European Commission

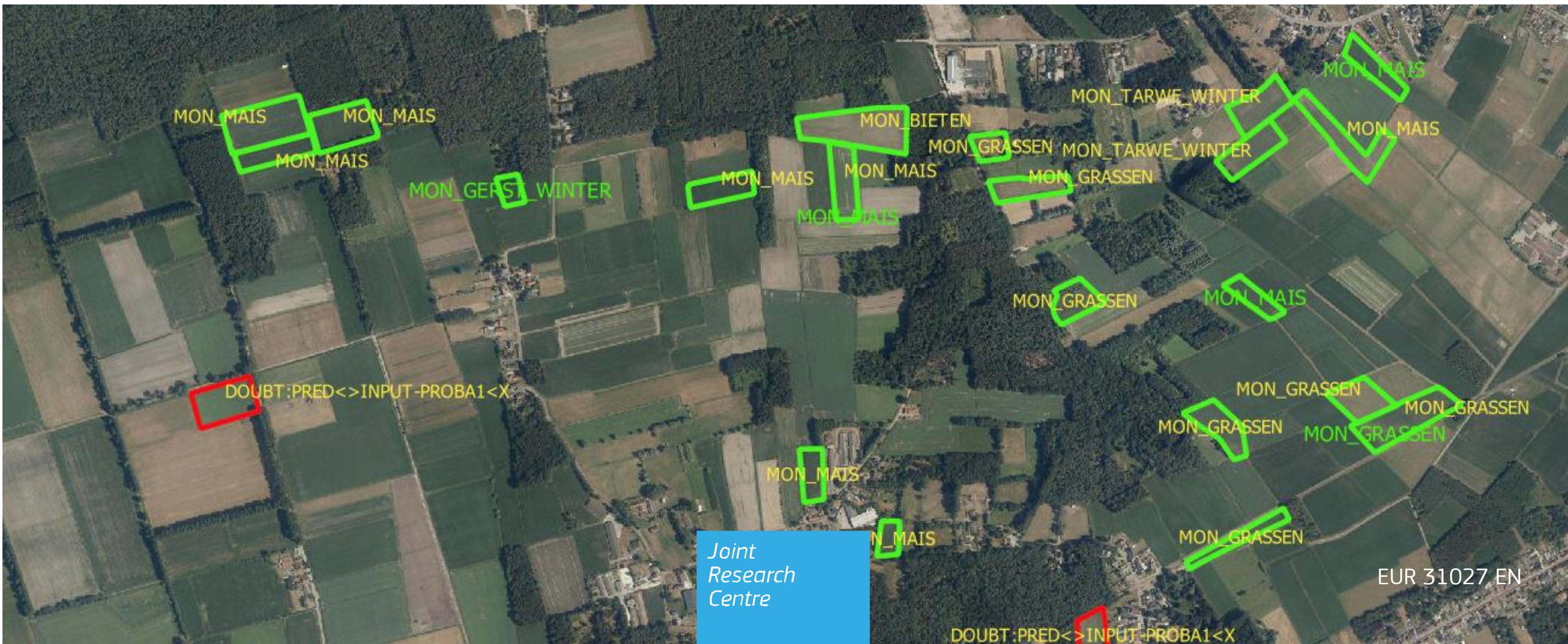
JRC TECHNICAL REPORT

Checks by monitoring Quality Assurance: Development activities

Summary of the work performed toward the CbM QA development

Luketic Natasa, Borio Daniele, Devos Wim, Sedano Fernando

January, 2022



Joint Research Centre

EUR 31027 EN

Contact information
Name: Wim Devos
Address: Via E. Fermi 2749, I-21027 Ispra/Italy
Email: wim.devos@ec.europa.eu
Tel.: +39 0332-789655

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

JRC128276

EUR 31027 EN

PDF ISBN 978-92-76-49632-8 ISSN 1831-9424 doi:10.2760/348125

Luxembourg: Publications Office of the European Union, 2022

How to cite this report: Luketic et al., *Checks by Monitoring Quality Assurance: Development activities*, EUR 31027 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-49632-8, doi:10.2760/348125, JRC128276.

Contents

Foreword

The Joint Research Centre (JRC) is the European Commission's (EC) science and knowledge service that supports EU policies with independent scientific evidence throughout the whole policy cycle.

The JRC's GTCAP (Geodata and Technologies for the Common Agricultural Policy) team is contributing to the development of methodological concepts and tools for the quality assessment of Check by Monitoring (CbM) systems adopted by Member States (MS).

These preparations of the methodologies and technical guidance was carried out through collaborative meetings and discussions with DG AGRI entities and MS administration experts. The operational setup of such collaboration takes the form of a quality assurance (QA) framework.

This document summarizes the activities and progress achieved in 2020 and 2021 by the GTCAP team on the CbM QA framework.

Acknowledgements

The authors wish to acknowledge the 5 Member States' Paying agencies for providing data and input.

The authors appreciate the support received from DG AGRI, which made possible to conduct needs assessments, to invite key attendees at meetings and workshop events, and to develop a policy relevant methodological document set.

GTCAP team participating in the methodological development were: Daniele Borio, Wim Devos, Dominique Fasbender, Nataša Luketić and Pavel Milenov. GTCAP colleagues supporting these activities were Laura Acquafresca, Paolo Isoardi, Philippe Loudjani, Fernando Sedano, Blanka Vajsova, Csaba Wirnhard and Rafal Zielinski.

Authors

Nataša Luketić, Daniele Borio, Wim Devos, Fernando Sedano.

Abstract

From early 2020 until the end of 2021, JRC's Geodata and Technologies for the Common Agriculture Policy (GTCAP) team developed the checks by monitoring quality assurance (CbM QA) methodology to address concerns expressed by the Member States (MS) and to follow a recommendation of the European Court of Auditors. During the two years, GTCAP set up dedicated activities to elaborate and frame this effort; these are described in this report.

Interaction of the CbM expert community was ensured by organising six plenary meetings, a dozen bilateral technical meetings with the MSs and one webinar. This collaborative approach brought together JRC scientists, MS quality managers, processing experts and DG AGRI officers to agree on purpose and scope of the development. The meetings also investigated the automated processing of the available Sentinel datasets in CbM as well as this data's potential for inspecting, testing and assessing the automated results.

The resulting CbM QA methodology offers a set of inspection tools and operational methods that streamline efforts to assist MSs to assure the robustness, transparency, reproducibility and veracity of their CbM systems. The methodology involves a three-step approach that consists of step 1: inspection of a sample from a homogeneous lot by visual image interpretation; step 2: eligibility check of errors found, categorisation per relevance and acceptance testing; and step 3: financial impact calculation of identified end-stage errors.

At the end of this report, we provide feedback from the first experimental implementation of the proposed CbM QA methodology provided by 5 MSs along with some proposals for improvements to accommodate all implemented CbM options.

This document summarizes the activities and progress achieved in 2020 and 2021 by the GTCAP team on the CbM QA framework.

1 Introduction

In 2017, the European Commission started to encourage the Member States' (MS) Paying Agencies (PAs) to use the potential of the EU-owned Copernicus Sentinel satellites for checks on the Common Agricultural Policy (CAP) payments. These new Sentinel satellites provide frequent high resolution optical and microwave data over the Globe.

This availability of frequent, wall-to-wall and free Sentinel data, drove JRC to initiate technical discussions and developments of high-level concepts, centred around the monitoring of agricultural activities and practices.

The CAP legislation was amended in 2018 with the basic idea to permit a substitution of the costly checks either by a single-date field visit or through the visual inspection of images, with automated checks based on Sentinel data streams. These can capture farmers' activities and provide timely information to the PAs. Moreover, to increase compliance and prevent possible sanctions, the amended legislative provisions counted on feedback to the beneficiaries on discrepancies between aid applications and detected farming practices. The beneficiaries could change the declaration data in their aid application before payment and this allows MS to operate at reduced overall cost and with lower administrative burden.

In 2019, five MSs notified the Commission on their decision for a CbM adoption and some of them initiated a gradual implementation, the so-called phasing-in. Phasing-in allowed a progressive adoption of technology as well as an incremental inclusion of parts of their territories.

A year after, in 2020, European Court of Auditors (ECA) carried out an audit and issued the special report([1]) "Using new imaging technologies to monitor the Common Agricultural Policy". This report formulated recommendations to the EC, including a call for completing developments on the CbM quality assessment framework (CbM QA) by the end of 2021.

By early 2020, JRC had already started to develop a concept for CbM QA that took into account the general aspects of quality assurance. This set up adhered to ISO standards for setting up the limiting quality, the sampling scheme and the acceptance criteria. Basic governance expectations and practical experience from the existing Land Parcel Identification System (LPIS) QA framework were also considered. This report documents the developments and activities performed by JRC's GTCAP group on the CbM QA methodology during 2020 and 2021. The detailed description of the resulting CbM QA methodology can be found in the draft document set CbMQA v.1.2., available on WikiCap: https://marswiki.jrc.ec.europa.eu/wikicap/index.php/CbMQA_TG_v1_2

([1]) https://www.eca.europa.eu/Lists/ECADocuments/SR20_04/SR_New_technologies_in_agri-monitoring_EN.pdftext

2 Motivations and general principles

To develop an effective QA framework for CbM, one should look not only at the outputs from the system (i.e. checks), but also at the technical choices and input elements used to obtain that output. Primary quality expectations should thus be aligned with this duality in mind. The objective is to create a QA methodology that is cost effective and technically simple to implement regardless of the CbM design chosen by the administration.

That is quite challenging in particular since the checks are intertwined with two additional components of the Integrated Control Administration System (ICAS): the Land Parcel Identification System (LPIS) and the Geo-Spatial Aid Application (GSAA). Quality considerations of the GSAA and the LPIS datasets became the pre-conditions for CbM monitoring. As both provide input information to the CbM, their potential design weaknesses can propagate to the CbM results. However, the CbM QA involves a standalone output assessment for the CbM system . The other two (GSAA and LPIS) are validated in separate exercises.

In addition to identifying issues that can be remediated, the CbM QA also offers a framework that enables the MS administrations to report to the Commission about the CbM performance as control mechanisms. This common methodology thus ensures an objective and comparable reporting from all MSs.

The general principles of the QA include three main sub-elements: 1) inspection, 2) assessment and 3) the evaluation carried out on a sampled data set.

1. Inspection deals with observing of the evidence of the actual reality ("truth") on the QA sample, and comparing it with the outputs from the system. In the specific case of the CbM QA, that evidence can be sourced independently from the CbM system or it may re-use the Sentinel imagery. The choice remains on the administration. Still, an independent procedure of observation is required to conclude on the nature and the type of the phenomena visible from the evidence data source. This inspection records findings as observed in the physical reality.
2. The assessment deals with the context of those findings in relation to the specified system requirements and business expectations. Hence, this part provides a final compliance conclusion of the mismatches found during the inspection.
3. In the end, evaluation regarding "fitness for purpose" should be done by the MS administration based on the assessment results. CbM could bring many benefits to the MS, but for the Commission, the financial impact stands out. An evaluation should involve an objective identification of the weaknesses found and remedial actions planned to eliminate them.

These general QA principles have to fit into a specific workflow of the CbM QA which we divided into three steps following the basic requirements of the developed QA methodology (described in Chapter 3):

- to measure the performance of the automatic detection system based on satellite Earth observation data,
- to measure the performance of the eligibility requirements and CAP compliance,
- to quantify the residual risk in area and financial terms.

The proposed three steps are implemented sequentially, so that the relevant results (errors found) from the step 1 are processed in the step 2 and further to step 3, narrowing processing time for the final quantification and thereby focusing only on relevant items.

Such a method has its limitations when it comes to the assessment of the elements which, due to their nature or technological immaturity, cannot assign a conclusive outcome by the automatic detection system. But at this early phase, the goal is to cover and to inspect the "biggest" part of the system and to deal with the rest after the first QA experiments when the inconclusive numbers surface up.

Apart from reporting requirements at the very end, the inspection method does not make any distinction between different aid schemes. The assumption is that, after a correct detection of a particular phenomenon as observed by the satellite data, the detection subject will be put into the context of the aid schema and its eligibility requirements, operating a final traffic light for the payment.

3 Methodology

This chapter provides an overview of the methodology developed which details the procedure on how to inspect, assess and evaluate the quality of the CbM system decisions, guidance on sampling principles, testing methods and acceptance criteria.

In particular,

- step 1 consists of inspecting the operation of the automatic algorithms for detecting specific phenomena by interpreting the Sentinel data (or equivalent).
- step 2 processes the output from the first step (Sentinel detection errors) and from other inspection data to eligibility cases. Inspection errors are linked to the corresponding agricultural parcels that quantify a distinct area for a particular scheme, allowing an assessment of the eligibility errors.
- step 3 combines the eligibility errors of the second part with the payment rates and holding rules, allowing an evaluation of the financial performance of the CbM operations.

3.1 Identifying the items for inspection (step 1)

In the first step of the QA process, a procedure is established to analyse the detection performance of Sentinel-based decisions. CbM relies on the detection of the presence of a specific phenomenon in the field following farmer's activities. The decision within the process is performed on a feature of interest (FOI ([2])) and it concerns a specific marker (observed behaviour) on the land unit representing that FOI. For example, CbM can be used to detect ploughing. In such marker([3])-based process, the detection of a specific target (phenomena) holds a binary output: positive detection or negative detection per FOI. In more complex detection processes, there may be a series of marker queries that are run sequentially targeting different types of information extraction. For example: a marker confirming ploughing, a marker confirming harvest and a machine learning-based crop classification marker may in combination evidence a specific crop eligible for Voluntary Coupled Support (VCS) aid scheme (e.g. tomato or rice). Hence the three markers together would conclude on the compliance for that specific declaration. However, each of the three marker queries generate separate observations on: 1) ploughing activity, 2) harvest activity, and 3) crop class, and each should be analysed separately for system diagnostic reasons.

For this reason and for practical handling, it makes sense to first identify and meaningfully group the information extraction algorithms that trigger decisions generated by the automatic CbM detection into a homogeneous type([4]). That type may correspond to individual marker-based decisions or a series of signal-based analytics (depending on the system design). Each identified type of information extraction will hold a dedicated lot of operational decisions for inspection, which is used for the sample pre-selection.

Since the lot is formed on the information extraction type, the actual item for inspection is the decision made on a FOI using the marker query during the season. I.e.

- whether a season offers image series to analyse (e.g. 100+ images), only one (final) decision will be inspected,
- if three independent marker queries operate on the FOI during the season, three decisions will be inspected.

In the CbM concept, such decision is operated by a switch. The output of switches forms the item of inspection in this methodology.

3.2 Sampling

For a large lot of size N, only a small sample of size n will be inspected during the CbM QA process. The lot is expected to contain two types of items: positives where the phenomenon actually occurred and negatives where the phenomenon did not take place. For example, an automatic detection process would face FOIs where either ploughing occurred or not. In those cases, when ploughing actually occurred (positive case), the automatic

([2]) See list of abbreviations and definitions on page 24.
([3]) See list of abbreviations and definitions on page 24.
([4]) For further description on the generic type of information extraction, please see Annex III.

detection process can commit an omission error (false negative). When ploughing did not occurred (negative case), a commission error (false positive) can eventually occur. Given the random nature of the sampling procedure, it is not possible to predict the exact numbers of items in the sample where the phenomenon is supposed to occur and not occur. To remain practical, some assumptions are needed and while the resulting sampling process could lead to a situation where one of the two subpopulations (of sizes N_0 and N_1 respectively) is underrepresented, this is only an apparent paradox. Justifications for using a single sample with paired observations are provided in Annex V and in the CbM QA discussion document (Devos at all, December 2020). To obtain a standard reliability of accept/reject verdict of the inspection lot, a value $n \leq 365$ is sufficient.

The sample size, n, is provided in Table 1 as a function of the population size (N). N represents the total number of decisions (FOIs subject to the marker query in the season), regardless whether the underlying phenomenon actually occurred or not. The table is derived in Annex V. of the QA methodology.

Table 1. Tabulated values for the sample size (n), as a function of the population size (N).

N	[1-124]	[125-199]	[200-399]	[400-2100]	> 2100
n	N	125	200	315	365

3.3 Testing

Each sample is inspected and tested for acceptance based on the ISO 2859/2 standard, in order to respond to the following criteria:

- Is CbM detection able to process a full population (N) with a number of type I errors (false positives) lower than Lq·N? Where limiting quality (Lq) is the quality limit set to 10%.
- Is CbM detection able to process the full population (N) with a number of type II errors (false negatives) lower than Lq·N? For this criterion, the limiting quality limit was set to 10%, as well.

LQ 10 or limiting quality 10 means that a lot with an error rate of 10% or higher, should not have a reasonable chance to pass the test and be considered acceptable.

As already discussed, a false positive (type I) error occurs when a marker/behaviour is erroneously detected as present. Conversely, a false negative (type II) error occurs when a marker/behaviour is erroneously detected as not present.

Details on the tests developed and on the decision thresholds are provided in Annex V, while the acceptance numbers for Lq 10% are provided the Table 5 of the QA methodology.

The final state of the detected and inspected results can be expressed in a confusion matrix:

Table 2. Confusion matrix in a paired lot

Paired lot		CbM values: CbM detected	CbM values: CbM not detected	sub-population sizes
Actual values	**True presence**	true positive N_{11}	false negative N_{01} or β	N_1
	True absence	false positive N_{10} or α	true negative N_{00}	N_0
sum		$N_{.1}$	$N_{.0}$	N

The inspection of the CbM decision is carried out per parcel/FOI, and yields a binary "detected/not detected" result applied to every single inspection. The inspection is applicable for the decisions of any lot determined as G1, G2, T1, T2, T3, T4 or C1([5]).

The inspection itself is done by means of visual image interpretation in combination with pre-defined tools for spatio-temporal image enhancement. It is an independent "blind check" interpreting the presence of a particular phenomenon (as defined per lot type) without an a-priori information from the CbM detection switch's output.

Such marker-based inspection routine implies interpreting the actual phenomena observed from the suitable imagery, checking it against the comparable class as declared in the GSAA. However, there could be cases in which the predicted class could be different from both the declared and found one. This is commonly the case of crop type mapping systems (e.g. ML-based classification algorithms). In these cases, the mismatch between predicted, found and declared crop types, cannot be captured by a binary output, as both declared and predicted classes could be erroneous and different from the one found during the QA process. Further details on the proposed methodology for non-binary classification algorithms can be found in chapter 6.3 of the QA methodology.

3.4 Eligibility check (step 2)

While the tests developed in step 1 allow to investigate the quality of the Sentinel-based detection process, they do not take into account their relation to the eligibility requirements of the declaration data within the assigned aid scheme. This is done in the eligibility check – processing all false positive and false negative errors for each lot at the system level.

In step 2, the methodology introduces the concepts of abatable/end-stage errors based on eligibility criteria.

To categorise/label detection error found in step 1 into either abatable or end-stage error depends under which aid support scheme a given inspection item was declared.

- **Abatable errors** are those detection errors that were affected by eligibility rules of a given aid scheme and may result in a claim reduction. These errors are expected to be corrected by the beneficiary as they are disadvantageous for her/him provided the beneficiary was timely informed.
- **End-stage errors** are those detection errors that were affected by eligibility rules of a given scheme and don't results in a claim reduction. Those errors represent an undue payment in direct financial loss for the payment Fund.

([5]) See Annex III.

Moreover, in the compliance context, a lot holds a binary behaviour of every phenomena detection in relation to the occurrence scenario:

- **manifestation scenario:** that requires the manifestation of the behaviour/ marker(s) detected in order to confirm payment eligibility (e.g. for compliance “green light” switches).
- **absence scenario:** that expects the absence of manifestation of the behaviour/marker(s) in order to confirm payment eligibility (e.g. for non-compliance ”red light” switches).

These scenarios may be defined per parcel/FOI, or per type of information extraction, again depending on the system design. For example, a manifestation scenario of ploughing activity is expected on all FOIs which have been declared as one of the crops on the arable land, while absence scenario of ploughing is expected on all FOIs which have been declared as permanent grassland.

In step 2, only errors found in the step 1 are further processed, where depending on the manifestation/absence nature of the scenario and its relevance can be further divided as abatable or end-stage errors.

In such a framework, the eligibility assessment of the errors is deemed simplified and should not impose any significant additional workload. Furthermore, in a fully matured CbM-system, the farmer feedback loop should allow for the removal of all abatable errors during the campaign.

Since the eligibility check is done per declared parcel, one should make sure that the final quantification of the end-stage errors represents a unique information per declared parcel. That means that duplicates are eliminated to prevent double counting before the analysis and testing of each error.

Step 2 is carried out at the aid scheme level and combines results from multiple step 1 lots. In this respect, a parcel could lead to an error in several step 1 lots. When these lots are combined at the scheme level, a parcel should be counted only once even if it has led to several step 1 errors. Testing of finally identified end-stage errors should be below the 5%([6]) of limiting quality per aid scheme. Acceptance numbers for Lq 5% are provided in the Table 6 of the CbM QA TG.

3.5 Financial impact (step 3)

Step 3 evaluates the financial risk of the CbM by quantifying the end-stage errors identified per support scheme. For each parcel leading to an end-stage error, the area for which an undue aid was paid is determined. The total area erroneously paid is calculated as the sum of the contributions from the different end-stage errors. This area is converted into the undue payment amount through the average payment rate of the scheme under analysis. The final financial impact is estimated as the ratio of the undue payment amount and the total amount considered in the samples analysed for the specific scheme as identified in step 2. Additional details on step 3 can be found in Section 7 of the CbM QA TG.

3.6 Boundary checks

In addition to the three CbM QA steps that address the key aspects of the CbM automated process, the QA procedures also includes boundary conditions as validity tests. These elements are boundary conditions because they deal with the FOIs that, for deliberate or other reasons, do not come to a correct conclusion by the automatic processes. As the QA process directly depends on the good LPIS/GSAA quality, some elements of CbM testing are designed to assess the consistency/cardinality between the process-derived decisions and the reality in the field: this is the role of the boundary condition P1. A first boundary, denoted as P1, deals with area management and aims at verifying the correctness of the location and area of the FOI as declared by the farmer.

The P1 type test will be mandatory for all systems that do not integrate an explicit consistency/cardinality test (type G1 – spatial cardinality) within the overall CbM process. P1 assure that the area component is dealt with upfront and, for most scenarios, that the observations made on each FOI are meaningful.

([6]) 5% is agreed limiting quality for the CY 2019 and 2020 between Member States and the Commission for the eligibility check

The testing procedure aims at detecting matches between the FOIs represented by the GSAA declaration (sketch) and the Sentinel data. This test does not generate neither type I nor type II errors per se, but detects matches between the FOIs represented by the GSAA declaration and the Sentinel-derived FOIs. More information on spatial cardinality can be found in Annex III. For the P1 test, Lq equal to 10% is applicable. Errors found in the P1 test will not be considered for steps 1 and 2, but added for the estimation of financial impact calculation in step 3.

Other boundary conditions, such as those on the discriminatory power of the Sentinel-based automatic extraction process are outside the scope of this methodology. In particular, there are cases where the Sentinel-based automatic extraction process is unable to take a decision with the information provided by the end of the season (non-monitorable aspects) or cases where FOIs are intentionally non-monitored (non-monitored cases). These cases that are not targeted by the CbM flow will end up in the “pool of non-conclusive parcels” and will be subject to non-Sentinel follow-up including sampling. A dedicated test, called “P2”, checks whether inclusion into that pool was appropriate.

3.7 Reporting

The outcome of the QA framework should involve an objective self-assessment of the functioning of the system itself. This reflection involves investigating weaknesses, both of the QA method and of the monitoring system itself. It will allow the optimization of the operational processes and to identify the subsequent appropriate remedial actions.

4 Chronology of the development process

A draft CbM QA methodology outline was presented in Brussels on 19 February 2020 at expert group meeting with the invitation to the CbM experts from 5 on-boarding MSs. Over following months JRC produced gradually more detailed documentation that allowed the experts from the 5 mentioned MSs to evaluate and test the proposed methodology. Their feedback was then embedded in the developing methodology. In general, the CbM QA exercise operated as a joint endeavor with many intermediate interaction points. The chronology of these interaction events is summarized in Table 3.

Table 3. Chronology of the major meetings and events leading to the development of the CbM QA methodology.

Date	Event	Participants
19 February 2020	Expert group meeting in Bruxelles – introduction of the CbM QA	JRC, DG AGRI, 5 MSs
29 April 2020	Dissemination of the CbM QA, Part 1. Sentinel based decisions, Technical Guidance - DRAFT	5 MSs
7th May 2020	Bilateral meeting	JRC, MS1
11th May 2020	Bilateral meeting	JRC, MS2
13th May 2020	Bilateral meeting	JRC, MS3
14th May 2020	Bilateral meeting	JRC, MS4
15th May 2020	Bilateral meeting	JRC, MS5
28th May 2020	Plenary meeting - CbM QA kick off	JRC, DG AGRI, 5 MSs
27th August 2020	Committee of direct payments – presentation of the ongoing CbM QA developments	JRC, DG AGRI, 5 MSs
27th August 2020	Dissemination of the CbM QA TG - DRAFT v.1.0	5 MSs
August - December	Exchange and revision of the CbM QA 2019 results	JRC, 5 MSs
5th November 2020	IACS webinar	JRC, DG AGRI, 27 MSs
9th December 2020	Dissemination of the CbM QA TG - DRAFT v.1.1	JRC, DG AGRI, 27 MSs
August ‘20 - March ‘21	Methodological experiment on MS1 2019 and 2020 data	JRC, MS1
23rd June 2021	Plenary meeting - Dissemination of the CbM QA TG - DRAFT v.1.2	JRC, DG AGRI, 5 MSs
7th July 2021	Follow up plenary CbM QA meeting	JRC, DG AGRI, 5 MSs
12th July 2021	Bilateral meeting	JRC, MS5
22nd Sep 2021	Bilateral meeting	JRC, DG AGRI, MS2
13th Oct 2021	Bilateral meeting	JRC, MS6
22nd November 2021	Plenary meeting - CbM QA 2021 kick off	JRC, DG AGRI, 11 MSs
17th December 2021	Bilateral meeting	JRC, MS5

The European Court of Auditors organized fact finding missions to two MSs and communicated their findings in May and June 2021. The Court did not formulate any objections.

The first draft of the technical guidance was produced in April 2020.

It served to help 5 CbM adopting MSs to set up and evaluate the quality assessment process, gain and share experiences, and point out weaknesses and possible improvements. Some elements of the first draft methodology presented were provisional, missing or incomplete. Nonetheless the draft methodology introduced a basic concept of the CbM QA inspection flow. The prime purpose was collecting feedback regarding feasibility, burden and relevance of the inspection activities done by MSs. That feedback was analyzed and where appropriate, taken on board in subsequent versions of the technical guidance.

The second (May 2020) and the third (June 2020) revisions of this draft were produced after the bilateral meetings with the 5 MSs.

The fourth draft version of the method was consolidated and named version 1.0 by August 2020. Then followed a December 2020 update (version 1.1.). and a version 1.2 from June 2021 which included feedback from actual results of the trial CbM QA inspections conducted by 4 MSs. This version brought in consolidated and fine-tuned acceptance criteria and limiting quality decisions for the "formal" 2020 implementation.

The core methodological instructions elaborated in the main text of the draft TG were complemented by six annexes with additional information, such as the rationale, detailed derivations and reporting templates, on the CbM QA process. These six annexes are:

- Annex I: Testing elements
- Annex II: Inspection workflow
- Annex III: Identification of CbM decisions
- Annex IV: Ground data
- Annex V: Statistical background
- Annex VI: Reporting data package documentation

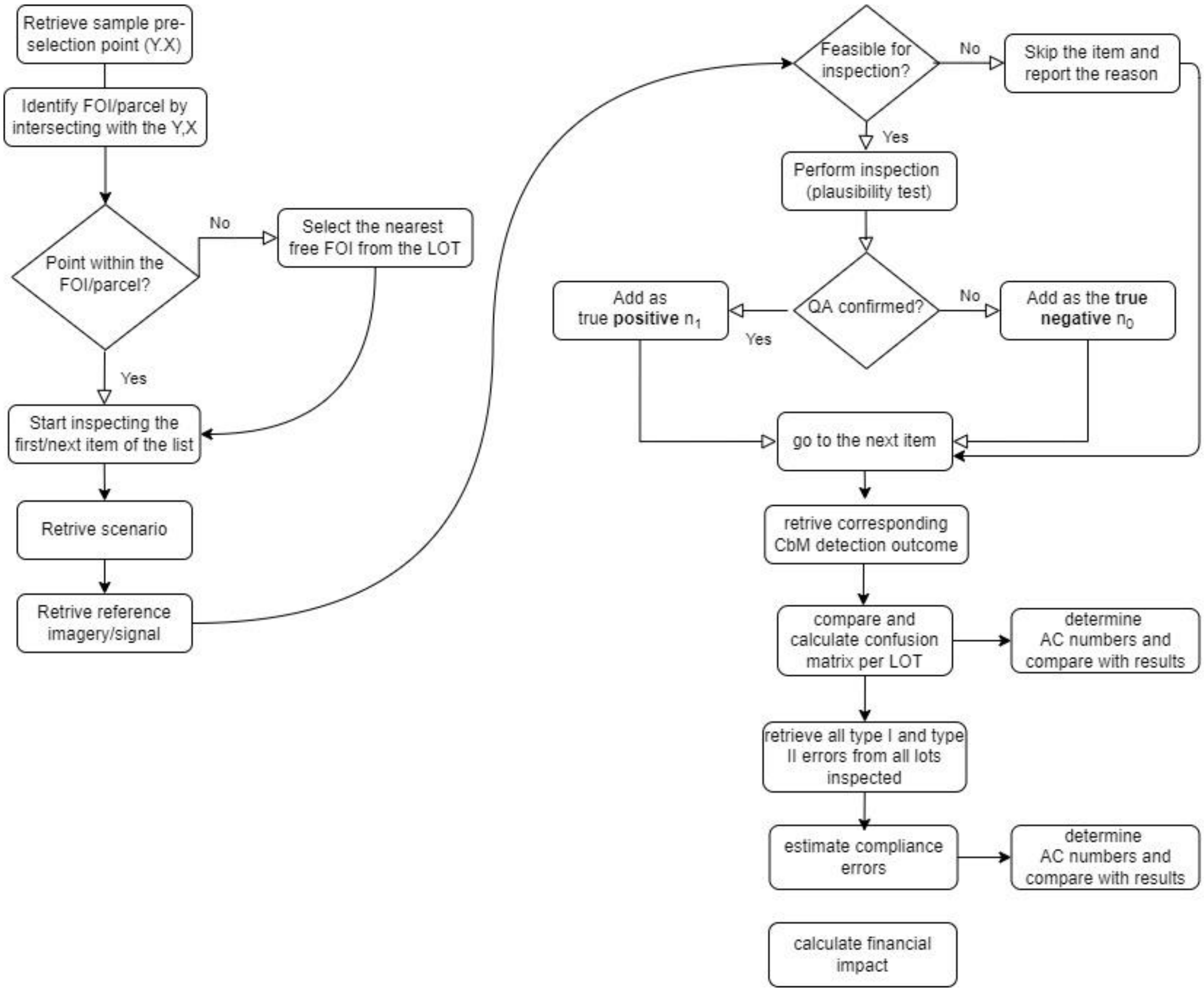
Annex I aims at providing the structure for a more formalized and standardized description and reporting of the quality elements and measures (in terms of definitions, inspections method, evaluation) included in the TG on CbM QA. It provides templates for CbM quality measures and discusses aspects such as feasibility of inspection, verification of boundary conditions/CbM detection outcomes for the different items, components of the acceptance tests and eligibility conditions at the system level.

Annex II elaborates the inspection workflow associated with the CbM QA process. A schematic representation of this workflow is provided in Figure 1.

Annex III provides a first set of instructions on how to identify and categorize all types of observation based decisions (output from switches) occurring in the CbM system and subsequently on how to establish the number of items for inspection per each applicable type. The annex discusses "spatial", "temporal" and "spatio-temporal" types of information extraction and guides to the correct lot definitions (as defined in the TG CbM QA) and population sizes for the QA.

In order to perform CbM QA assessment activities, reference ground information, mainly taken by a human expert, is needed. These references are compared with the outcomes of the automatic process implemented in the CbM system. The protocols to obtain that reference information are, in turn, developed on the basis of ground truth data. Annex IV discusses the main properties of such ground data, the advantages/disadvantages of different data sources and best practices for the definition of the ground truth.

Figure 1. Schematic representation of the inspection workflow as defined in Annex III.



In Annex V, the statistical background and the derived operational tables for both CbM QA step 1 and step 2 are provided. The annex details the procedure adopted to determine the sample size for each lot inspected and provides the derivations of the acceptance numbers for the different tests. The acceptance numbers are derived according to the mindset, rationale and targets of the ISO 2859-2 standard, and are based on the binomial statistics.

Annex VI defines the reporting data package to be adopted by the MSs to describe the outcome of the CbM QA process. The reporting data package has been organized in several XML schemas, which are listed in Table 4. below. The 'x' indicates that the specific reporting item is mandatory for the lot. The different reporting items covers the three stages of the inspection process: (1) lot definition, (2) CbM QA sample pre-selection generation, (3) CbM QA inspection and reporting. These formats allow for lossless data exchanges needed for the assessment and fine tuning of the methodology itself.

Table 4. Reporting package items for the different LOT types. Mark 'x' indicates that the specific reporting item is mandatory for the LOT.

Reporting package item	P1	G1	G2	T1	T2	T3	T4	C1
LpisPointZeroState	-	x	x	x	x	x	x	x
CbmQaSamplePreselection	-	x	x	x	x	x	x	x
CbmItemStatus	-	x	x	x	x	x	x	x
CbmQaReferenceImagery	x	x	x	x	x	x	x	x
CbmQaSamplePreselectionStatus	x	x	x	x	x	x	x	x
CbmEligibilityRules	-	x						
CbmItemLog	-	x						
CbmObservationLog	x	x	x	x	x	x	x	x
CbmQaReport	-	x	x	x	x	x	x	x
CbmQaReportP1	x	-	-	-	-	-	-	-

5 Experiments

The CbM QA pilot phase was successfully completed by four of the MSs taking part and contributing to the 2019 and 2020 exercises. Overall, there has been a good acceptance of the CbM QA process and the feedback from the MSs positively contributed to the CbM QA methodology. The concepts introduced in the TG v.1.0 and developed in the annexes were generally well-received and understood: continuous exchanges between the JRC and the MSs helped to clarify CbM QA concepts and improve the description of the methodology which led to the next TG updates v.1.1 in December 2020 and v.1.2 in June 2021.

A summary of the data considered for the 2019 and 2020 campaigns is provided in Table 5 and Table 6. Most of the MSs decided to adopt the verification of the P1 pre-condition (boundary check) in absence of an explicit G1 test to verify potential cardinality issues between the GSAA declarations and the actual parcel geometries observed by Sentinel-1 and -2. From the preliminary results obtained for the 2019 campaign, it emerged that all the P1 pre-conditions were fulfilled for all the MSs who performed the test, i.e. the number of declarations affected by cardinality issues were below the acceptance number defined by the CbM QA TG. Most MSs participating in the exercise recognized, that those few cardinality errors do substantially propagate into the other tests done in parallel. For example, a cardinality error can lead to biased or misleading signals that can cause wrong decisions in temporal/spatial tests. The implementation of an explicit G1 algorithm can help identify these issues and avoid errors in general CbM detection.

All this simply confirms the key-precondition of the CbM approach, namely that the spatial/area handling (i.e. of FOIs) must be appropriate to obtain meaningful results.

Preliminary results from 2019 showed that most alpha errors (verification of false positives) did not pass the acceptance test. This was mainly due to

1. The very small subpopulation size of negatives (true and false) in the lots, and
2. the stricter acceptance threshold for alpha errors as compared to the beta errors: the limiting quality (LQ) was set to 5% for the former whereas a 10% value was adopted for the latter.

These initial tests in step 1 (automatic detection) were designed to accommodate the (worst case) equal (50/50) proportions of positive and negative items in the lot population. Moreover, tests were designed in order to detect a match rather than an error.

Table 5. Lots considered by the different MSs for the 2019 CbM campaign.

MS	Lot	Lot size	Sample size	MS	Lot	Lot size	Sample size
MS1	P1	411.000	365				
	T4	411.000	365	MS4_2	P1	7.544	350
MS2	P1	500.000	365		T3-1	7.544	350
	T4	500.000	365		T3-2	775	255
	T3	200.000	195		T4	742	255
MS3	P1	368	125	MS4_3	G1	1.592	160
	T4	368	125		T4	1.592	160
MS4_1	P1	301	115	MS4_4	P1	22.187	365
	T3	301	115		T4	22.187	365

From the preliminary results observed during the 2019 campaign, it emerged that most declarations were truthful and monitored via appropriate manifestation scenarios, which resulted in small populations of negative items and total absence of false negatives. In some cases, the number of negative items was practically negligible. In general, the automatic detection systems developed by the MSs were able to properly verify the correctness of declarations with a low number of false negatives (beta errors, missed detections). On the contrary, the low number of negatives available for training the automatic decision algorithms led to high percentages of alphas errors (false positives) as compared to the negative population size. Note that despite the high percentage of alphas errors, the number of alpha errors was, in absolute terms, small. This is due to the small size of the overall population of negative items. Moreover, acceptance numbers are a function of the subpopulation sizes: small subpopulations leading to very conservative acceptance numbers. These conditions explain this, rather high, number of nominal test failures observed for the 2019 tests on alpha errors.

In order to address this issue, two changes were proposed for the 2020 campaign:

1. verification of the subpopulation sizes: if the size of a subpopulation is insignificant, then a failure of a test is waived.
2. increase of the LQ for alpha tests to 10% which is the same value adopted for the beta tests.

The first modification recognizes the low relevance and low statistical significance of a test performed on a population of negligible size. The second change accounts for the fact, that alpha and beta errors can be swapped by simply changing the definition of the tests and complies with the recommendation from the ISO 2859-2 standard to have LQs set at least 3 times higher than the desired quality level (hence the name "limiting" quality). This helps to avoid too stringent acceptance numbers.

Table 6. Lots considered by the different MSs for the 2020 CbM campaign.

<table>
<tr><th>MS</th><th>Lot</th><th>Lot size</th><th>Sample size</th><th>MS</th><th>Lot</th><th>Lot size</th><th>Sample size</th></tr>
<tr><td rowspan="2">MS1</td><td>P1</td><td>476.279</td><td>365</td><td rowspan="3">MS4_2</td><td>G1</td><td>324</td><td>200</td></tr>
<tr><td>T4</td><td>476.279</td><td>365</td><td>T3</td><td>324</td><td>200</td></tr>
<tr><td rowspan="4">MS2</td><td>P1</td><td>506.706</td><td>365</td><td>T4</td><td>324</td><td>200</td></tr>
<tr><td>T2</td><td>33.707</td><td>365</td><td rowspan="4">MS4_3</td><td>G1</td><td>7.756</td><td>365</td></tr>
<tr><td>T4</td><td>506.706</td><td>365</td><td>T3_A</td><td>7.756</td><td>365</td></tr>
<tr><td>T3</td><td>229.546</td><td>365</td><td>T3_B</td><td>2.025</td><td>315</td></tr>
<tr><td rowspan="2">MS3</td><td>P1</td><td>336</td><td>200</td><td>T4</td><td>1.943</td><td>315</td></tr>
<tr><td>T4</td><td>336</td><td>200</td><td rowspan="2">MS4_4</td><td>G1</td><td>25.741</td><td>365</td></tr>
<tr><td rowspan="2">MS5</td><td>T3</td><td>1.347.903</td><td>365</td><td>T4</td><td>38.253</td><td>365</td></tr>
<tr><td>T4</td><td>410.099</td><td>365</td><td rowspan="4">MS4_5</td><td>G1</td><td>578.078</td><td>365</td></tr>
<tr><td rowspan="3">MS4_1</td><td>G1</td><td>21.861</td><td>365</td><td>T3</td><td>578.078</td><td>365</td></tr>
<tr><td>T3</td><td>1.793</td><td>315</td><td>T4C</td><td>544.015</td><td>365</td></tr>
<tr><td>T4</td><td>5.822</td><td>365</td><td>T4A</td><td>575.322</td><td>365</td></tr>
</table>

The preliminary 2019 QA results also revealed that the QA methodology should be expanded to cover the multi-hypothesis testing for those systems, where detection results yield beyond assumed binary outcomes. This is the case where a CbM detection method provides a land cover/land use classification for the GSAA-derived FOIs through machine learning. Hence, QA hypothesis testing should cover three instead of two comparisons between the declared class, detected class and the actual class (found during the QA). This issue affected step 1 inspection and error identification, which was added in the latest TG, v.1.2, in 2021.

While in the first year of the QA trial, in 2020 on the 2019 campaign, QA experts from the Member States were focused on learning how to execute the QA exercise, in the second year, in 2021 for the 2020 campaign, they were more focused on identifying methodological gaps concerning their implementation practices. Apart from steps 1 and 2, also step 3 (financial impact calculation) was introduced later in 2021 in TG v.1.2. Therefore, Member States were able to apply and experiment with the full methodology during 2021. From the second round of the QA results obtained from the 2020 campaign, additional technical details were communicated to the JRC. Below is a summary by MS feedback.

MS1 2020

In the second year of CbM implementation, MS1 found the following elements as non-monitorable in relation to their CbM design option: permanent crops, hemp, Agri-environmental Climate Measure (AECM) not set-aside, Ecological Focus Area (EFA) wetlands or shallow soils with set-aside, mini wetlands, fallow with flowers; set-aside, early mowing, EFA fallow with melliferous plants, hen yard with bare soil, melon, food grape, tomato, cucumber, green house, and vegetable. 5% of these elements have been checked on the spot.

Regarding their QA inspection experiences, MS1 reported several issues:

- The interpretation of grazing activities on permanent grasslands are challenging or even impossible by using the Sentinel data. The only useful evidences were actually complementary ortho images with very high resolution (i.e. 0,5 m) where the animals could have been identified on the inspected parcel.
- Sentinel 2 data gaps (lack of cloud free S2) may coincide with the mowing activities which are relatively short term activities.
- Vegetable growing cycles are differently spread across the season which could be difficult to interpret in annual crops detection marker (T4 lot).
- Some parcels have been identified with partially compliant area. The QA methodology does not take this into consideration in step 2 and step 3.
- 20 px rule for spatial cardinality was found to be practical for bigger parcels. For smaller parcels an area % threshold would be beneficial.

MS2 2020

To overcome the small parcels issue in the detection process, the Administration operated on self-sourced "better than Sentinel" Planetscope data and tried to implement an aggregation of the agricultural parcels of the same land cover type to construct a bigger unit for monitoring (FOI concept). However, they found that aggregation may not work on arable land where the sowing of a crop occurs in distinct phases (partial FOI activity) which is often the case in MS2 using traditional methods; or where the farmer cultivates multiple crops within the same reference parcel without differentiating the agricultural parcel.

MS3 2020

In the second year of CbM implementation and in relation to the CbM design option, MS3 found the following elements as non-monitorable: minimal maintenance of (natural) grassland, fallow land, permanent crops, field margins with AEM and greenhouses. These elements have been checked through the LPIS (no specific details were provided).

Regarding QA inspection, they reported:

- Calculation of the financial impact of an exemption end stage error doesn't work with averages.
- The error on exemption is unrelated to compliance with greening requirements. So MS decided to first look at the actual impact of the error on the holding level.

— The inspections were done by means of ground truth collection (and not the visual image interpretation on Sentinel data). Field visits were performed during the follow-up physical inspections. Note that JRC did perform the corresponding Sentinel based inspection to guarantee uniformity in the reporting.

MS4 2020

MS4 has performed methodological trials through the various regional PAs, who had introduced different CbM methods. For each of them, the CbM QA methodology was found to be useful to test the implemented design option. They expressed their feedback by individual stakeholder:

— **MS4_1** reported a number of issues with the proposed QA methodology.

For step 1:

- MS4_1 considered that the proposed acceptance numbers are unsuitable for unbalanced populations as they are too strict in these cases which lead to many tests failing. As a remedial action, MS4_1 proposed to review the tables and adapt for cases with unbalanced samples.
- MS4_1 brought up the difficulty of properly identifying crop types with similar phenological profiles based on photo interpretation of Sentinel-2 images during the T4 test of the QA. MS4_1 said that, while these errors may not be frequent, this is likely to be a persistent limitation of the system.
- MS4_1 deemed it unreasonable to check cardinality for very small geometries/FOI or with very complicated shapes, as results are totally random.

For step 2:

- In general, MS4_1 noticed some inconsistencies in the guidelines of the proposed QA methodology. For instance, they pointed out that this methodology does not consider that besides Basic Payment Scheme (BPS), other schemes are also subject to CbM.
- MS4_1 stated that eligibility rules should be linked to processing lanes of particular aid schemes. MS4_1 believes that the current eligibility rules, linked to manifestation/absence scenarios (Lot-based sampling) do not always reflect reality accurately. To solve this issue, MS4_1 proposed selecting a single sample of geometries for the whole QA procedure. Because of the above observation and due to the high number of aid schemes and processing lanes of the MS4 monitoring projects, its implementation entailed a high workload. MS4_1 proposed a simplification of the CbM QA methodology better suited for MS4 conditions. This simplification includes a single random sampling for the whole QA procedure. Selected parcels would be considered for all their aid schemes and corresponding processing lanes.
- MS4_1 underlines that, rather than following a sequential (step by step) process, a hierarchical system should be established to waive errors. For instance, cardinality should be at the highest priority, as a cardinality error would remove the parcel from subsequent analysis in other lots.
- MS4_1 found that the allowed error-rates to be very restrictive, proposing a 20 % error rate for step 2.
- Templates for reporting should be revised and simplified.

MS4_1 has been in close contact with the JRC to discuss and solve these standing issues. Some of their recommendations have been already accepted while others are currently (December 2021) being discussed.

— **MS4_2** reported the difficulties of separating certain crops, moreover when atypical growth patterns are found. They pointed out that the main weaknesses of the QA system were associated with the intrinsic characteristics of the Sentinel-2 data (e.g. spatial resolution, scarce cloud free observations during some key periods). For instance:

- spatial cardinality issues related to parcel size and shape, relief and the use at parcel level of the pixel-based crop classification map.
- bare soil detection is hampered by the limited number of cloud free observations available during the required acquisition window. Moreover, MS4_2 found challenging discriminating bare soil and senescent vegetation due to their similar spectral response.

Equally, MS4_2 found that the unbalanced subpopulations represent a serious limitation to reach meaningful conclusions in the QA.

MS4_2 believes that due to differences in the eligibility, it will not be possible to achieve a proper financial impact per scheme for the 2020 campaign. Moreover, they proposed a modification of the original methodology to solve existing inconsistencies (see MS4_1 section above).

— **MS4_3** claimed that the current acceptance numbers are perceived as very restrictive. In addition, MS4_3 sees the use of only Sentinel 2 imagery for QA as a limitation for small parcels.

The overall process has been performed in a collaborative and constructive way. The continuous interaction between the JRC and the MSs has led to a progressive understanding and resolution of reported issues. The MSs have appreciated the support provided by the JRC. Moreover, the JRC has tested and analyzed the data provided by the MSs to better understand the relative weight and potential impact of issues. The results of the analyses have been communicated to the MSs leading to positive comparisons and discussions.

6 Discussion and proposal for further developments

In general, the CbM QA methodology was introduced and developed by the JRC to respond to concern and request of MS. The initiative tentatively aligned with recommendation 1.2 of the ECA's special report SR04/2020([7]) to be completed by 31st December 2021. Since, COM didn't provide so far no legal provisions during the current programming period, those Member States who opted for CbM have no legal obligation to follow By the end of 2021 a new CAP regulation was adopted in the EU Parliament and Council, heralding the shift towards the new delivery model and performance based area monitoring system (AMS) makes legal provision unlikely.

As a result, these developments on CbM QA methodology will no longer translate into a technical guidance (there is no legal requirement to adhere to) but will remain a technical proposal for an inspection, assessment and evaluation methodology.

Furthermore, where CbM operates a control on the beneficiaries of payments, AMS aims at providing system level indicators, irrespective of the beneficiary processes. These are substantially different perspectives. Still, it is very likely that those Member States which already developed CbM will have a huge technological advantage in setting up the new AMS. It is expected that in this transitional period they accumulated knowledge on Sentinel data processing, automatic detection and monitoring of agricultural activities.

As regards CbM QA, its future lies in voluntary self-assessments by Member States or at least until CbM doesn't phase-out. Nevertheless, some methodological components developed in this process could and should be recovered for the future AMS QA. All Member States involved in the CbM QA developments have expressed that they gained an enormous amount of new insights not only on the inspection and testing methodology which supports the self-assessment exercise. Thus, the proposed CbM QA achieved the purpose to assess, to fine-tune and to further develop PA's system for automatic detection. It also triggered more correct parametrization of the eligibility criteria and rules defined on the national level.

In relation to the JRC's CbM QA 3-step methodology itself, four major topics of discussion can be identified: 1) suitability of Sentinel data as all-purpose inspection data, 2) LPIS based sampling or lane (support scheme/practice) based sampling; 3) debate on compliance (parcel) based or switch based inspection and 4) importance of the financial impact assessment. Each of the topics are elaborated in the subchapters below.

6.1 Suitability of Sentinel data as all-purpose inspection data

As the inspection process needs evidence-based data to be confronted with the declared data or the one generated from the automatic detection process, there were two different approaches expressed by the Ss on which data can be considered as the most suitable for independent observations. The QA methodology proposes the Sentinel data to be used as the main source for visual image interpretation (with an independent observation method), some MSs argued that Sentinel data have poor image resolution quality for a specific crop identification. Such views were mostly driven by some specific detection implementations, mostly black box type of detection systems, in which the detection outcome generates a prediction on specific crop classes, rather than detection of agricultural activities. For them, the most suitable method for validation of the detection outcomes were sporadic and well targeted field visits.

The other, completely opposite opinion from the MSs, was that detection outcome testing should be done on the very same input data with the same image resolution since otherwise the inspection observations obtained would not be relevant to remedy the detection processing (or algorithms). The challenge here is obviously linked to the detailed information that is normally visible as the actual (ground) situation in comparison with the limitations of depicting the same level of details from high image resolution data, and the concern, that such found differences would not contribute to the detection mechanism per se.

Both approaches were identified by the JRC as confusions in terms of the observation method. Namely, the QA methodology clearly separates the evidence data against the observation method. As for evidence data it was recommended to use freely available Sentinel data as the main source of the QA inspection, and where necessary to be complemented with auxiliary data sources. As for the observation method, it was recommended to perform a visual image interpretation as it would be a different (independent) mean to come to a conclusion

([7]) "Using new imaging technologies to monitor the Common Agricultural Policy" on the testing hypothesis. That would mean that the computer based results would be replaced with a human based interpretation involving a logical cognition and interpretation of the "truth" and compared against machine-based results. As the automated (much faster) process usually needs to replace a human and much slower process, it makes sense to use a human interpretation for validating the automatic detection results on the same and/or complemented evidence datasets.

For practical implementations, JRC prefers precise (low variability) methods over accurate (low bias) methods, and the use of common source data (the Sentinels) for both CbM operation and CbM QA inspection is believed to enhance precision. Different sources facilitate the identification of bias issues.

6.2 LPIS based sampling versus lane (support scheme) based sampling

Three options are available for sampling the CbM populations entering into the proposed methodology. The most controversial option was to use the LPIS population as the basis for sampling. The idea behind this option was to minimize the compilation workload of the Administration since the LPIS population was assumed to represent all declared reference parcels in the assessment year and in parallel had already been prepared for the legally required LPISQA exercise. This option implied no additional data exchange burden for the MS.

However, inspection items (parcels/FOIs) that were pre-selected from the LPIS population would have to be further analyzed and matched with the items (decisions) that are subject to specific types of information extraction as grouped and identified within a homogeneous lot. It seems that the combined practice was more of a complication than a simplification when it comes to the overall data preparation. Therefore, JRC acknowledged the need to have one to one relation between the lot populations and sample pre-selections determined for each lot. This would mean that the Member States would prepare the lot population file upfront, from where the JRC would randomly select a dedicated QA lot sample. In self-assessment setups, MS would sample themselves.

Another challenging aspect was revealed in the second half of 2021, where MSs included more support schemes in their phasing-in operations, and as a consequence, had to break down the QA results per each aid scheme (per lane). A proposal was given to consider again more specific and dedicated populations grouping per lane instead of per type of information extraction. Given the estimated amount and complexity for the support schemes, this last proposal was so far not considered as optimal for the time being. Further work is needed.

6.3 Compliance (parcel) based versus switch based inspection

In order to facilitate detailed diagnostic analysis of potential errors within the CbM system, especially within the automatic detection and data processing, the QA methodology defines the switch decision item as the basic item for inspection. The switch decision item is a generated outcome from a unique processing logic, when completed, would yield a positive or negative decision as an answer to its programmed block or algorithm. In practice, it will be a single fragmented outcome from a time series processing in a form of one or more marker detection result and similar analysis. The concept of grouping all generated decisions into a homogeneous lot facilitates a faster inspection routine and at the same time, when erroneous, instantly provides feedback on which algorithm needs to be corrected. This is quite a novelty for QA inspections in the IACS community. Previous OTSC and CwRS exercises were directed more towards inspection of the compliance findings on the level of the holdings and its parcels, hence, similar approach was expected in the community to be continued also in the CbM QA.

For that reason, it was proposed by a MS that the item for inspection would be the declared parcel and the inspection would simply compare the final CbM compliance result (traffic light) with the QA observations which would compile all necessary field elements as well as eligibility elements in a single observation approach. This oversimplified approach would, however, miss the assessment of the specific algorithms which lead to the final CbM compliance output on the level of the parcel. In such inspection identification of effective measures for system correction would not be sufficient.

In this context, it's also worth mentioning, that the CbM conceptual framework explicitly introduces marker performance in terms of false negatives/positives, to be assessed by independent ground truth and relevant for documenting the design options and detecting bias. By contrast, the CbM QA switches approach deal with operational decisions of the overall process (compliance and beyond) and favor precision.

6.4 Importance of the financial impact assessment

The introduction of the 3rd step was also intensively discussed, not only with the MSs, but also within the DG AGRI. Some MSs expressed their concerns that the calculated error values would be used to sanction the work done on the CbM. However, the Assurance and Audit Unit required such calculations for the Annual Activity Report. This Annual Activity Reports informs Parliament and Council on the financial performance of the policy.

That discussion brought some valuable aspects for further consideration. One of them was to consider that a particular FOI subject to inspection may be partially compliant and partially incompliant. Hence, only that partial incompliant area share should be measured to quantify the financial impact on errors found. However, area measurements were not initially foreseen in neither CbM or CbM QA In cases were area measurement is needed for correct calculation of the financial impact, MSs may use all measurement protocols valid for IACS purposes.

The second aspect was also, raised by DG AGRI mostly, that step 3 is to extend the QA scope to other elements of the controls which were not covered by the automatic processing within the CbM. Those elements were identified as either inconclusive CbM results, non-monitorable, or not-monitored elements, all of which may or may not be currently present in a given system. It is expected that in time and with the technological improvements, all of those currently not monitored elements will be included in the fully automatic detection system. They are the subject of the traditional "grey zone" discussions that traditionally take several years of exchanges and progressive clarifications prior to reaching an agreement.

References

- DS-CDP-2018-18 technical discussion: 2nd discussion document on the introduction of monitoring to substitute OTSC: rules for processing applications in 2018-2019

  https://publications.jrc.ec.europa.eu/repository/handle/JRC112913

- Devos Wim et all, "Conceptual basis of checks by monitoring: Guide to the Checks by monitoring tool set", 2021

  https://marswiki.jrc.ec.europa.eu/wikicap/images/8/87/JRC127678_final.pdf

- Annex IX - Technical guidance on LPIS population for LPIS QA inspection

  https://marswiki.jrc.ec.europa.eu/wikicap/index.php/LPIS_TG_Population

- ISO 2859/2: Sampling procedures for inspection by attributes – Part 2

- Devos Wim et all., "Checks by Monitoring quality inspection: EU requirements and methodology", JRC Technical Report, Aug 2020

  https://marswiki.jrc.ec.europa.eu/wikicap/images/0/0a/TG_CbMQA_1_1.pdf

- European Court of Auditors, Special Report 04/2020: "Using new imaging technologies to monitor the Common Agricultural Policy"

  https://www.eca.europa.eu/Lists/ECADocuments/SR20_04/SR_New_technologies_in_agri-monitoring_EN.pdf

List of abbreviations and definitions

Abatable error	Abatable errors are those detection errors that were affected by eligibility rules of a given aid scheme and may result in a claim reduction. These errors are expected to be corrected by the beneficiary as they are disadvantageous for her/him provided the beneficiary was timely informed.
Absence scenario	Scenario where the absence of a behaviour and/or its markers is required to meet eligibility conditions
AC	Acceptance number;
AECM	Agri-environmental Climate Measure;
AL	Arable Land;
BPS	Basic payment scheme;
CAP	Common Agricultural Policy;
CAPI	Computer assisted photo-interpretation;
CbM	Checks by monitoring; a process substituting the OTCS; a procedure of regular and systematic observation, tracking and assessment of all eligibility criteria, commitments and other obligations which can be monitored by Copernicus Sentinels satellite data or other data with at least equivalent value (as defined in Art 40a of Regulation EU No 809/2014);
Compliance	Accordance with the eligibility rules within an aid application or a system;
Confusion matrix	Tool to determine the performance of a classifier. It contains information about actual and predicted classifications;
EFA	Ecological focus areas;
End-stage error	End-stage errors are those detection errors that were affected by eligibility rules of a given scheme and don't results in a claim reduction. Those errors represent an undue payment in direct financial loss for the payment Fund.
FOI	Spatial "footprint" of the observed land phenomenon; i.e. the space occupied by the (bio)physical object on the earth. Its spatial representation in the CbM system is derived/constructed from GSAA/LPIS. On individual Sentinels images it is captured, as a continuous patch of pixels associated with (bio)physical object.
Ground truth	Data and information on physical reality obtained by direct observation or measurement in the field; used to derive the rules and parameters for extraction of the relevant information from remote sensing data.
GSAA	Geospatial aid application;
GTCAP	Geodata and Technologies for the CAP (a JRC work package)
Item	Individual processing decision based on Sentinel observation of field conditions, that impacts an explicit conclusion within the CbM process flow or within its sub-process;

	unit of the assessment/inspection;
LOT	Quantity produced together and sharing the same production costs and specifications;
LPIS	Land parcel identification system;
Manifestation scenario	Scenario where the manifestation/presence of a behaviour and/or its markers is required to meet eligibility conditions
Marker	Record on “what” has been observed on the signal. It holds both the nature of the observation and the time it was manifested on the FOI. The nature of the observation either reflects a behaviour or a (provisional) status quo in the bio-physical manifestation of the FOI where the signal was calculated.
Monitorable	Reference to either an eligibility condition, a parcel or other element that is a subject of automated process in the CbM and as such can produce observable verdicts on Sentinel based processing;
MS	Member State;
N	Population size;
n	Sample size;
N_0	Sub population of negatives;
N_1	Sub population of positives;
OTSC	On the spot check; control process of the farmers’ applications on 5% sample in a given year;
PA	Paying Agency;
QA	Quality Assurance;
TG	Technical Guidance;
Type I error	α, false positive of the automated process in the system; on a sub process level a type I error occurs when a decision (or a switch) identifies falsely presence of phenomenon ;
Type II error	β, false negative of the automated process in the system; on a sub process level a type II error occurs when a decision (or a switch) identifies a falsely absence of phenomenon
VCS	Voluntary coupled support.

List of figures

List of tables

KJ-NA-31027-EN-N

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

EU Science, Research and Innovation

EU Science Hub

Publications Office
of the European Union

doi:10.2760/348125

ISBN 978-92-76-49632-8