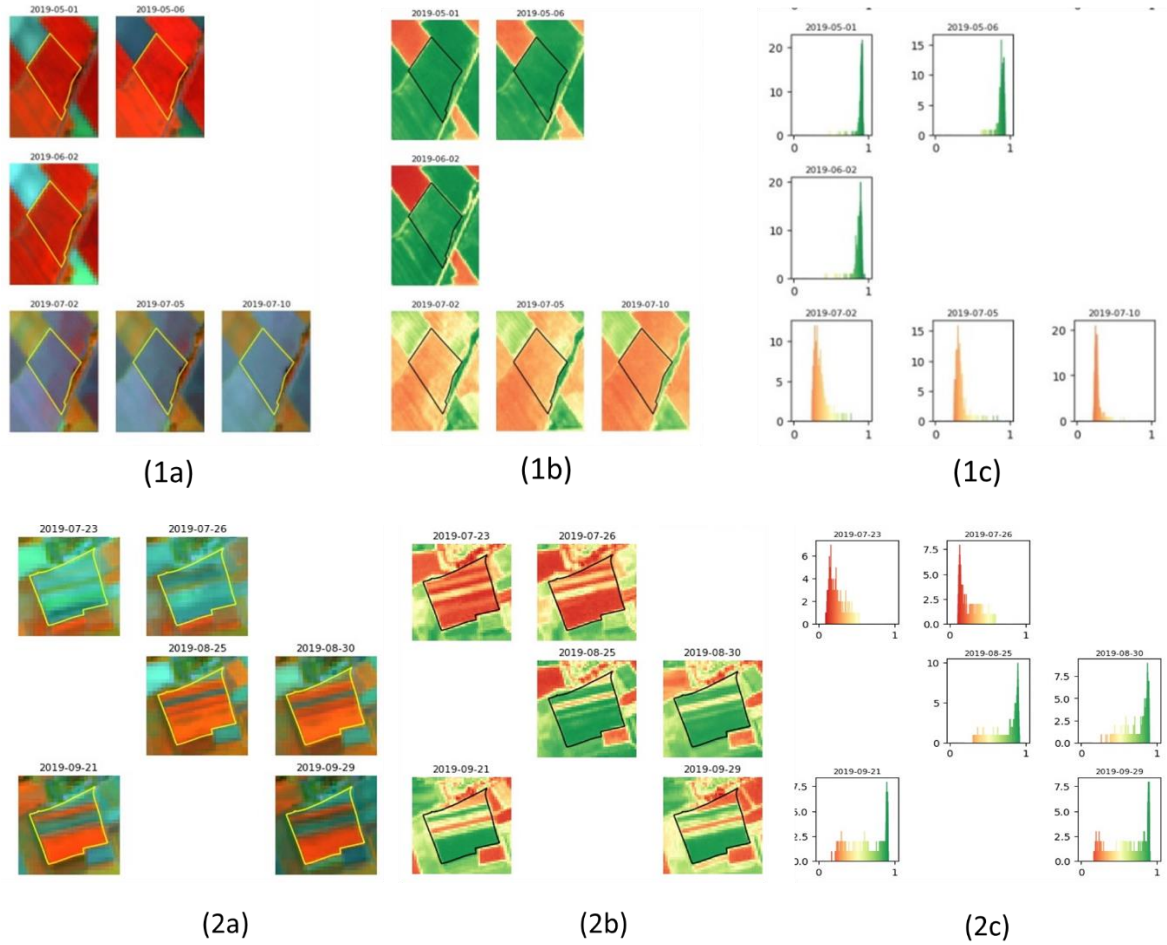


Figure 33: Example 1 of homogeneous parcel (declared as winter wheat). (1a) Calendar view of the imagettes in S2 false colour composite, (1b) calendar view of the imagette with the NVDI heat map, (1c) calendar view of the imagettes of the histogram, where the distribution of pixels is unimodal (most grouped near a single peak. Example 2 of heterogeneous parcel (declared as potato). (2a) Calendar view of the imagettes in S2 false colour composite, (2b) calendar view of the imagette with the NVDI heat map, (2c) calendar view of the imagettes of the histogram, where the distribution of pixels is multimodal (grouping of some pixels at the two opposite sides of the range is visible)

Pixel value distribution is expressed through the quartiles p25, p50 and p75, which are calculated for each Sentinel-2 acquisition. The nature of the pixel distribution determines the "shape" of the quantile function. On homogeneous FOI, where pixel values are rather uniform and follow usually normal distribution, the p25, p50, p75 values are expected to increase monotonically (unless pixel values are constant), following a linear trend. The method actually checks the linearity of the calculated quantile trend (Figure 34). Any notable deviation from the linear increase is an indication that the pixel value distribution does not follow uniform or normal distribution and the initial hypothesis of FOI homogeneity is challenged.

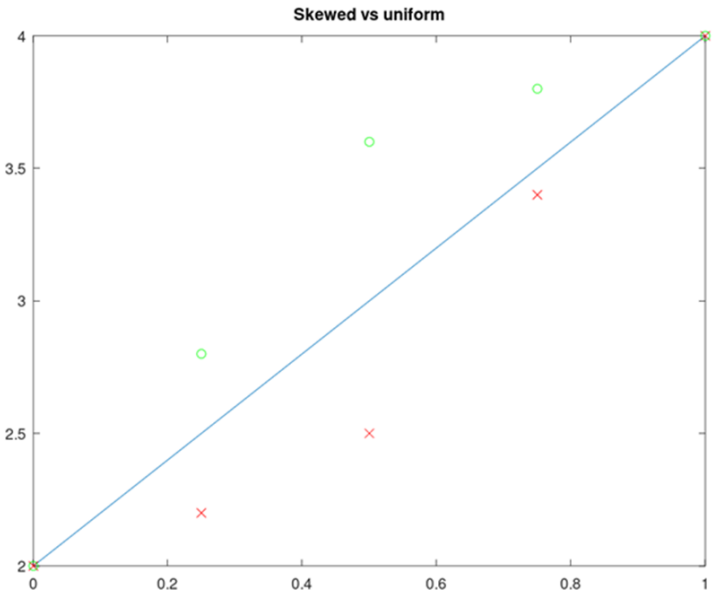


Figure 34: Quantile function for a population of uniform random variate (blue line). Quartile values for asymmetric (skewed) population (green circles and red crosses)

The method calculates the quartiles (p25, p50 and p75) from the histogram of RED channel (Band 4) and NIR channel (Band 8), for each Sentinel 2 acquisition having purely cloud free pixels within the FOI. Working with the individual bands and not with their ratio (as in the NDVI) gives certain advantages in revealing distinct clusters of vegetation and bare soil, which might be otherwise supressed/smoothed by the ratio.

The following two metrics are further calculated:

Interquartile range (IQR) = p75 – p25 = pdiff

Distance of the average of p75 and p25 from median, in absolute terms =|(p75+p25)/2 - p50| = |pdist|

Quartiles are calculated from the Level 2A product of Sentinel-2, which is atmospherically corrected. For each S2 acquisition, the FOI is flagged as potentially heterogeneous, when the values of pdiff and pdist indicate a notable deviation from the linear trend.

As with the previous methods, in order to confirm heterogeneity, the interquartile analysis method should indicate a persistence of a notable deviation in a number of (preferably consecutive) S2 acquisitions.

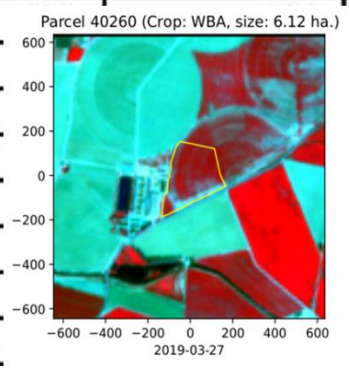
The periods defined for the Sentinel 2 observations were adapted to the relevant part of the growing season of the crop/land use declared.

6.3.2 Meaning of the output produced

The method was written in Python and compiled as Jupiter notebook on the JRC CbM GitHub repository. It uses as input, the extracted signatures (min, p25, p50, p75, max, mean, stdev and count) of the B4 and B8 bands for each Sentinel 2 acquisition and for each observed FOI, stored in a separate file. The table may contain 100s of millions of records, but since it is indexed and clustered, the data retrieval is extremely fast. It automatically generates tables with the outputs, together with a graph with the temporal evolution of the image bands and indices, as well as the image chip highlights.

Table 10: Output from the histogram analysis method per given FOI and all Sentinel 2 acquisitions in the relevant period

date_part	pdiff_B4	pdist_B4	pdiff_B8	pdist_B8
2019-03-27 11:50:31.025000	0.299	0.068	0.302	-0.051
2019-04-11 12:50:29.025000	0.218	0.038	0.409	-0.012
2019-04-16 12:50:31.024000	0.301	0.061	0.439	-0.022
2019-04-26 12:50:31.024000	0.195	0.06	0.505	0.028
2019-05-01 12:50:39.024000	0.121	0.041	0.484	0.015
2019-05-06 12:50:31.024000	0.123	0.042	0.496	0.008
2019-05-11 12:50:39.024000	0.118	0.034		
2019-05-16 12:50:31.024000	0.163	0.051		
2019-05-26 12:50:31.024000	0.084	0.023		
2019-05-31 12:50:39.024000	0.122	0.038		
2019-06-10 12:50:39.024000	0.312	0.032		
2019-06-15 12:50:31.024000	0.218	0.0		
2019-06-20 12:50:39.024000	0.101	0.004		
2019-06-25 12:50:31.024000	0.091	0.002		



The main output parameters are as follows:

pdiff Interquartile range per S2 image band (B4 and B8)

pdist Relative different to median per S2 image band (B4 and B8)

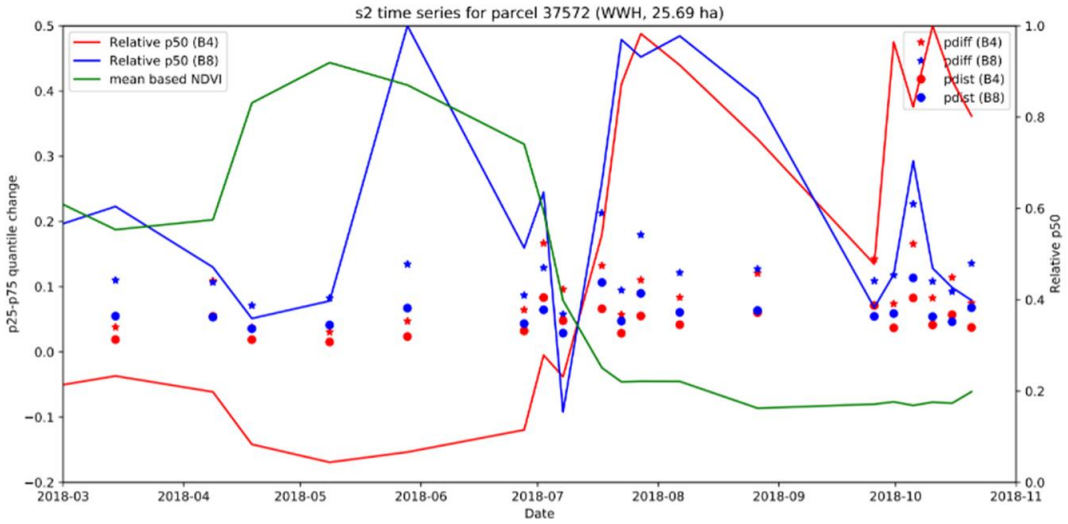


Figure 35; Example of temporal evolution of the images bands and indices for a given FOI with winter crop. As seen from the graph, the blue dot (for pdist) is systematically lower than 0.1 and occasionally close to zero, during the crop growth and crop maturation period (April-June). The observed difference is within the limits of the expected intra-parcel variability of the winter crop cover; thus, not sufficient to invalidate the winter wheat scenario. The Relative p50 is the "pdist" and "p25-p75 quantile change" equals -pdiff (as values are positive).

6.3.3 Calibration and validation

The method has undergone several major revisions before reaching the current proposed settings. Previous versions based on the rather arbitrary difference in 25 and 75 percentile values, relative to the mean, have not shown promising results. The latest version presented in this document still needs to be assessed more rigorously. Nevertheless, visual assessment made on a small sample revealed that the method is quite robust in relation to the possible notable differences in FOI shape and boundary and seems to cope with different parcel sizes (except those being very small). However, it is sensitive to some border effects, especially the case of noise introduced by the still occurring and unwanted geometric shifts (1-2 pixels) between S2A and S2B sensors. Since the method tries to exploit every possible cloud free acquisition, it suffers also from the deficiencies of the SCL classification used to detect and flag cloudy pixels of Sentinel-2. This leads to a high number of commission errors.

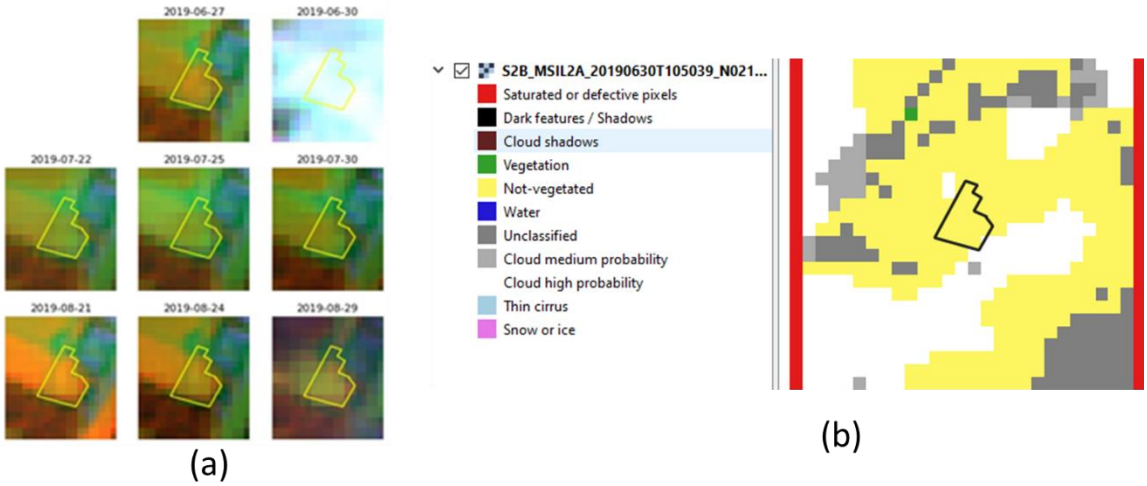


Figure 36: Example of errors in the SCL classification (a) calendar view of imagettes with a cloudy acquisition for 03.06.2019 (b) the SCL classification of the same acquisitions has classified the cloud as non-vegetated area

It would be worth mentioning another similar method for assessing parcel heterogeneity, developed and tested previously in GTCAP, based on the detection of bimodal and multimodal distribution of the pixel DN values within parcel geometry. The underlying idea was that if there are mixtures of sub-objects within the FOI, then it will be manifested in the histogram with the presence of two or more peaks (i.e. the mode in statistics)[3]. Detection of peaks in the histogram could be a sign for several sub-objects in the FOI. A prototype of this method was developed in R.

6.4 Future prospects

The outcomes of the FOI project showed the complexity involved in assessing the spatial context of the FOI using Sentinel data. FOI relates to phenomena that are multi-dimensional and very much context specific (landscape, parcel structure, management practices). In this respect a "one-size fits all solution" is unfeasible. For these reasons, any future endeavour towards more tailored FOI assessment method should account for the following three aspects:

Keep things simple and modular: As seen from the project outcomes, the designed methods are addressing different aspects/characteristics of the bio-physical phenomenon and associated unit of management, represented by the FOI. The more comprehensive is the characteristic, the more complex and resource-demanding the method is. To trace properly the evolution of FOI conditions, the CbM would require a combination of several methods, which complement each other. In this respect, it would be practical and efficient to apply a tiered (or reductive) process: (a) start with a simple statistical method (S1 Gamma, S2 SNR, S2 IQR) applied to all FOIs under monitoring to separate the homogeneous from heterogeneous cases; (b) apply more sophisticated clustering methods to depict those heterogeneous FOIs containing different objects; (c) analyse these multi-cluster FOIs with object-oriented (segmentation) methods to assess the cluster size and their topological relationship (Figure 26).

Collaborate with community: Methods presented in this document were never considered to provide a fully-fledged, ready-to-use, solution. They were rather regarded, as generic prototypes designed to be customized further by the users (EU Member States), in their local context. To understand the observable behaviour depicted from the Sentinel signal, the local expert knowledge is required. Thus, any future development should be in close collaboration with the CbM adopters and the relevant expert community (applying the multi-actor approach). The project team included the notebooks on FOI assessment as part of

[3] http://www.brendangregg.com/FrequencyTrails/modes.html

the library of JRC tools to be shared with CbM technical community. This will foster the exchange of ideas on new solutions, novel approaches and optimization, bringing also the existing experience of the CbM adopters (Malta, Spain) on the spatial validity checks and of the technology partners (Tragsatec, Sinergise) on the design of heterogeneity markers[4].

Think out of the box: So far, the focus of the methods was on their ability to reflect the CbM boundary conditions (correctness of area component and recorded agricultural land cover) and to depict the spatial variability associated with the most common EU CAP scenarios (permanent grassland and annual crop shares, in the context of greening). However, it would require a little shift of the viewpoint to see that they capture information on the field conditions, very relevant to other EU CAP domains, such as the cross-compliance and green infrastructure (variable soil condition, presence of landscape features), or the usefulness in the frame of the LPIS update cycle (detection of non-active/abandoned land). The following chapter provides examples of such use cases.

[4] https://medium.com/sentinel-hub/area-monitoring-homogeneity-marker-742047b834dc

7 Emerging specific use cases

7.1 Findings from the quality assessment of the unsupervised segmentation method

During the evaluation of the performance of the FOI assessment methods, the project team invested additional time to obtain more insights into the typology of the errors found, and the possible causes for observed disagreement. Segmentation methods were of particular interest, for its ability to provide an indication of the location and distribution of objects of potentially different nature.

The project team performed a more thorough visual analysis of the segmentation results for the FOI cardinality, from the multi-temporal S2 data acquired in the period of March to May, over the test site in Catalonia. The main goal was to understand the distribution of the errors between the different FOI cardinality types (Figure 20).

The rules applied in the visual interpretation of the calendar view of Sentinel-2 imagettes were as follows:

1. more than one segment with different class in the FOI;
2. difference between class numbers is more than 2
3. area share is substantial (more than 10-15% of the FOI)

During the assessment the operator took into account also the geometry (shape, complexity) and used the VHR imagery for confirmation of doubtful cases.

The results showed that the type 4 cardinality problem, associated with non-agricultural areas, was always detected by the segmentation method. Type 2 was not detected in the 2/3 of the cases found by the QA; while type 1 and 3 were not detected in the 1/6 of the cases, each (Figure 37).

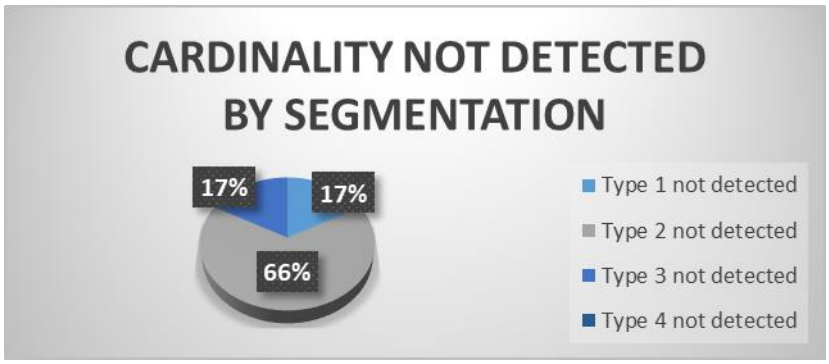


Figure 37: FOI cardinality detection through segmentation: Results from the assessment of the error typology

Closer look on the FOI cardinality type 4, being always detected, showed that they were corresponding to cases of persistently bare natural or artificially sealed surfaces, present within the FOI. In some occasional cases, where the surfaces could be considered agricultural, they were corresponding to either (1) area covered with dry vegetation present through the entire period, or (2) permanent

tree crops having intra-row space sufficiently wide to reveal the bare soil beneath (Figure 38). This gave the idea to check whether the successful detection of such bare/inactive area could be useful in some cases related to LPIS update and cross-compliance.

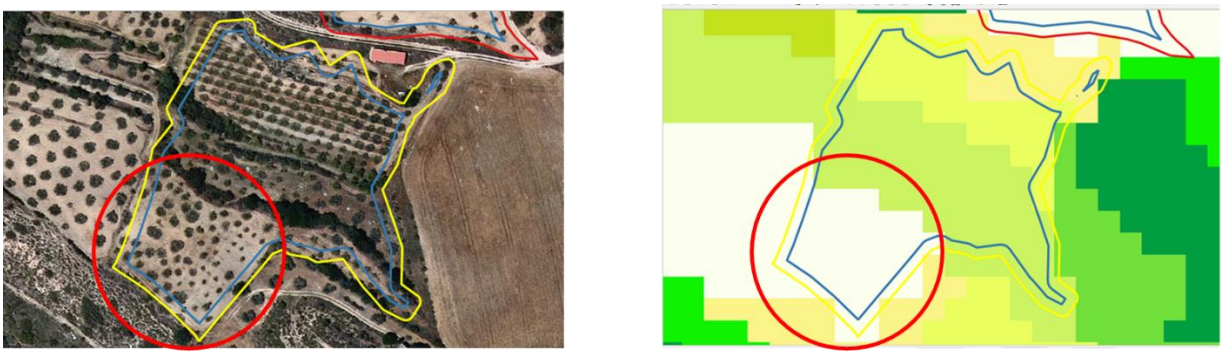


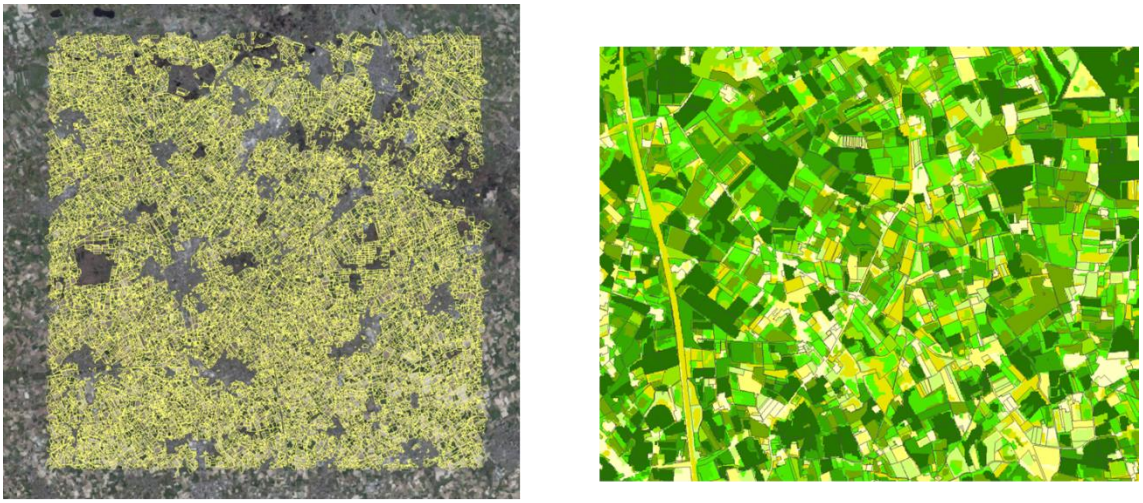
Figure 38: (a) Example of an FOI with permanent tree crop on sloping terrain. (b)The part in south-west corner is labelled with class 10 (bare soil, inactive).

7.2 Checking ineligible area for LPIS update

During the discussions on the 2019 and 2020 quality assessments of the CbM, both JRC and a MS Administration discussed the possibility for the introduction of explicit FOI validity check (G1) in their CbM. Currently this MS administration applies CbM mostly for BPS and greening exemptions, where the detections of the agricultural land cover and annual use rely on supervised machine learning. The validity of the FOI is guaranteed by the LPIS annual update cycle, where a set of automated techniques are regularly run on the up-to-date orthophoto, to verify for presence of unaccounted ineligible areas (built-up surfaces, abandoned land). In the light of the introduction of a possible G1 check, the MS administration expressed interest to explore the potential of some of the JRC clustering methods for timely detection of the non-agricultural areas, invalidating the FOI integrity. The project team decided to assess the usability of the segmentation method in this respect, on dedicated area of interest (AOI), provided by this MS administration (Figure 39).

The procedure consisted of three main steps: (1) run the unsupervised segmentation method over the AOI, using Sentinel-2 data from 2020; (2) assess thematic raster file through the FOI notebook available on the JRC DIAS hub; (3) analyse the outcomes.

Input data were: (1) GSAA extract from 2020, covering all types of land cover (agri, non-agri); (2) Cloud free Sentinel imagery acquired in the period March to August 2020.



(a) (b)

Figure 39: The test area: (a) 2020 GSAA extract; (b) section of the thematic raster file produced

The thematic raster file, derived from the unsupervised segmentation, "maps" areas with distinct behaviour. Those areas having significantly different behaviour (on the opposite side of the max diff histogram) are assumed related to distinct land cover features. Following this logic, features (clusters) with very little difference in the brightness change within the period could be assumed as related to areas within the FOI, being "non-active" in agricultural sense. The internal validation and calibration made, showed that clusters labelled with class 10, refer in 95% of the cases to built-up surfaces and associated areas, or agricultural areas being in "non-active" state, such as bare soil and long-term fallow. FOIs having clusters of class 10 occupying more than 25% of the FOI (equals 0.1 ha for FOIs with minimum area of o.4 ha) were flagged as candidates for further verification.

Several runs were made with different observed periods, different S2 band combinations, and algorithm settings. Each run provided as output, the resulted thematic raster file and the associated tables with the list of FOIs and related statistics derived (see point 2.5). The outcomes were checked against 365 randomly selected FOIs, which were assessed visually using the national orthophoto.

The best results were achieved with multi-temporal S2 data from March to August and S2 bands B2, B3, B4, B6, B8 and B11. The assessment made, following the reporting template adopted for the QA CbM (see Annex II) shows considerable agreement with respect to the detection of artificial built-up area (Figure 40). The commission errors were within the LQ of 5%. The omissions were above the acceptance threshold; however, the sample is too small to draw definite conclusions (occurrence of this phenomenon in the FOI is considered being rare).

Table 11: Results from the detection of the built-up clusters within FOI, using the unsupervised segmentation method (GSAA sample)

Detection of built-up clusters		Automated method		Sub-population
		Detected	***Not detected***	
Actual situation	***Present***	8	2	10
	Not present	9	346	355
	Sum	17	348	365

More important was the fact that all clusters of class 10 found in the 17 FOIs detected by the method in the observed period, were truly related to areas, where potential anomalies in LPIS context might occur. They deserve attention and subsequent tracking in the course of the season. Figure 40 below provides an overview of the typologies found.

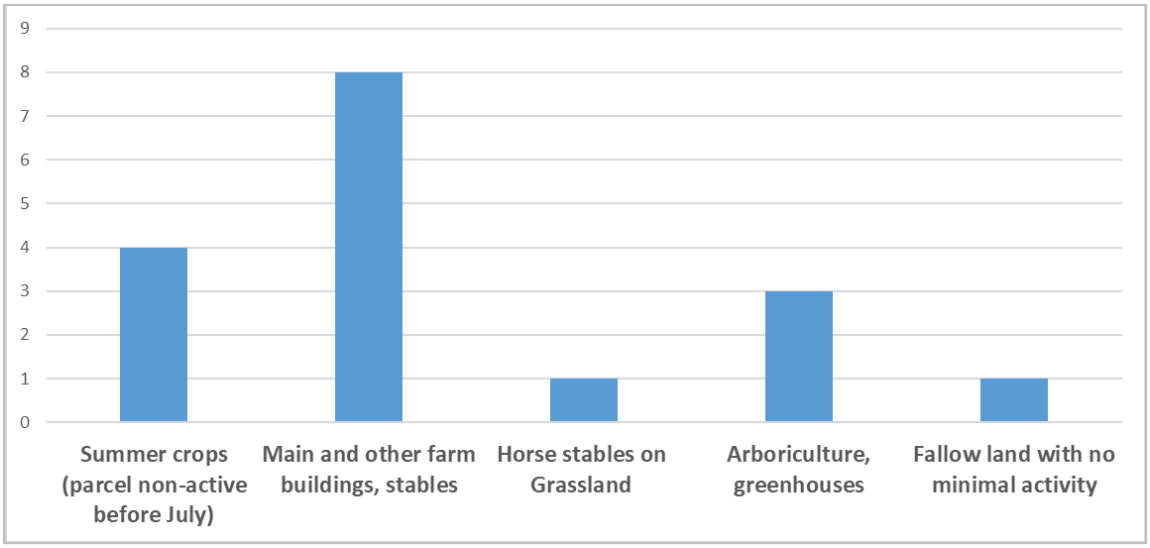
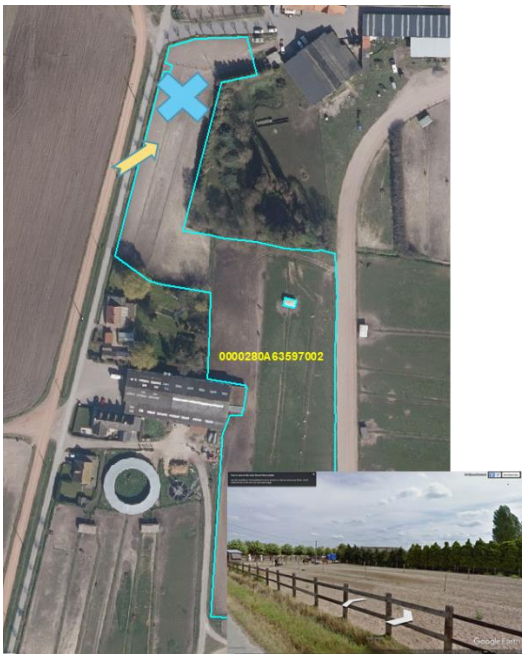
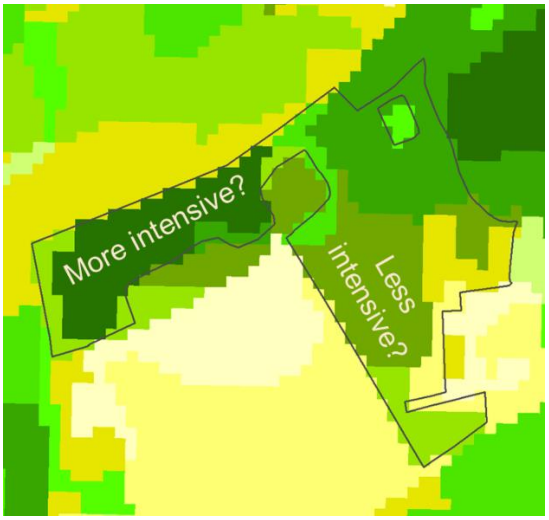


Figure 40: Number of detected FOIs with more than 25% of "non-active" agricultural areas inside (GSAA sample)

Since the purpose of this use case was to test the detection performance of the method and not the quality of the GSAA/LPIS, the full set of GSAA parcels were used, including those related to non-agricultural areas. In fact, all 8 cases of FOI detected with buildings, were already recorded properly as non-agricultural (non-eligible) in the LPIS. Still there were isolated cases, as for the presence of horse stable, which were not recorded as non-agricultural in the system. This is most probably due to the difficulty to depict such areas automatically using one single observation, even if using an orthophoto (Table 12). Another interesting observation was the ability of the clustering method to depict areas of different density of vegetation cover, which in the case of grasslands, could be related to different grazing intensities.

Table 12: Example of detected cases of FOIs with non-active area

Detection, through the thematic raster file, of potential anomaly in a GSAA parcel with grassland. The persistently "non-active" area is marked with X	Paddock visible on the 2020 national orthophoto and on the Google StreetView
	
Clusters of different dynamic detected within GSAA parcel declared as grassland	Same grassland on the 2020 national orthopho. Detected clusters correspond to area with different vegetation cover.
	

7.3 Towards the cross-compliance

The project team decided to apply the same setup designed for the previous use case, but in the context of Catalonia. Here, in this case, the reasons for the detection of non-active areas within the FOI are different. Contrary to the previous area located in North Europe, where the levelled terrain and the intensive agriculture ensure a relatively homogeneous vegetation cover of the FOIs, the relatively poor soil and rugged landscape in Catalonia creates conditions for a non-uniform cover of the herbaceous vegetation. Thus, we could expect that most of the persistently non-vegetated areas would reflect areas prone to soil erosion with poor or absent crop cover, or structural elements designed to regulate the terrain and make it suitable for cropping (ex. terraces). These are elements of interest in the context of CAP cross-compliance measures: minimum land management reflecting site specific conditions to limit erosion (GAEC 5) and retention of landscape features (GAEC 7). Furthermore, linking the heterogeneity detected with the inherent variability of the FOI due to the natural conditions, would allow for the elimination of much of the false positive detections related to the FOI cardinality problem.

The procedure used consisted of three main steps: (1) run the unsupervised segmentation method over the CwRS site NOUR using Sentinel-2 data from 2019 (Figure 41); (2) assess thematic raster file through the FOI notebook available on the JRC DIAS hub; (3) analyse the outcomes.

Input data were: (1) GSAA extract from 2019, with extent over the agricultural areas only; (2) Cloud free Sentinel imagery acquired in the period March to July 2019.

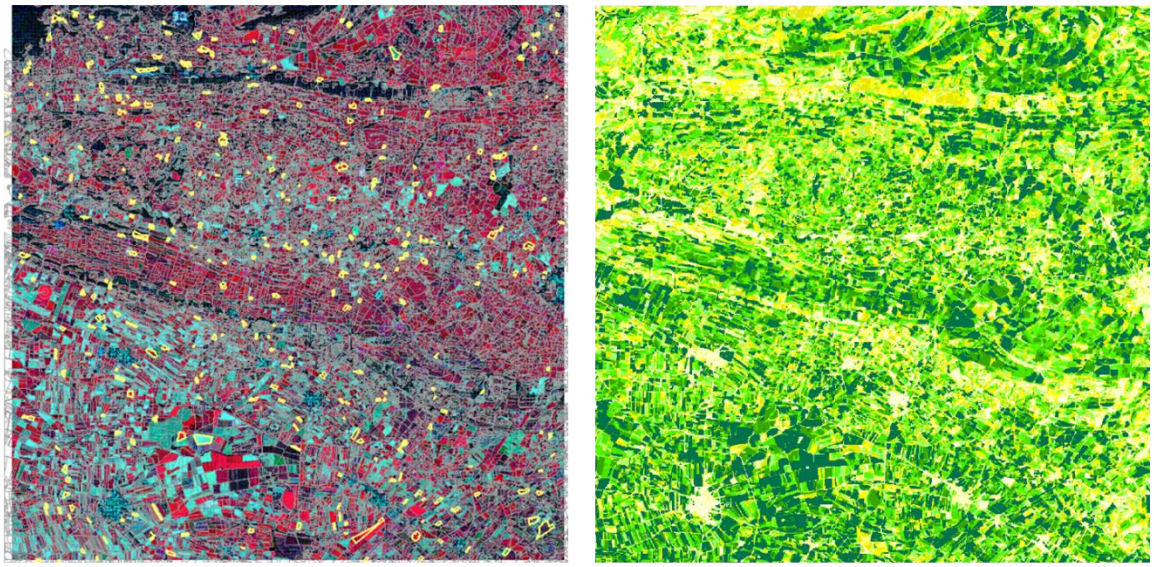


Figure 41: The test area in Catalonia, Spain (Castellsera) (a) 2019 GSAA extract in grey, with selected sample in yellow; (b) Thematic raster file produced

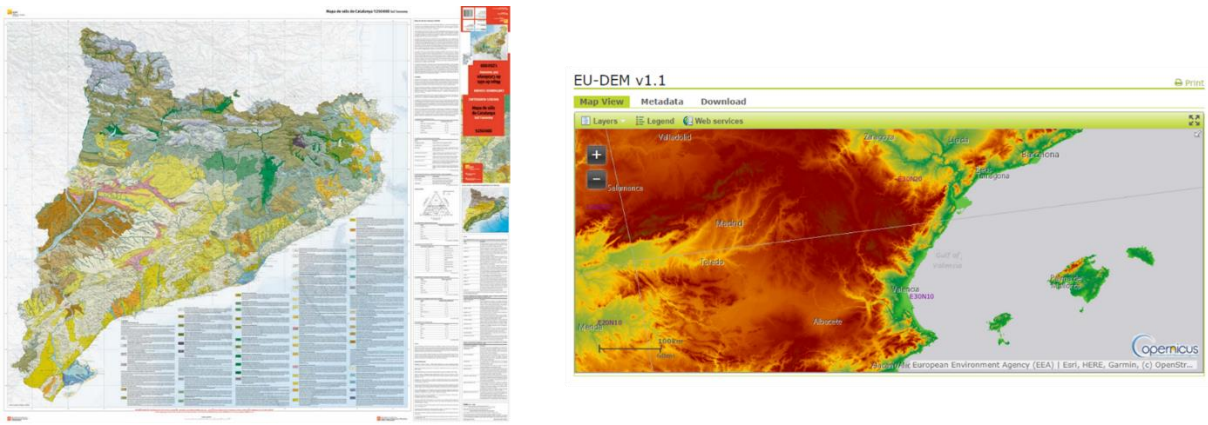
As with the previous use case, the focus was on clusters from the thematic raster file, having little difference in the brightness change within the period, and associated to non-active agricultural areas within the FOI. However, since, these were assumed to be mostly natural bare areas within the cropped parcel,

occasionally covered by scattered dry vegetation, the allowed range of brightness was expected to be larger. The internal validation and calibration made, showed that clusters labelled with classes 8, 9 and 10 refer in 95% of the cases to bare, sparsely vegetated areas or eventually artificial surfaces. FOIs having the sum of clusters labelled with any of the above mentioned classes of more than 30% of the FOI area were flagged as having a potential cardinality problem.

$$\sum_{k=8}^{10}\sum_{i=1}^{n} A_{ik} > 30\% A_{FOI}$$

A_{ik} – area of the cluster i, labelled with class k within FOI
k – class label [8;10]
A_{FOI} – Area of the FOI

To be able to conduct meaningful analysis in the cross-compliance context, the project team asked and received from the Catalonian Administration, the openly available digital soil map of Catalonia, at scale 1:250 000. Furthermore, the project team downloaded the relevant tiles of the EU Digital Elevation Model, available on the Copernicus Land Service Portal (Figure 42)



(a) (b)

Figure 42: (a) Catalonian Soil map 1:250.000; (b) Copernicus Land Service: EU-DEM v1.1

Several runs were made with different time periods, different S2 band combinations, and algorithm settings. Each run provided as output the resulted thematic raster file and table with the list of FOIs and related statistics derived (see point 2.5). The outcomes were checked visually against reference dataset of 365 randomly selected FOIs, prepared previously on the basis of Sentinel-2 and the 2019 VHR imagery.

The better results were achieved with multi-temporal S2 data from March to July and S2 bands B2, B3, B4, B6, B8 and B11. The assessment made for the detection capacity with respect to built-up areas alone, following the reporting template adopted for the QA CbM (see Annex II), shows an expected disagreement between the results and the reference (Table 13), due the prevalent presence of natural bare features that could be erroneously detected as artificial objects. The commission errors were 3 times above the accepted threshold for LQ of 5%. The omissions were above the acceptance threshold too; however, the sample is too

small to draw definite conclusions (occurrence of this phenomenon in the agricultural FOIs is very rare).

Table 13: Results from the detection of the built-up clusters within FOI, using the unsupervised segmentation method (Catalonian sample)

Detection of built-up clusters		**Automated method**		**Sub-population**
		Detected	***Not detected***	
Actual situation	***Present***	2	1	3
	Not present	66	296	362
	Sum	68	297	365

The more important in this case was the typology of the detected clusters found within the flagged FOIs. There were indeed related to: (1) wide intra-row spacing on permanent crops (revealing the underlying bare soil); (2) bare (probably eroded) spots in parcels with annual crops or fallow land, located on relatively poor soil and at least 5% slope; (3) landscape features (terraces); (4) small impoundments, temporally inundated. All these objects are inherent to the landscape, and should not be considered as "alien" to the FOI. Figure 43 below provides an overview of the typologies found.

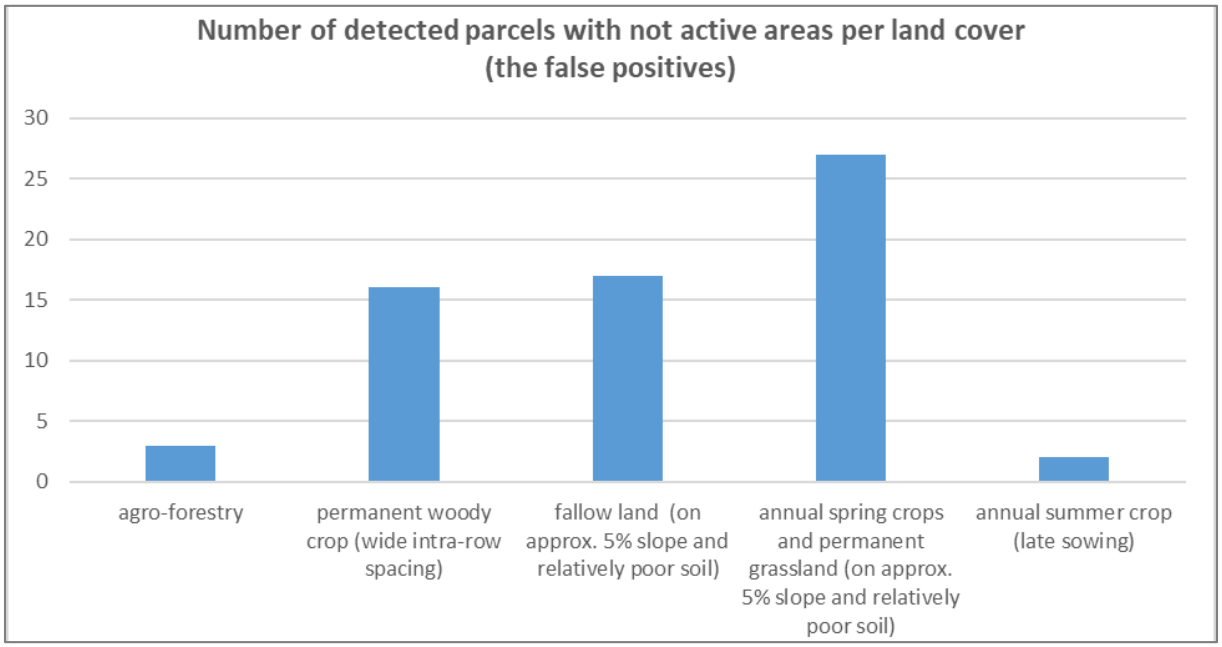


Figure 43: Number of detected FOIs with more than 25% of "non–active" agricultural areas inside (Catalonian sample)

As evident from the soil data, the whole test area resides on soil types, such as calcisols, gypsisols and regosols, rather typical for the Mediterranean and arid climates (Figure 44). The inland area of Lleida, where zone NOUR is located, is characterized with Mediterranean climate, which is slightly more arid and continental than the one on the coastline. There is presence of groundwater

bodies[5], which creates the conditions for precipitation of gypsum and carbonates from the soil solutions and their evaporation at the surface. Depending on the degree of accumulation of the gypsum and calcium carbonate in the upper soil horizons, there could be adverse conditions for water and plant root penetration; thus poor crop development in certain areas of the parcel. Regosols are also common on sloping land, subject to erosion. The terrain variations are also considerable, with average slope of 6.3%.

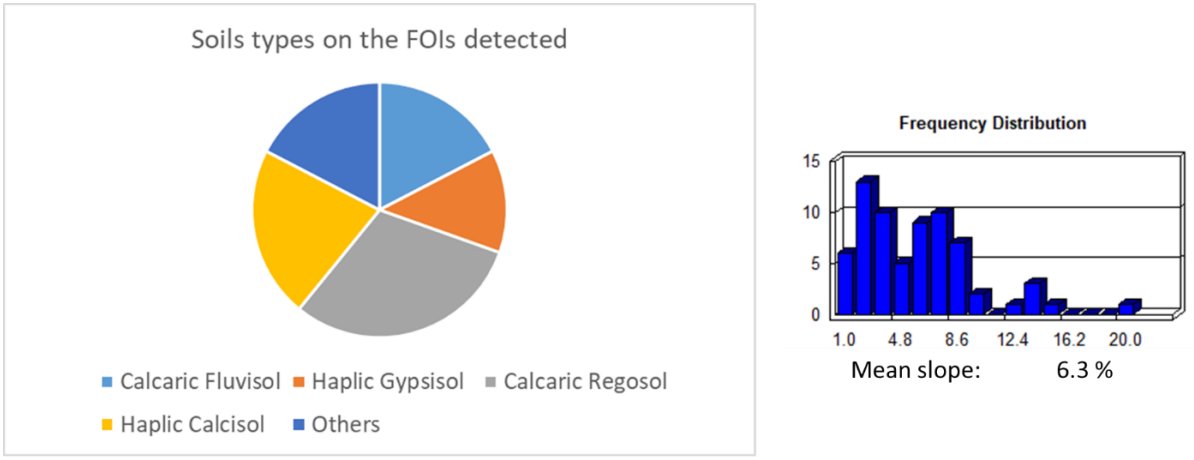
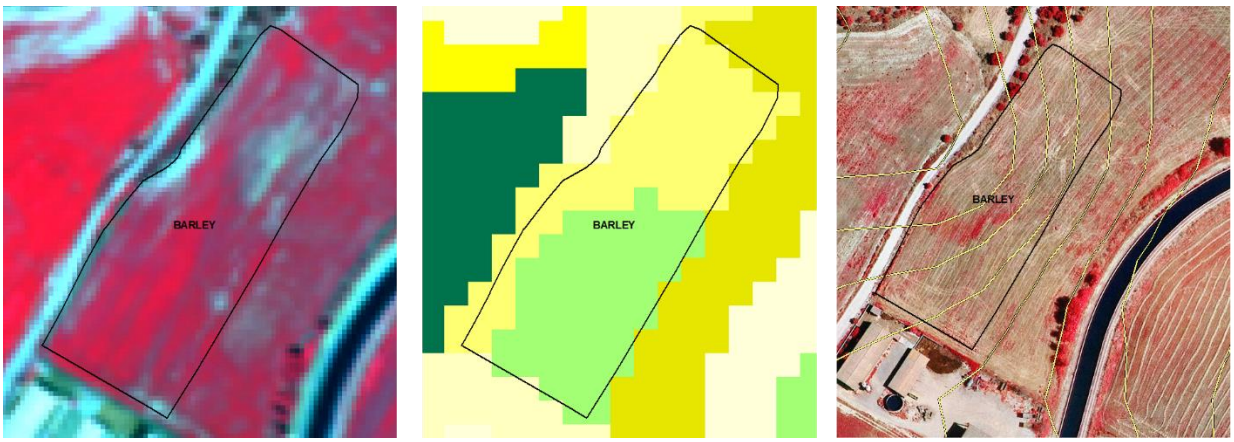


Figure 44: *Test site in Catalonia: (a) Distribution of the soil types and (b) frequency distribution of DEM slope [%] in FOIs detected with non-active areas*

As seen, the extracted qualitative and quantitative data on the persistently abiotic objects within the FOI could provide valuable information on the agricultural and environmental conditions of the agricultural parcels in the test area. The stronger brightness from soil could be associated occasionally with its higher gypsum and calcium carbonate content. It could provide an indication of the water retention capacity and productivity in general. When combined with soil and terrain data, cluster occurrences seem to correspond to the area on the parcel where water erosion processes are expected (Figure 45). Last but not least, in permanent crops, the zones with stronger contribution of the bare soil to the overall reflectance could provide an indication of wider intra-row spacing applied or poor/immature development of the tree crop.



[5] https://doi.org/10.1080/07900627.2014.938260

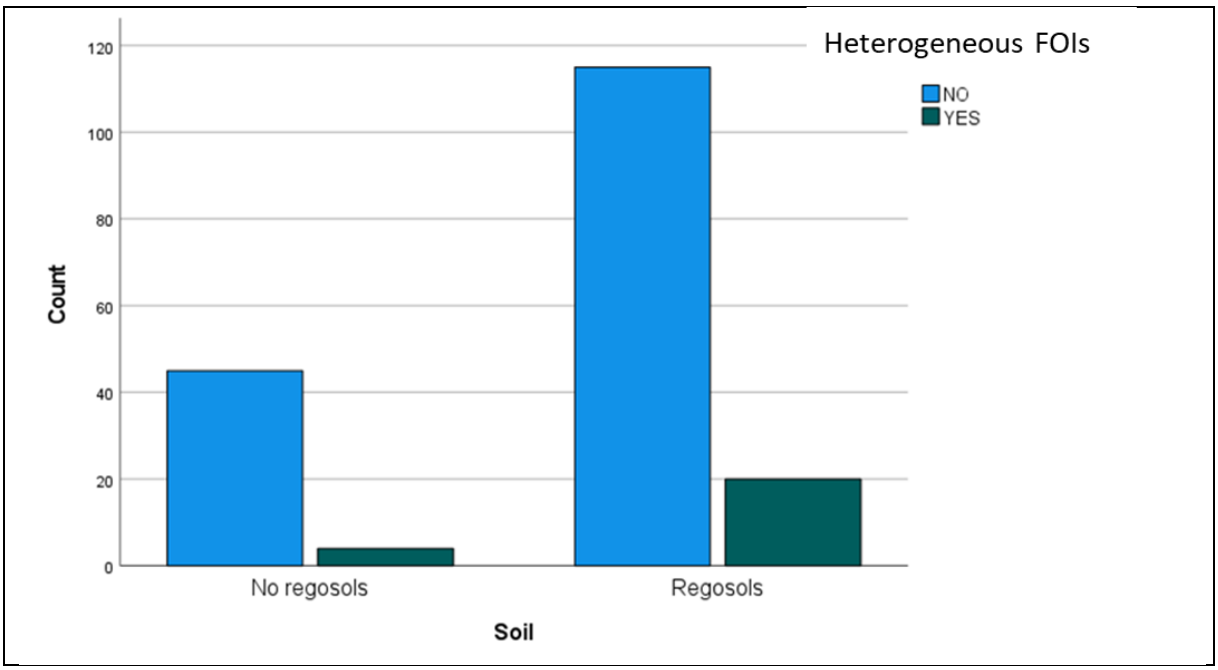
Figure 45: Example of non-uniform vegetation cover within parcel with barley:(a) Bare and probably eroded spots, in light blue, are clearly visible on the VHR Imagery 2019; (b) Thematic raster file, where the sparsely covered upper parcel area is separated in another segment; (c) Orthophoto from 2020, with DEM contours overlaid, showing stepper slope present in the Northern parcel area (mean slope is 8.2%). The soil type is Calcaric Leptic Regosols.

The project team made some further attempts to investigate for possible relations between the FOI heterogeneity detected and the ground conditions related to soil and slope. The first analysis (Table 14), using chi-square test, was to check whether there is statistically significant difference between the abundance (frequency) of the detected heterogeneous FOIs on regosols and other (non-regosols) soil types. Only FOIs on arable crops were selected in the population, for which we could expect uniform crop coverage in ideal terrain conditions.
The second analysis, using non-parametric Mann–Whitney U test, was to check whether there is statistical significant difference between the slope distribution for two groups of arable FOIs:

- flagged as homogeneous (FOI_hetero=0)
- flagged as heterogeneous (FOI_hetero=1)

Neither the first (*Table 14*), nor the second analysis (*Table 15*) revealed any particular relation between the soil type/slope and the FOI heterogeneity detected. The p-values given below are well above the alpha of 0.05. The reason partly could be explained by the coarse resolution of the soil map (1:250 000) and elevation data (25m grid, Vertical RMSE > 7 meters) used. With relation to soil data, it would be more relevant to use the particular morphological characteristics, such as the mechanical content and structure, rather than the soil classes themselves.

Table 14: Analysis of the possible relation between FOI heterogeneity detected and the soil type, using Chi-square test (in SPSS)



Chi-Square Tests

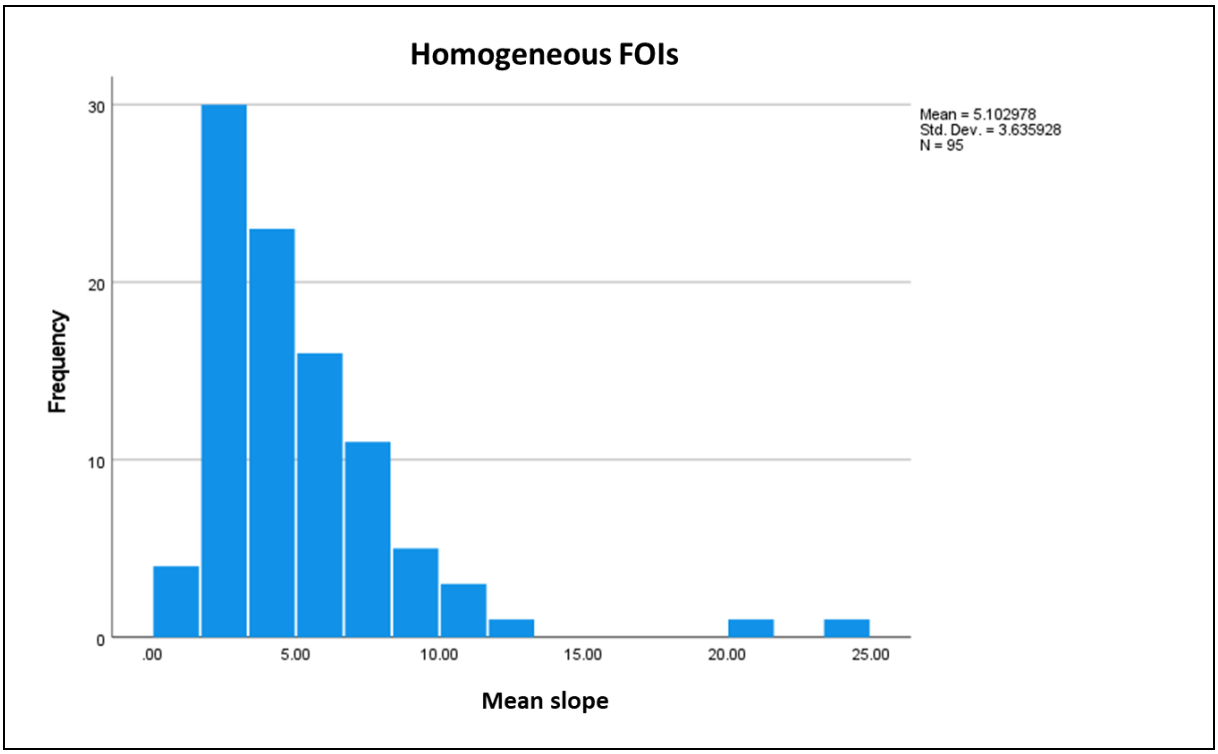
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.402[a]	1	.236		
Continuity Correction[b]	.877	1	.349		
Likelihood Ratio	1.525	1	.217		
Fisher's Exact Test				.324	.175
N of Valid Cases	184				

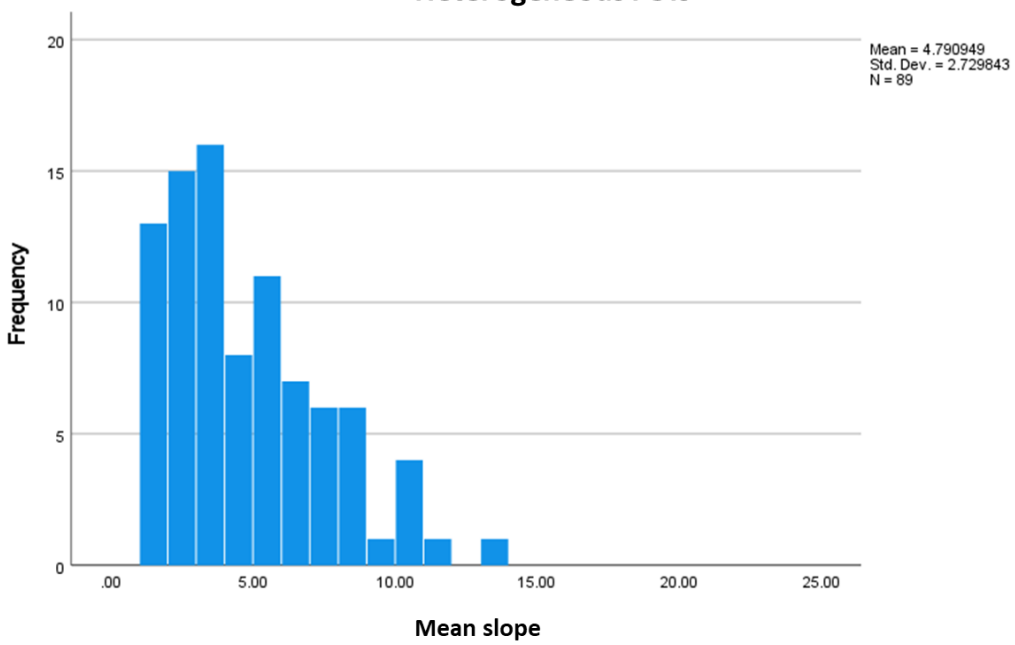
a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.39.

b. Computed only for a 2x2 table

P-value

Table 15: Analysis of the possible relation between FOI heterogeneity detected and the slope distribution, using Mann–Whitney U test (in SPSS).





Ranks

	V1	N	Mean Rank	Sum of Ranks
V2	1	95	93.34	8867.00
	2	89	91.61	8153.00
	Total	184		

Test Statistics[a]

	V2
Mann-Whitney U	4148.000
Wilcoxon W	8153.000
Z	-.222
Asymp. Sig. (2-tailed)	.824

a. Grouping Variable: V1

P-value

ANNEX I Initial heterogeneity tests on simulated backscattering data

Already in the beginning of the FOI assessment project, the team found that in order to understand how to "read" properly the outputs of the S1 gamma distribution method and adjust its initial parameters, it should conduct some tests on simulated data.

The test consisted of an array of 100 equal rectangular clusters of pixels, organized in 10 rows and 10 columns. Each of the clusters was divided in different sub-blocks. The division was the same in all clusters. Different gamma distributions were applied in each of the blocks of the cluster.

Ten different scenarios were prepared, with different values for the following test parameters:

- Number of pixels in each block (2x2, 4x4, 5x5, etc.)
- Number of blocks within cluster (2, 3, 4)
- Ratio between blocks in a cluster (50/50; 25/75; 20/80; etc.)

The method made attempt to detect heterogeneity on each of them. The detection rule was that to confirm heterogeneity, at least on 95 of the 100 clusters (for alpha equal to 5%) should be detected.

The results obtained show that detection rate is function of:

1. The number of pixels in the block (information content)

2. The number and ratio between the blocks

3. The delta between the sig0 of the Gamma distributions (smaller delta would require larger amount of pixels in each block)

The preliminary assessment indicated that increasing the block size from 2x2 towards 20x20 with 50/50 block ratio, increases the detection success rate as well. Adding more blocks, seems to make the detection less sensitive due to the kind of "connection" built between the blocks, which creates a sort of continuum between the sig0 (i.e. from sig0=0.03 to sig0=0.05 then sig0=0.07 then sig0=0.1 etc...), while overall variance stays the same.

By running in future more and more simulations with different combinations of number of pixels (n1), ratio (n2) and with different number of blocks, we will be able to define the best conditions for which the S1 gamma distribution methods would be able to detect the heterogeneity.

2 blocs of 2x2 pixels
detection rate = 45%

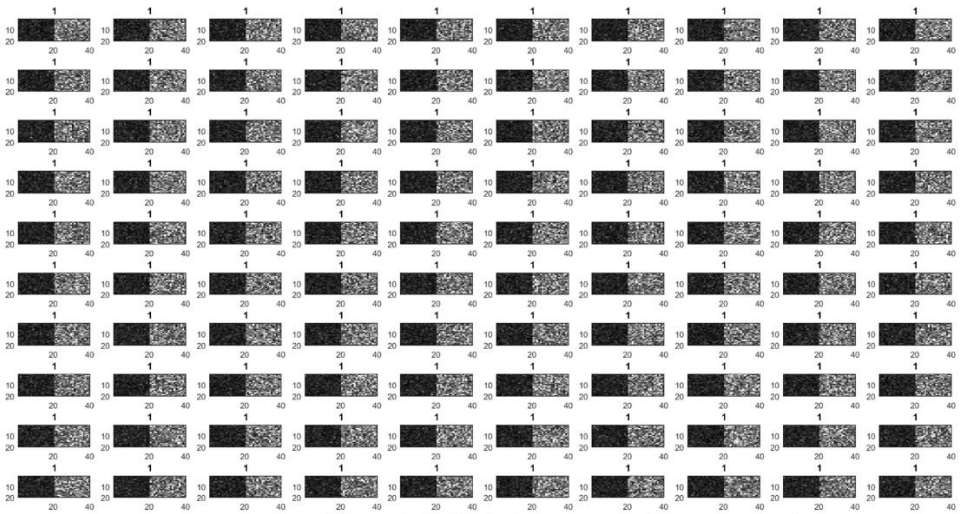
Gamma parameters:
k = 4.4 and sig0 = 0.03
k = 4.4 and sig0 = 0.1

2 blocs of 4x4 pixels
detection rate = 94%

Gamma parameters:
k = 4.4 and sig0 = 0.03
k = 4.4 and sig0 = 0.1

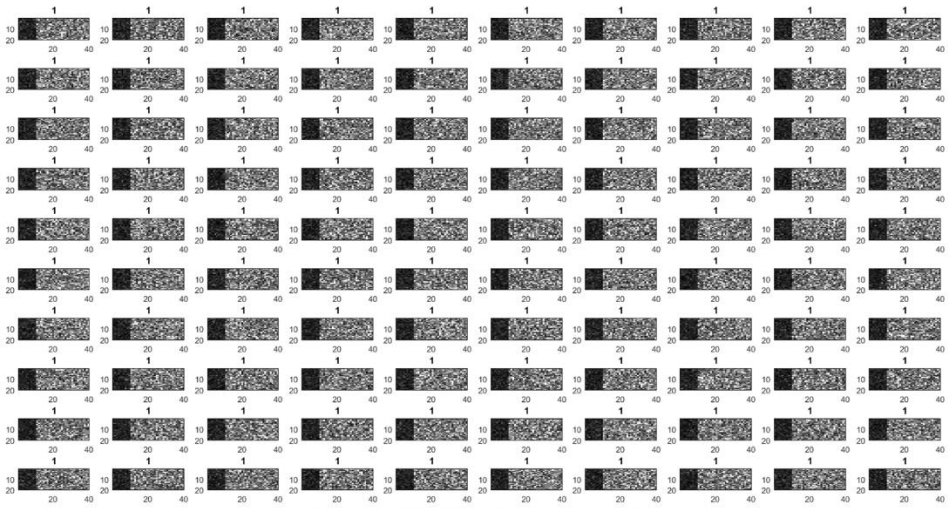
2 blocs of 5x5 pixels
detection rate = 99%

Gamma parameters:
k = 4.4 and sig0 = 0.03
k = 4.4 and sig0 = 0.1



2 blocs of 20x20 pixels
detection rate = 100%

Gamma parameters:
k = 4.4 and sig0 = 0.03
k = 4.4 and sig0 = 0.1

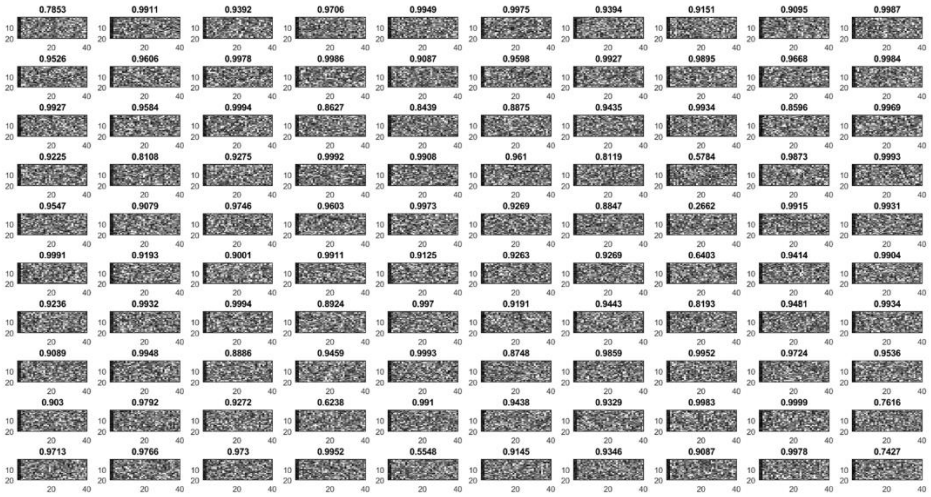


2 blocs of 10x20 and
30x20 pixels
detection rate = 100%

Gamma parameters:
k = 4.4 and sig0 = 0.03
k = 4.4 and sig0 = 0.1

2 blocs of 5x20 and
35x20 pixels
detection rate = 100%

Gamma parameters:
k = 4.4 and sig0 = 0.03
k = 4.4 and sig0 = 0.1



2 blocs of 2x20 and
38x20 pixels
detection rate = 52%

Gamma parameters:
k = 4.4 and sig0 = 0.03
k = 4.4 and sig0 = 0.1

3 blocs of 2x2 pixels
detection rate = 27%

Gamma parameters:
k = 4.4 and sig0 = 0.03
k = 4.4 and sig0 = 0.1
k = 4.4 and sig0 = 0.07

3 blocs of 4x4 pixels
detection rate = 87%

Gamma parameters:
k = 4.4 and sig0 = 0.03
k = 4.4 and sig0 = 0.1
k = 4.4 and sig0 = 0.07

4 blocs of 2x2 pixels
detection rate = 48%

Gamma parameters:
k = 4.4 and sig0 = 0.03
k = 4.4 and sig0 = 0.1
k = 4.4 and sig0 = 0.07
K = 4.4 and sig0 = 0.13

ANNNEX II: Structure of the reference data of NOUR for checking the FOI methods and conduction of the accuracy assessment (revision based on the decision from 19 June 2020)

Name: NOUR_DUM_2019_FOI_March_July_QA_365_card_final_21062020.shp

Number of Item: 365

Format: ESRI SHP file

Structure of the attribute table

OBJECTID	First type of FOI ID taken from the original file (the 2019 GSAA)
OBJECTID_1	Secord type of FOI ID taken from the original file (the 2019 GSAA)
PROD_NOM	The name of the crop in Catalan as recorded in the 2019 GSAA
Land_Use	The name of the crop in English as recorded in the 2019 GSAA
Land_Cover	Name of the generic agriculture land cover category to which the given crop belongs to
VHR_use	Boolean flag indicating when VHR imagery was used to confirm the S2 decision or INDEPENDENTLY from S2 (in case S2 data was found not sufficient to come to a decision) – approach 1 (S2 OR VHR)
FOI_hetero	Boolean flag indicating presence of heterogeneity (1 if present, 0 if not, and 2 if decision cannot be made)
FOI_Cardin	Boolean flag indicating presence of cardinality problem (1 if present and 0 if not, and 2 if decision cannot be made) – approach 1 (S2 OR VHR)
FOI_hete_n	The numbers of useful Sentinel 2 images used to come to the decision for presence of heterogeneity (numbers are sequential from March to July; roughly 4-5 images per month)
FOI_homo_n	The numbers of Sentinel 2 images used to come to the decision for presence homogeneity (numbers are sequential from March to July; roughly 4-5 images per month)
FOI_Card_n	The numbers of Sentinel 2 images used to come to the decision for presence cardinality problem (numbers are sequential from March to July; roughly 4-5 images per month)
FOI_Type	Type of FOI cardinality, according to the legend provided in the document FOI_cardinality_typology.docx - approach 1 (S2 OR VHR)
Observ	Observations made during the QA inspection
SHAPE_Le_2	Perimeter in meters
SHAPE_Area	Area in square meters
FOI_Card_2	Boolean flag indicating presence of cardinality problem (1 if present and 0 if not, and 2 if decision cannot be made) – approach 2 (S2 AND VHR)

VHR_USE_2	Boolean flag indicating when VHR imagery was used ONLY AS COMPLEMENTARY to confirm the S2 decision AND NOT INDEPENDENTLY (in case S2 data was found not sufficient to come to a decision) – approach 2 (S2 AND VHR)
FOI_Area	Type of FOI cardinality detected, according to the agreed approach during the meeting on 19.06.2020. Codes are: 0 – no cardinality problem; 1 – problem with non-agricultural area present; 2 – problem with more than one agricultural land cover category present; 3 – problem with more than one crop group present. NOTE 1: However, the "10 pixels in cluster" rule is applied. NOTE 2: 3 crop groups are defined: winter crops; summer crops; fallow land and annual EFA
FOI_Card_3	Boolean flag indicating presence of cardinality problem (1 if present and 0 if not) – approach 3 (according to FOI_area input).

Inspection protocol for P1

1. Prepare sample of 365 items (FOI representation from GSAA)
    a. Select them randomly
2. Prepare reference data
    a. structured array of Sentinel imagettes overlaid with the GSAA geometry
    b. VHR Imagery, mostly to confirm a phenomenon already detected on Sentinel
3. For each item in the sample
    a. Verify if item can be inspected with the available reference data. Check:
        i. Whether same FOI has not already been inspected for the same phenomena
        ii. For presence of any image artifacts, which prevents the inspection of the item
        iii. If meaningful CAPI is possible (ex. land is not flooded)
    b. If the inspection if not feasible, flag the parcel as skipped (code 0 in the "FOI_Card" field, report the reason for skipping in "Observ" field, and go to the next item.
    c. Else, proceed with step d.
    d. Check visually for presence within the GSAA of
        i. two or more different crops/agricultural land cover (AL, PG, PC), or
        ii. unaccounted non-agricultural area
    e. on a sequential number of imagettes/observations 5 in a row) on temporal profile and from the following period
        i. for arable winter crops – March-May
        ii. For grasses and permanent crop – March-July

f. If found, flag the given item as having cardinality problem (FOI_Card = 1)

g. Else, flag the item as being valid (with FOI_Card = 0), OR that decision cannot be made (FOI_Card+2)

h. Fill-in other relevant fields with the complementary information required.

4. For those FOIs with area >= 4000 m2. and FOI_Cardin = 1

   a. Check visually where the different crops/land cover found represent contiguous clusters of equivalent areas of at least 20 Sentinel pixels (1 pixel = 10 meters)

   b. If found, flag the given item as having AREA cardinality problem (FOI_Area_c = 1)

   c. Else, flag the item as FOI_Area_c = 0

The accuracy assessment should follow the instructions given in CbM TG v5 (https://jrcbox.jrc.ec.europa.eu/remote.php/webdav/Documents%20(2)/CbM_QA/CbMQA_TG_v5.3.pdf), Chapter 6.2

The focus will be to check the performance of the FOI methods in capturing the FOI cardinality problem, as defined in FOI_cardinality_typology.docx. Thus, the field that need to be used from the reference data is "FOI_Cardin".

For each of the FOIs in the reference data (365 in total), compare the values given in the "FOI_Cardin" (0 or 1) with the results from your method (normalized to Boolean: 0 or 1).

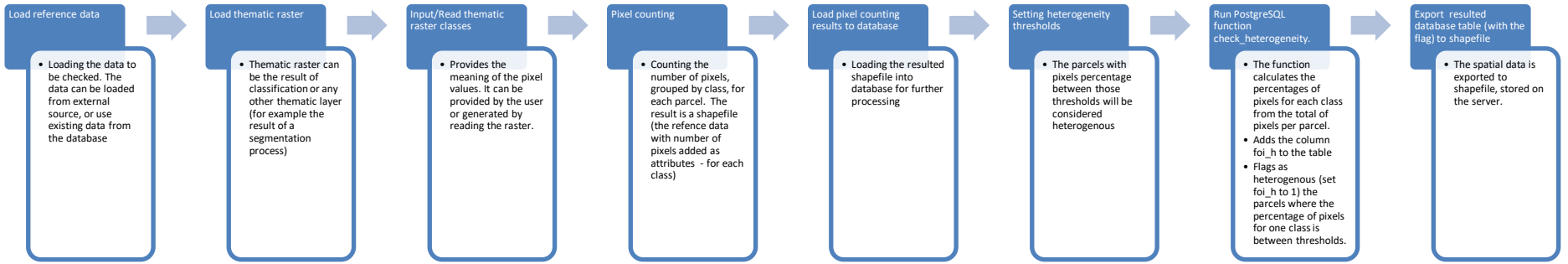
Summarize the results for all FOIs in the table given below (Table 2 of CbM TG):

FOI cardinality lot	**FOI cardinality problem detected by your method**	**FOI cardinality NOT detected by your method**	sum
FOI_Cardin Confirmed	true positive n_{11}	false negative n_{01} or β'	$n_{1.}$
FOI_Carding not confirmed	false positive n_{10} or α'	true negative n_{00}	$n_{0.}$
Sum	$n_{.1}$	$n_{.0}$	n

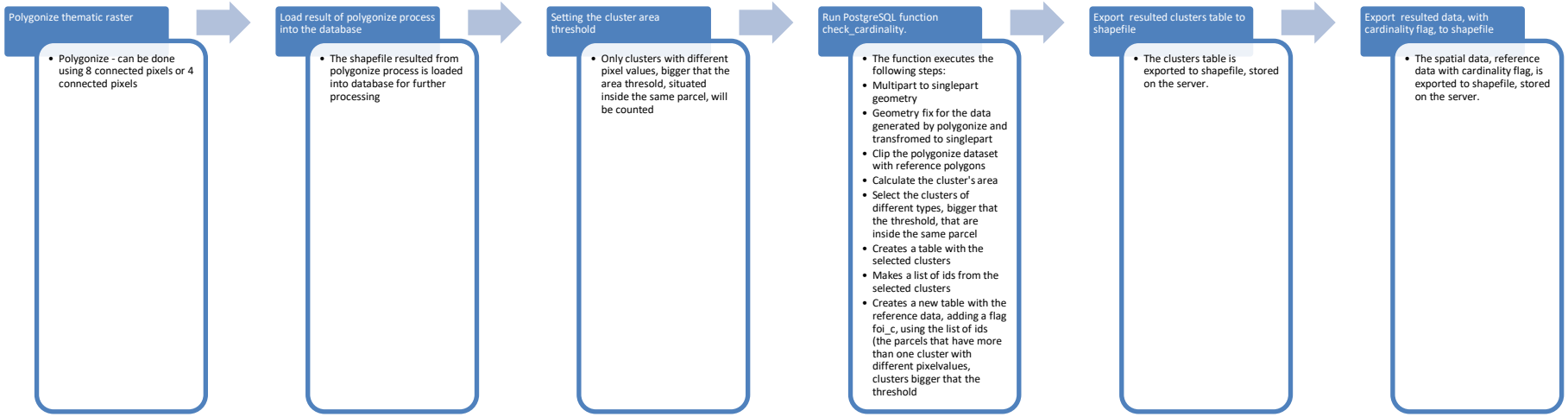
Finally, go through the testing flow given on pages 15 and 16 CbM TG v5 (https://jrcbox.jrc.ec.europa.eu/remote.php/webdav/Documents%20(2)/CbM_QA/CbMQA_TG_v5.3.pdf) to derive the relevant acceptance numbers and to compare them to your outcomes (false positive and false negatives).

ANNNEX III: FOI notebook processing workflow

FOI heterogeneity workflow



FOI cardinality workflow



References

A Comber, P Fisher, R Wadsworth. 2005. "What is land cover?" *nvironment and Planning B: Planning and Design* volume 32, pages 199 - 209.

Chauhan, S., & Srivastava, H. 2016. "Comparative Evaluation of the Sensitivity of Multi-Polarised SAR and Optical Data for Various Land Cover Classes." (International Journal of Advancement in Remote Sensing, GIS and Geography) 4, 1–14.

Harfenmeister, K., D. Spengler, and C. Weltzien. 2019. "Analyzing Temporal and Spatial Characteristics of Crop Parameters Using Sentinel-1 Backscatter Data." *Remote Sensing* 11, 1569. doi:https://doi.org/10.3390/rs11131569.

Hartigan, J. A. Hartigan. P. M. March, 1985. "The Dip Test of Unimodality." *Ann. Statist.* 13 (1) 70 - 84. doi:https://doi.org/10.1214/aos/1176346577.

JRC. 2020. *Checks by Monitoring Quality Assurance Framework, ANNEX III, Identifying the CbM decisions, version 1.1.* Technical Report, Joint Research Centre, Ispra: European Commission.

JRC. 2015. *Technical Guidance for Management of Layers in LPIS.* Technical Guidance, European Commission: Joint Research Centre.

Luketić N., Borio D., Devos W., Milenov P. 2020. "Technical Guidance on Quality Assessment of the CAP Checks by Monitoring." *JRC Technical Report.*

Milenov P., Devos P., Luketic N., Vajsova B., Loudjani P., Zielinski R. 2020. "Checks by Monitoring Quality Assurance Framework, ANNEX III - Identifying the CbM decisions, version 1." *JRC Technical Report.*

Nezry, E. 2014. "Adaptive Speckle Filtering in Radar Imagery. In Land Applications of Radar Remote Sensing." IntechOpen. doi:https://doi.org/10.5772/58593.

Roshan Pande-Chhetri, et al. 2017. "Object-based classification of wetland vegetation using very high-resolution unmanned air system imagery." *European Journal of Remote Sensing* 50:1, 564-576. doi:DOI: 10.1080/22797254.2017.1373602.

Schnuerch, M., Nadarevic, L. & Rouder, J.N. 2020. "The truth revisited: Bayesian analysis of individual differences in the truth effect." *Psychon Bull Rev.* doi:https://doi.org/10.3758/s13423-020-01814-8.

Tasdemir K, Milenov P, Tapsall B. 2012. "A hybrid method combining SOM-based clustering and object based analysis for identifying land in good agricultural condition." *Computer and electronics in Agriculture* (Elsevier Sci Ltd, England) (ISSN 0168-1699): pages 92-101.

Tomas, Blanka & Fasbender, Dominique & Wirnhardt, Csaba & Lemajić, Slavko & Devos, Wim. 2020. "Assessing Spatial Limits of Sentinel-2 Data on Arable Crops in the Context of Checks by Monitoring." *Remote Sensing* 12. 2195. 10.3390/rs12142195.

List of figures

List of tables

End of report

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications
You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub

ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub