GSE Land

Date Issued: 28.04.2008 **Revision/Issue: I.00**

C5 - Service Validation Protocol

ITD0-421-RP-0003-C5

Issue 1.00

GSE Land Information Services

ESRIN/Contract No. 19407/05/I-LG

Stage 2 of the Earthwatch GMES Service Element
Scaling Up Consolidated GMES Services

Book Captain: Stefan Kleeschulte (UAB)
Contributing authors: Charlotte Steinmeier (WSL)
Michael Koehl (sfm)
Susanne Meirich (ibm)
George Büttner (FÖMI)
Oscar Gomez (UAB)

Author:	Stefan Kleeschulte (UAB)	Sign	Date
Task manager:	Stefan Kleeschulte (UAB)	Sign	Date
Review:	J. Janoth (ITD)	Sign	Date
Approval:	S. Kuntz (ITD)	Sign	Date
Approval:	F.M. Seifert (ESA-ESRIN)	Sign	Date

Distribution:	All GSE Land partners	ESA		

Change Record

Issue/Rev	*Date*	*Page(s)*	*Description of Change*	*Release*
-	13.01.2006	All	Release of first Draft Document	D1.00
D1.00	03.03.06	All	Addition of statistical sampling concept, validation approach	D1.01
D1.01	28.03.2006	All	Completion of validation concept	D1.02
D1.02	31.03.2006	All	Comments ITD and update on validation concept	D1.03
D1.03	13.09.2006	Chapter 3.4	Review of description of validation concept and addition of details	D1.04
D1.04	11.10.2006	All	Review and comments by ITD	D1.05
D1.05	13.10.2006	Chapter 3	Update on point sampling approach and urban atlas validation	D1.06
D1.06	26.07.2007	All	Update of chapters: 3.3.1.1, 3.4.3.2, 3.4.3.3; update of Table 6	D2.00
D2.00	28.04.2008	-	Release of Issue	I1.00

TABLE OF CONTENTS

1.SCOPE

The goal of this document is to define the quality assurance concept for the whole GSE Land project. GSE Land will produce a number of different products derived from remote sensing data; these products can be classified in two different groups: mapping products, and downstream services. Mapping products are created from remote sensing images, and downstream services are derived from mapping products.

In the context of GSE Land project, only mapping products are considered in QA. The validation of the downstream products will be done by the final users of the service based on specific criteria and information that cannot be generalised.

The main goal of quality assurance concept is to guarantee quality of the product. The quality assurance concept is not only approved by the service provider network, but also by the user working group.

It is important to produce reproducible and traceable results. In this sense, quality assurance not only guarantees quality, but also helps to reduce cost in production.

Figure 1: QA concept

Qualification & Quality Assurance concept

Main focus:
Service Qualification & Quality Assurance concept approved by:

- Service Provider Network
- User Working Group

Goal:

- Basis for guaranteed quality
- Reproducible and traceable results
- Builds up trust
- Basis for cost reduction by industrial production

Key Elements	Mapping Products	Downstream Products
• Qualified interpreters	X	
• Independent Process / Production Qualification	X	
• User QA	X	
• User Acceptance	(X)	X

1.1 GSE LAND MAPPING PRODUCTS

Inside GSE Land mapping products, there are two different approaches to obtain them:

- Visual interpretation products
- Automatic or semi-automatic products

Visual interpretation products are the products that include a visual interpretation process as one of the steps inside their production chain. In these products, the most important point in which quality must be assured is visual interpretation.

Automatic or semi-automatic products are the products that do not include a visual interpretation process. They have an automatic process to classify remote sensing images into the mapping product class. The semi-automatic procedures include a manual threshold definition.

GSE Land mapping products are the following:

a) Visual interpretation products:

- M1.1 Urban Atlas Map (very high resolution mapping)
- M2.1 Regional Land Cover
- M2.4 Land Take Map (update / downdate)
- M2.6 Land Take Map (first inventory)

b) Automatic or semi-automatic products:

- M2.3 Forest Parameters
- M2.5 Agricultural Land Use
- M3.1 Arable Acreages Map (medium resolution seasonal mapping)
- M3.2 Phenology (seasonal biophysical parameters)

These two different groups of products will have different approaches for quality assurance. In the case of visual interpretation products, the results of the semi-automatic pre-processing and the subsequent visual interpretation must be checked. Ensuring visual interpretation quality means, that products are created a comparable way independent of its production site or production time or target area. So in this case it is crucial to check that a common nomenclature is agreed, and that interpreters are trained equally and that they are networked enough to ensure they all map the same way.

In the case of automatic and semi-automatic products, it is easier to ensure quality, as long as the process is automatic. The difficulty here is more linked to the quality of the automatic process as it acts as some kind of a black box. The usability of the end product should be checked by the end user, the QA team can only control the correct use of the proposed method, but not its result.

2.RELATED DOCUMENTS

2.1 INPUT

Overview of former deliverables acting as inputs to this document.

Document ID	Descriptor
	C7 – signed SLA
	S5 – Production chains for selected products
	C6 – reports from production

2.2 OUTPUT

Overview of other deliverables for which this document is an input.

Document ID	Descriptor
	C6 – Service Validation Report
	Certificate for the audited processing chains

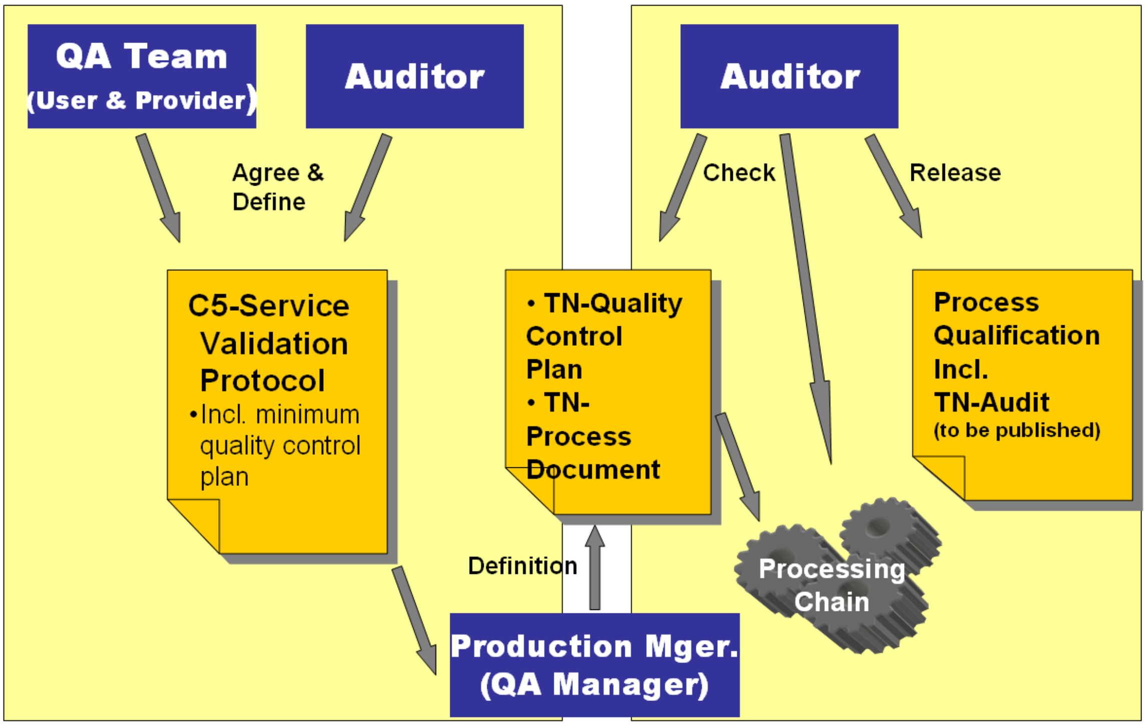
3.GSE LAND QUALITY ASSURANCE / QUALITY CONTROL

3.1 THE VALIDATION PROCESS

The actors involved in the Quality Assurance process in GSE Land are:

- The service provider, who provides the products to the client.
- The QA team, who is responsible for assuring quality, as an external organization.
- The Auditor, who is responsible for checking that the service providers are producing accordingly to the quality assurance guidelines, and to their own standards.
- The Client organization, who will review the product according to the requisites he defined (U7 - Service Utility Report).

Figure 2: Validation process



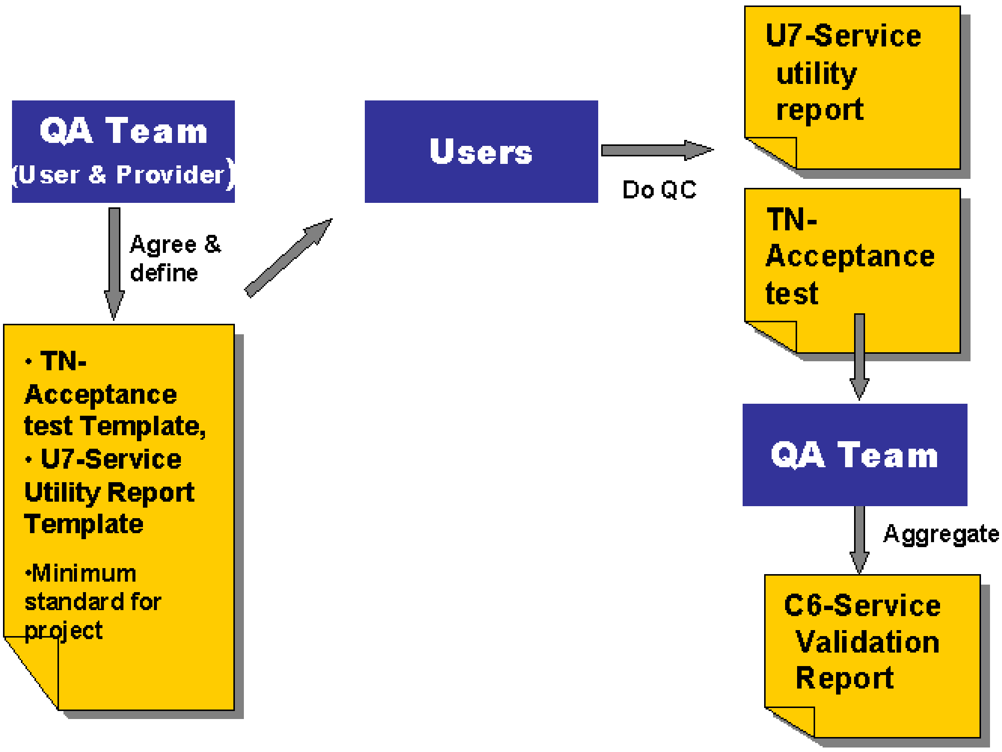
As described in the figure above, there is a first step, in which the QA team agrees and defines the Service Validation Protocol (this document). After that, and before starting the production, the service providers must define their internal quality assurance/ quality control plan accordingly, integrated into their processing chain.

A second step will comprise the following: the auditor will check that the quality control plan exists, integrated in each processing chain, and including all the requisites defined in the Service Validation Protocol. Once the production has finished (or there is one block finished), the auditor can also check if the production/process documents agree with service providers' defined processing chain and quality control plan.

Another main step is the validation of the product. It includes three main components:

- A qualitative verification by the QA technical team in which some intermediate results of the service provider interpretation will be commented by the QA team and potential deviations from the specifications will be highlighted. The verification will be done during the course of production and is meant to increase data quality.
- A quantitative validation: The QA team will perform an independent external technical quality control after the finalisation of the mapping product to check whether the products reached the desired quality.
- Requisites validation: the user will validate the product if it fulfils the needs that the user itself has included in the product requisites. The figure below illustrates the validation process.

Figure 3: Collaboration QA team / user and related documents



3.2 TECHNICAL CONCEPT

As described above, the quality assurance concept will be different for the visual interpretation and the automatic and semi-automatic procedures. Common to both approaches are the following points:

- There will be two different quality control processes: an internal one, performed by the service provider and external one, performed by the QA team.
- The external QA team is led by the European Topic Centre on Terrestrial Environment (ETCTE from here onwards), and the Quality Control will be performed by the CORINE Land Cover technical team.

- At the beginning of the production process each service provider is requested to define the principles of the production chain he plans to follow and the accompanying quality control breakpoints. The adherence to these production chain guidelines and the correct documentation of the results of the internal quality control will be evaluated by an independent body. This independent body is represented by the German TÜV (an independent service organisation for security, quality and environmental protection). Due to the lack of an existing international standard, the TÜV will not be able to "certify" the production chain of the service providers, but only to verify if given rules are adhered to or not.
- After finalisation of a significant part of the production process (> 30%) the QA team will perform a qualitative verification of the mapped products. In this verification comments about main "errors" are provided to the service providers in order to improve and harmonise their production. This quality control is based on a sample of the actual mapping product and a reinterpretation of a certain area by an external expert.
- The final quantitative validation of the mapping products based on a statistical sampling design and the provision of defined accuracy measures (see chapter 3.4.3 for full details).

In order to start the production process for each mapping product, guidelines have to be created (legend description, interpretation guidelines [if apply]) which should ensure a harmonised production process across the different service providers and the delivery of comparable products. Here, a similar approach as in Corine Land Cover is followed.

3.3 INTERNAL QUALITY ASSURANCE / QUALITY CONTROL

Internal QA/QC is even more important than external QA/QC. Although the quality will be checked through the external QC, the quality of the product will be ensured internally (it should be produced according to the quality requirements) by each service provider. Each production chain for each product must meet the general principles described in this document: interpreters training and interpretation guidelines (if apply), main steps of the production and quality control breakpoints.

As a very first quality assurance guideline, and because some of the products of GSE Land are produced by different service providers covering different areas, the following issues must be considered for both visual interpretation and automatic or semi-automatic products:

- There should be two copies of each product: one using the national projection, and another one using a standard pan-European coordinate reference system (see http://crs.bkg.bund.de/crs-eu/);
- Field names and types should be compliant with upcoming INSPIRE specifications (see http://www.ec-gis.org/inspire/), and with the naming conventions included in the interpretation guidelines of the respective products. Agreed naming convention for the land cover attribute ("item") is "GSELxx_2005";
- Field names and types must always match for the same product when the datasets are splitted by region.

In addition, the processing chain for any product, visual interpretation or automatic or semi-automatic, should include at least the following minimum quality control breakpoints:

- Satellite data check for correctness and completeness
- Ancilliary data to be used (raster or vector) quality check for correctness (geometric accuracy, contents, plausibility, …) and completeness.
- A final product will be delivered to the production manager; that product will be checked by the production manager before it is delivered to the user.

The following table shows these minimum Quality Control breakpoints in detail:

Name	Description/ Function	Quality Control criteria	Method	Data used for QC	Action if no-go
Import data check	EO-Data will be checked for correctness, completeness	Data will be checked for: correctness (readability, location, data type, channels) completeness (information content, e.g cloud coverage, projection information, acqu. Date, ilumination) Valid for base-EO- and additional raster data	Parameter reporting	Product specification, purchase order forms	Repeat import; Contact data provider
Ancilliary data check	Ancilliary data will be checked for correctness, completeness	Data will be checked for: correctness (readability, location, geometric accuracy, data type, channels, …) completeness (information content, e.g full coverage of desired area) Valid for main (e.g. project area) and additional (e.g. grid) vector data	Parameter reporting	Product specification, purchase order forms	Repeat import; Contact data provider; Update of vector data layer if necessary, selection of suitable classes for implementation
Final product check	A final product will be created for: Allow the production manager to check it Send it to the final user after that check	The final product will be checked for plausibility, homogeneity	Plausibility check, completeness , correctness	Ancilliary data	Repeat the creation of the final product

3.3.1 Visual interpretation products

In order to ensure quality of the visual interpretation products, the way of how the interpretation is made must be homogeneus among service providers. In order to ensure this, the following must be considered:

- A document containing **interpretation guidelines** must exist, including both the agreed **nomenclature** and the **mapping/interpreting guide**. This document must be unique for each product, although more than one service provider delivers the same product. The chief interpreter is responsible for creating and updating this document.

- Each interpreter must be trained to ensure interpretation is done homogeneously.
- The production chain must contain the main steps described in this document.
- The quality control breakpoints described in this document must exist and the defined quality thresholds must be met.

3.3.1.1 Interpretation guidelines

For all mapping products interpretation guidelines and nomenclature descriptions have been created by the responsible task managers in cooperation with the partners. The quality assurance team has reviewed all these guidelines for completeness and consistency.

The different applicable documents are:

- M1.1 Urban Atlas Map (very high resolution mapping)

  The Urban Atlas nomenclature has been discussed and agreed with DG Regional Policy. The agreed version represents a European minimum standard which can be derived from EO data and limited ancillary information (TeleAtlas street network, topographic maps).

  The thematic details of products for local clients (cities) may go beyond this European standards (i.e. additional classes), but the compatibility with the European minimum standard needs to be guaranteed.

  Interpretation Guidelines:

  ITD_0421_GSELand_Mapping_guide_M11_E_3.2_SKL.doc (European minimum standard)

  The mapping guidelines for the local version will be defined as appendix to the European version.

  Nomenclature:

  UAnomenclature_Final_20070530.xls (European minimum standard)

- M2.1 Regional Land Cover and M2.6 Land take

  Interpretation Guidelines:

  ITD_0421_GSELand_Mapping_guide_M26_M21_E_V1.8.doc

  Nomenclature:

  GSELand_Digitisation_product_M2.1_E_V1.4.xls

  GSELand_Digitisation_product_M2.6_E_V1.5.xls

- M2.4 Land Take Map (update)

  Interpretation Guidelines:

  ITD_0421_GSELand_Technical_guide_M24_E_V1.1.doc

  ITD_0421_GSELand_Mapping_guide_M26_M21_E_V1.8.doc

  Nomenclature:

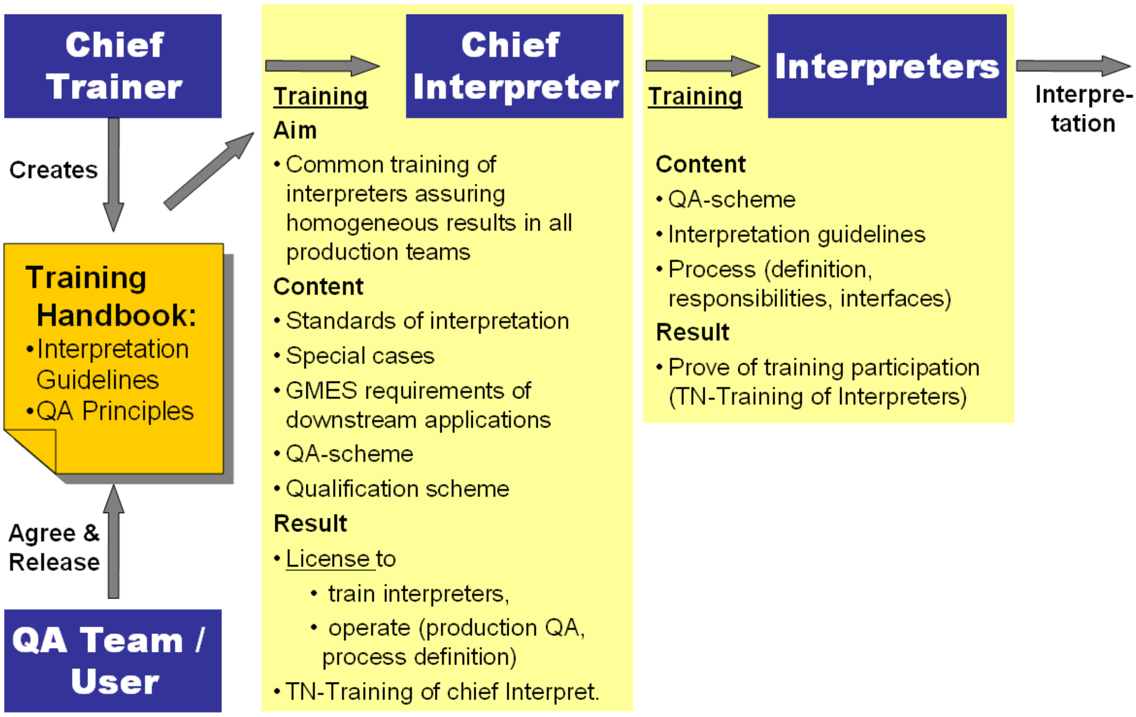
  GSELand_Digitisation_product_M2.6_E_V1.5.xls

3.3.1.2 Chief Interpreters and Interpreters training

In order to have a homogeneous approach and understanding for interpretation, interpreters should be trained. The suggested strategy is the following:

- The chief trainer, Susanne Meirich, a member of the QA team, prepares the interpretation guidelines and nomenclature, assisted by the QA team and service providers if necessary.
- The chief trainer trains the chief interpreters (at least one for each product and service provider)
- Each chief interpreter trains each interpreter that will work in interpretation

Figure 4: Training



This is enough to start the interpretation and to have a common basis, but might not be enough when a doubt comes up or something has not been foreseen in the interpretation guidelines or the nomenclature (special cases). The particular decisions taken each time should be available for all the interpreters, to take the same decision everywhere, every time. The QA team should be able to access these decisions in the QC process, or to give their opinion. The following actions are recommended for this:

- the establishment of an internet-based platform, as a discussion forum called 'interpretation', with access to the QA team and the interpreters. This can be used as an input to update interpretation guidelines
- Regular or on demand teleconferences between chief interpreters and QA team
- Regular training updates (i.e. annually)
- Annual review of the production process with feedback from service providers and QA team on possible improvements of the process.

3.3.1.3 Production chain

As part of the QA procedure, all service providers delivered information about their production chain. For each specific mapping product the production chains are rather similar among the different service providers involved in the implementation of a specific mapping product. By consequence, a standard processing chain can be defined for visual interpretation products. This standard processing chain must be followed by all service providers with respect to production steps and related quality control breakpoints.

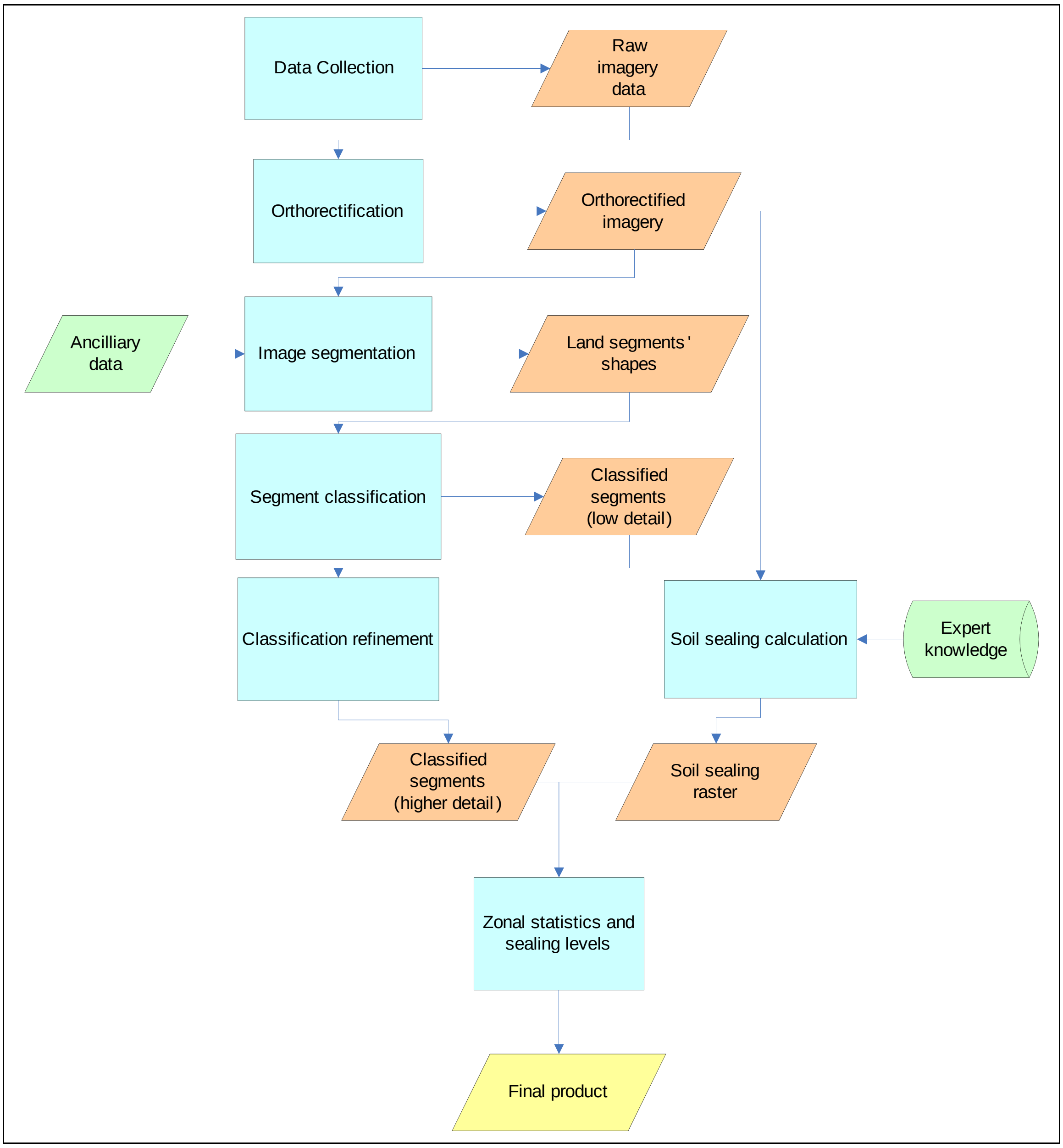
Service providers who do not follow the standard production chain are requested to document where they deviate from the standard and what are the impacts on the quality control breakpoints. In these cases, the QA team should evaluate whether this change is acceptable or not in terms of quality assurance.

If not documented differently, the TÜV will assume the use of the standard production chain when evaluating the individual service providers.

The standard processing chain for visual interpretation products comprises the following steps:

- Step 1: Data collection
- Step 2: Orthorectification
- Step 3: Image segmentation
- Step 4: Segment classification
- Step 5: Classification refinement
- Step 6: Soil sealing calculation (for M2.6 / M2.4 only)
- Step 7: Zonal statistics and sealing levels (for M2.6 / M2.4 only)

Figure 5: Flowchart of the standard processing chain



3.3.1.4 Internal quality control milestones

All the GSE Land mapping products that are created using visual interpretation should include at least the quality control breakpoints described in the table below.

For each quality control breakpoint in which some reporting must be done (i.e., parameters reporting), a template for each of those should be created and agreed among service providers and the QA team.

Table 1: Quality breakpoints

Name	Description/ Function	Quality Control criteria	Method	Data used for QC	Action if no-go
Import data check	EO-Data will be checked for correctness, completeness	Data will be checked for: correctness (readability, location, data type, channels) completeness (information content, e.g cloud coverage, projection information, acqu. Date, ilumination) Valid for base-EO- and additional raster data	Parameter reporting	Product specification, purchase order forms	Repeat import; Contact data provider
Import metadata-check *(additional data)*	Metadata will be checked	Data will be checked for: correctness (readability, location, data type, channels completeness (information content, e.g full coverage of desired area) Valid for main (e.g. project area) and additional (e.g. grid) vector data	Parameter reporting	Product specification, purchase order forms	Repeat import; Contact data provider; Update of vector data layer if necessary, selection of suitable classes for implementation
Orthorectification accuracy	Orthorectified images will be checked	Orthorectification process: RMSE <= 1 pixel measured with the GCPs In x, y and xy Validation process: RMSE <= 1,5 pixels measured with check points. RMSE maximum in one point <= 2 pixels Correct projection	Statistic evaluation / Parameter reporting	Spatial references, e.g. TK, GPS points	Repeat orthorectification
Data fusion check (optional)	Correct execution of data fusion	Data will be checked for: correctness (readability, geometic accuracy of datasets fused) completeness (information content) Also check of RMSE after fusion of data to verify that geometry has not changed after fusion.	Parameter reporting / plausibility check (visual)	Original image data before fusion, original histogram distribution of input images	Repeat fusion process Repeat orthorectification

Name	Description/ Function	Quality Control criteria	Method	Data used for QC	Action if no-go
EO subset check	Correct execution of EO subsetting	Data will be checked for: - correct coordinates of subset	Parameter reporting	Original image data before subsetting, Spatial references, e.g. TK,	Repeat subset process
Check data availability for Visual Interpretatio n	Correct execution of data reading: satellite data, preclassification result, topographic maps, vector datasets, any other data	Data will be checked for: completeness of data delivery correct position of images, correct set alias for processing chain, coordinates of subset , projection, …	Plausibility check	Original data	Repeat until succeed
Segmentati on/ PreClassific ation check	Correct execution of processing steps, interaction by interpreter (set thresholds)	Interim results are checked: - correct delineation of polygons	Parameter reporting, visual check, plausibility check	Product specification, interpretation guideline, additional spatial references	Repeat
Classificatio n check	Correct execution of assignments	classification content, - class coding	Visual check	intermediate classification goals (e.g. urban mask, forest mask, …)	Repeat
Data fusion check (to be done only if data fusion has been done)	Implementation of external data	Fusion product will be checked for: topology delineation (quality after GIS processing)	Visual/ plausibility	Input ancillary data files	Repeat until succeed, GIS post-processing
Data Post-classificatio n check: Lcover -> Luse	Correct post-processing of land cover classes using AUXdata to check if the LC to LU conversion was correct / makes sense	Clasification result will be checked for: classification, - class coding/ topology MMA - comparison to land cover class	Visual/ statistics/ plausibility	Aux data, Plausibility check	Repeat until succeed/ in-plausible objects are not transferred

Name	Description/ Function	Quality Control criteria	Method	Data used for QC	Action if no-go
Final Accuracy Assessment	Check of thematic and geometric accuracy	Thematic accuracy Overall acc. M1.1 : >= 90% M2.4, 2.6 : >= 95% M2.1 >= 80% Class nomenclature and geometric accuracy (delineation), file format		Independent reference data derived from e.g. aerial photographs, ground truthing, Topogr. maps	Identify and repeat responsible production steps

3.3.2 Automatic or semi-automatic products

Automatic or semi-automatic products do not have visual interpretation. This means that the two only issues to assure quality are the production chain and the quality control milestones.

3.3.2.1 Production chain

In the case of automatic and semi-automatic procedures, only for product M2.5 more than one service provider is producing the same product. In the case of M2.5, processing chains will not be compared because the products obtained in both production sites will be completely different in terms of the classes that will be obtained. Consequently, it is not necessary to define a standard for each product, as long as there is only one production chain for each product. Nevertheless quality breakpoints need to be defined and will be controlled by the external validation.

One of the goals of quality assurance is to produce reproducible results. In this sense, the steps where human intervention is required, such as setting threshold manually, is the weakest point. These steps should be very well documented during the production to be able to reproduce the same results from the same input information. Quality assurance guidelines must tackle these steps particularly, providing a methodology to ensure that the result can be reproducible. The definition and use of customized templates for each of these steps is recommended.

3.3.2.2 Internal Quality Control Milestones

As described above, internal quality control milestones will vary from one product to another. The tables below show the internal quality control breakpoints for each product. **During the first phase of GSE Land, the function or description of the breakpoint should be present, but the actual naming (B-01, B-02, …) might be different**:

Table 2: M2.3 Forest Parameters

ID top-level	ID	Name	Description/Function	Check criteria	Interaction	Hardware Platform	Operating System	Software	SW type
1	B-01	Check input EO data	Control of original EO data derived from data service provider (no bands, area etc.)	Data OK	semi-automated				COTS/prop.
1	B-02	Quality geocorrection	Control of geocorrected EO data. Based on GCP, RMS error	Quality OK	semi-automated				
2	B-03	Check input EO data, difference input	Check that all bands needed are included, geometry ok and that area to be mapped are included in both images.	Quality OK	manual	PC	Windows XP	Erdas Imagine	Image processing SW
2	B-04	Check cloud mask	Check that cloud mask covers all cloudy (incl. Haze) areas and that verified not cloudy area aren't declared as cloudy.	Data OK	manual	PC	Windows XP	Erdas Imagine	Image processing SW
2	B-05	Check difference image	Check that the output of the difference image is ok.	Data OK	manual	PC	Windows XP	Erdas Imagine or Enforma	Image processing SW
3	B-06	Check input data before processing	Check that all bands needed are included, geometry ok and that area to be mapped are included in both inputs and that pixel size is identical	Data OK	manual	PC	Windows XP	ArcGIS / ArcView	GIS SW
3	B-07	Check prototypes	Check of prototypes on subsample of reference data	Data OK	Manual	PC	Windows XP	S-PLUS	Exploratory data modeling and statistical analysis
4	B-08	Check quality of classification	Quality check. Evaluation of entropy & probabilities. If classification is NOT OK --> reclassification. If OK, no further processing needed	Quality OK	Manual	PC	Windows XP	S-PLUS	Exploratory data modeling and statistical analysis
4	B-09	Quality check final product	Final evaluation of entropy	Quality OK	Semi-automated	PC	Windows XP	S-PLUS	Exploratory data modeling and statistical analysis

Table 3: M2.5 Agricultural Land Use

ID	Name	Description/ Function	Quality Control criteria	Interaction	Hw Plat-form	Software used	To Report no.	Data used for QC	Respon-sible	Action if no-go
B-01		Check cloud coverage and radiometry of images	Cloud cover < 5%	Manual	PC	ERDAS / ER	R-01	Product specification, purchase order forms	Specialised technician	Repeat import; Contact data provider
B-02		Aereal Coverage	Image covers correct data	manual	PC	ERDAS / ER	R-02	Product specification, purchase order forms	Specialised technician	Repeat import; Contact data provider
B-03		Orthorectification quality	RMSE < 30 m	automated	PC	ERDAS / PCI	R-03	Spatial references, e.g. Orthoimages, GCP reports, Check point reports		Repeat orthorectification
B-04		Classification accuracy check	Accuracy > 80%	Manual, statistical	PC	ERDAS / ER /ARC Info	R-04			Repeat fusion process
B-05		Export Quality	Legibility of data	manual	PC	ARC Info	R-05	GIS compatible map	Specialised technician	Check reports, repeat process

Table 4: M3.1 Arable Acreages Map (medium resolution seasonal mapping)

ID	Name	Description/ Function	Quality Control criteria	Method	Software used	To Report no.	Data used for QC	Responsible	Action if no-go
B-01	MERIS image	Control quality of corrected images and cloud cover, and area covered	Radiance, clouds, location of the image	Visual	N/A	R-01	Purchase order forms	EOAS	Contact data provider
B-02	Geometry and area of interest	Control quality of images and correct extraction of UA's	Geometry correctness, correctness of extracted UA's (area of interest)	Visual	N/A	R-01	Area of interest (Units of analysis) shape	EOAS	Analyse shape files used in input and repeat ingestion
B-03	Biophysical	Control quality of model fit at biophysical inversion	Error of the inversion process (difference between modelled and observed spectral radiance)	Automatic detection of pixels with poor fit performance Control of biophysical results on images	Inhouse (included in Inversion SW)	R-02	Canopy reflectance model	EOAS	If not many pixels affected : Invalidation of pixels with poor fit performance If many pixels affected : (i) analyse image in order to detect specific atmospheric conditions entailing atmospheric model adaptation to be applied, (ii) analyse spectral radiance over various samples on the scene with Envi software
B-04	Mask of arable and pastures	Control correctness of raster mask, compared to CLC 2000	Mask of arable and pastures areas compared to CLC	Visual, acreages comparison in case of	N/A	R-03	CLC 2000	EOAS	Analyse topological configuration of arable and pastures objects that can be the cause of masking distorsions

ID	Name	Description/ Function	Quality Control criteria	Method	Software used	To Report no.	Data used for QC	Responsible	Action if no-go
				doubt					
B-05	Percentage	Control crops proportion	Proportions realistic with respect to local conditions Model Biophysical profiles fit compared to observed profiles	Tables analysis Plot and compare biophysical profiles (control fit between model and observed)	N/A	R-04	Agricultural statistics Crops development models	EOAS	Analyse observed profiles, recalibrate crops phenological models to local conditions if needed
B-06	Arable acreages	Verify acreages compared to agri stats	Comparison to agricultural statistics	Semi automatic comparison, visual inspection of GIS maps	Inhouse, spatialisation of agricultural statistics to obtain validation data	R-04	Agricultural statistics	EOAS, PM, PA, CUS	Analyse spatial variability of errors to detect possible geographical effects Review crops phenological models to local conditions if needed Analyse crops statistics validity (date, spatial distribution)

M3.2 Phenology (seasonal biophysical parameters)
No input was received for the product as production won't start till year 2 of the project and the processing chain may change. This document will be updated when the required information becomes available.

3.4 EXTERNAL QUALITY ASSURANCE / QUALITY CONTROL CONCEPT

3.4.1 Introduction

The service validation protocol comprises each single geo-information product. For the QA/QC team there are two different product types which will be handled differently – visual interpretation products and automatic / semi-automatic products.

The first group products are:

- M1.1 Urban Atlas Map (very high resolution mapping)
- M2.1 Regional Land Cover
- M2.4 Land Take Map (update / downdate)
- M2.6 Land Take Map (first inventory)

The products of this first group are created by the following main processing steps:

- Orthorectification of images – semi-automatic segmentation – preclassification – manual interpretation and checking. For this group it will be important to focus on the quality of the manually done work like interpretation but also on the so-called "correction" of the previous automatic or semi-automatic segmentation steps.

The second group products are:

- M2.3 Forest Parameters
- M2.5 Agricultural Land Use
- M3.1 Arable Acreages Map (medium resolution seasonal mapping)
- M3.2 Phenology (seasonal biophysical parameters)

The content of the products is very specific and highly time dependent. It therefore is very difficult to validate this information from a distance. Typical validation information is only available on site and therefore should be performed by the end user of the product.

The QA team sees no reasonable way of validating the above-mentioned products in a central way. During the user workshop in July 2006 it was agreed that these products will be evaluated by the end user only.

3.4.2 General concept

The external auditor (TÜV) will check the production chain itself and the adherence of the service providers to their self-defined production rules and quality breakpoints. The Corine Land Cover team will verify and validate the final mapping results (q.v. chapter 3.4.3).

The production manager will perform a general check for data homogeneity to avoid major differences in content and appearance of the individual products. This homogeneity is important for the production of the same product by different service providers as well as across different products which are produced for the same region.

The quality of the mapping products will be reviewed by the QA technical team. Two different quality control steps are recommended related to production:

- Verification, in order to improve the production, more or less at the 30 to 50% of the production. The aim of this verification (qualitative feedback) is to see whether the standards are kept and to guide some improvements. A verification will only be done once per product and phase to harmonise service provider and QA team visions about the products.
- Validation, with the aim of the final quality assessment, at the end of the production. The result of this step will be a qualitative expression of the actual data quality versus the proposed data quality.

In order to perform a really independent and realistic validation several points have to be considered:

- First, it is absolutely necessary to **use data that was not included in the processing chain**. Otherwise there will be dependencies in the result product which are not reproducible and above all, it will pretend a better result than it actually is.
- Second, the **sampling design has to be independent** from the quality control of the service provider.
- Third, the more **manual work** is done within the processing steps, the closer the result **has to be verified**.

Additional project internal boundary conditions for the development of the QA method were:

- Development of the methodological approach in parallel to the start of the production (normally the QA method should be known to the service provider before the start of production).
- Design of a sampling design and selection of validation areas before the end of the production in the service area, i.e. the actual land cover class distribution is not known when the validation areas are defined.
- European-wide applicability of the method.
- Minimisation of cost and effort for the provision of independent reference data.

The quality of the product will be described quantitatively. Standard accuracy measures nowadays are:

- Overall accuracy
- User accuracy (commission error),

- Producer accuracy (omission error)
- Kappa coefficient

During stage 1 of the SAGE, GUS and Coastwatch projects (whose results form the basis for the GSE Land project) the product quality was assessed by using a stratified random point sampling approach.

The approach proposed by GSE Land goes beyond the merits of a point sampling, i.e. the consideration of geometric and thematic accuracies. In order to guarantee the comparability of the results of the QA/QC in GSE Land, both measures (old and new) are provided. After phase 1 of GSE Land the point sampling can be provided by the QA team upon request by the service provider. The extra cost must be born by the service provider requesting this extra QA service.

3.4.2.1 Visual interpretation products

Considering the before-said, special attention will be given to the visual interpretation products. The validation of the result comprises therefore two criteria: the class assignment **and** the outline of the classified object. In other words: the thematic accuracy as well as the geometric accuracy of the delineated polygons will be validated. The above-mentioned criteria will help in showing potential weaknesses in special classes and the kind of problem. It will be necessary for the final product to meet both accuracy levels (geometry and content).

3.4.2.2 Automatic or semi-automatic products

The quality of the automatic or semi-automatic products will be difficult to assess, because no independent data to crosscheck the results will be available. It is therefore proposed to leave the validation of this product range to the end user. This was confirmed by the user workshop in Barcelona in July 2006.

3.4.3 External Quality Control

The external quality control will differ from the internal quality control. In the internal quality control, milestones are defined between the main steps to ensure the quality all along the processing chain. External quality control will only validate the final product to ensure it fits with the defined quality criteria.

External quality control comprises two main steps:

- Selection of the areas where quality control will be performed, i.e. **sampling**
- Doing the **quality control** itself over the samples.

Table 5: Accuracy levels (performance parameters) proposed for the different products

	CORINE Class(es) [Level I, No.]	Level(s) provided	MMU	Overall Thematic Accuracy	Positional Pixel accuracy
M1.1 Urban Atlas	1	I - IV	0,25 ha	>= 85%	<= 5m
	2 - 5	I	0,25 ha	>= 80%	
M2.1 Regional Land Cover	1	I - II	1 ha	>= 80%	<= 30 m
	2 - 4	I - II (III)	5 ha		
	5	I	5 ha		
M2.3 Forest Parameters	3	III - V	1 Pixel	NA	Pixel level
M2.4 Land take map (change)	1	I - II (plus density for built-up area)	0,25 ha	>= 95%	<= 20 m
	2 - 5	I	1 ha		
M2.5 Agricultural Land Use	2	II - IV	1 Pixel	>=80%	<= 30 m
M2.6 Land take map (first inventory)	1	I - II (plus density for built-up area)	0,25 ha	>= 80% (based on cluster method) >= 95% (based on point validation)	<= 10 m
	2 - 5	I	1 ha		

3.4.3.1 Sampling design

The sampling procedure for QC will be based upon the following considerations:

- It will cover a representative amount of territory

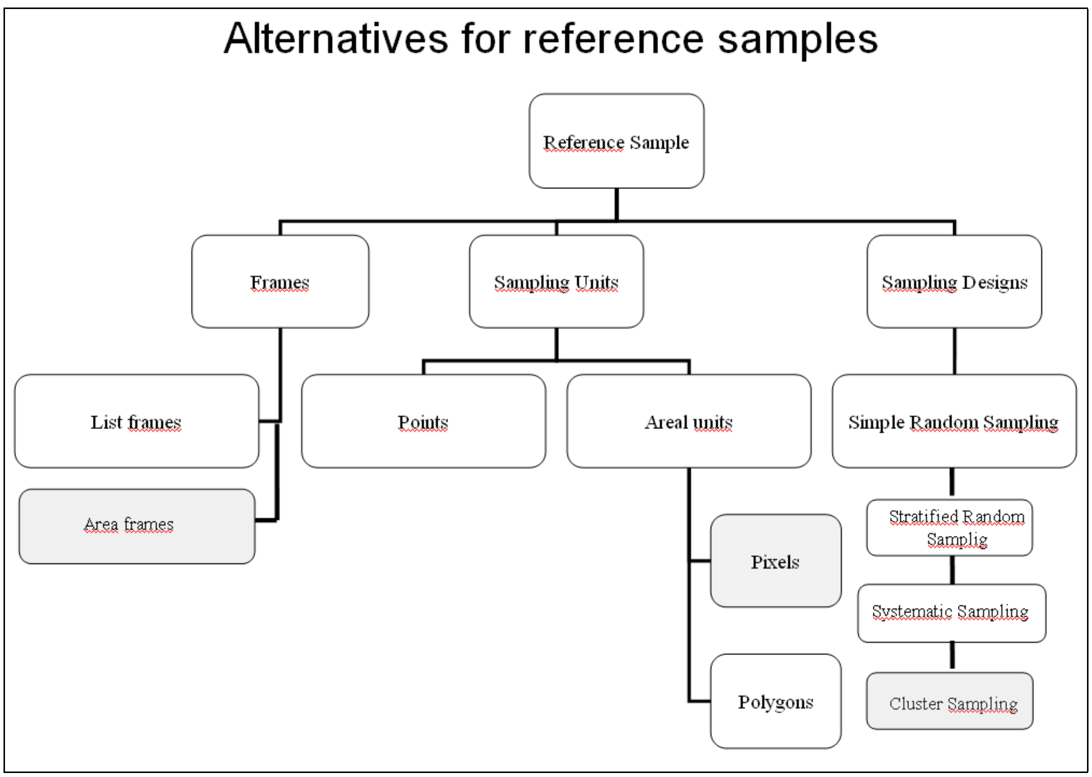
- If more than one SP is producing one mapping product, sampling will cover all the SPs with a similar share over the overall product.

When many different areas are produced (i.e. cities), it's not necessary to sample each area if all the above issues are covered. In the case of the urban atlas product it will be ensured that a sample of mapping results for each service provider is quality controlled.

Sampling Units and Sampling Design

The selection of the reference sample requires the specification of the sampling design, the definition of the sampling frame and the decision about the sampling units. The following figure presents different alternatives; the boxes in grey shade show the alternatives selected for the sampling selected for the GSE Land QA concept:

Figure 6: Alternatives for reference samples



Before a sample can be selected the Population must be divided into parts that are called ***sampling units*** (Cochran, 1977[1]). The sampling units must cover the entire population and every element in the population belongs to one and only one sampling

[1] Cochran, W.G., 1977: Sampling Techniques, John Wiley & Sons, New York, 428 p.

unit. The sampling unit is the fundamental unit on which the accuracy assessment is based. For each sampling unit the response design is applied to obtain the reference and the comparison of the map and reference classification is conducted on the scale of the sampling unit. Sampling units can be points or areal sampling units (pixels, polygons, fixed area plots). In the concept presented here a cluster of minimum mapping units (MMUs), which can be interpreted as pixels, are selected as sampling units. The size of the cluster is 25*25 MMUs, which cover an area of 156,25 ha at 0.25 ha MMU respectively 625 ha at 1 ha MMU.

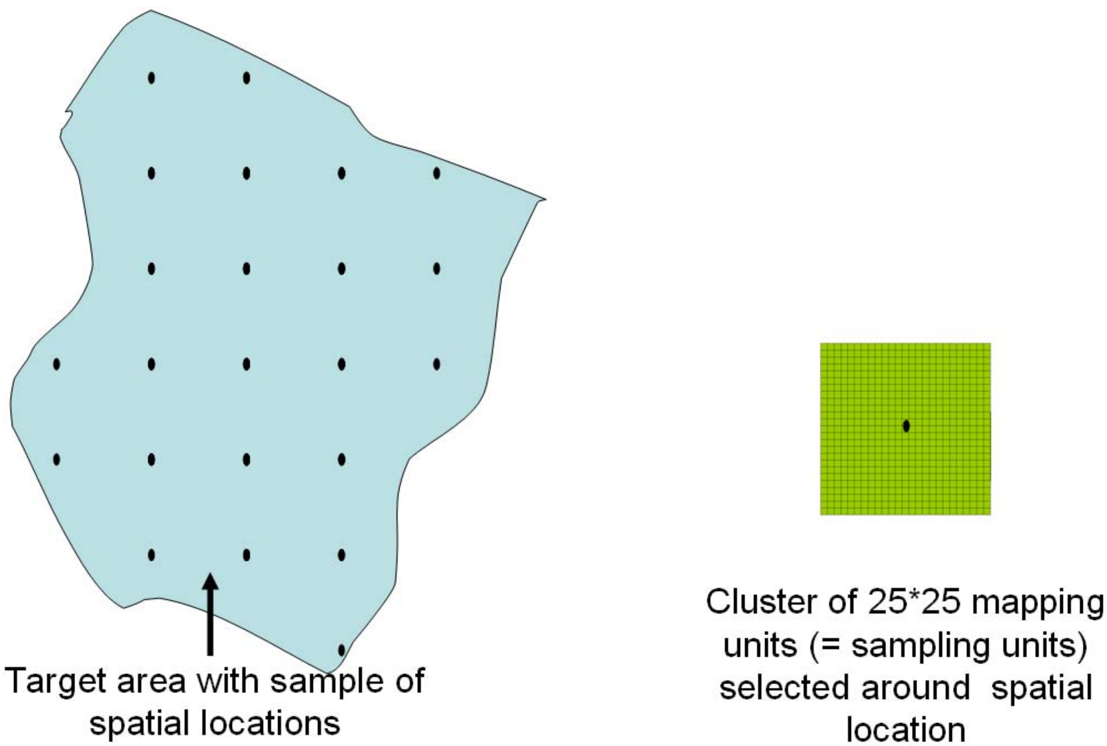
According to Särndal et al.[2] (1992) a ***frame*** consists of the materials or devices which delimit, identify and allow access to the elements of the target population". In traditional surveys frames have been seen as the construction of a list of the sampling units in the target population, i.e. list frames. The sampling element is directly sampled from the list. In areal lists the sampling elements are selected indirectly. First a sample of spatial locations is selected and then a sampling unit is associated with each sample location. This requires an explicit rule for associating a spatial location with a sampling unit. In the application described here the spatial locations are systematically distributed in a 4*4km grid and an area frame of 25*25 MMUs is constructed around each spatial location (3.4.3.1). For each of the selected MMUs the response design is applied to get the reference classification.

At an initial stage a random sample of clusters selected. In order to reduce the time and cost needed for the assessment of reference data, the pixels are grouped to clusters of 25*25 MMUs.[3]

[2] Särndal, C.E., Swensson, B., Wretman, J., 1992: Model-Assisted Survey Sampling, Springer-Verlag, New York

[3] An alternative would have been to choose a stratified sampling design. For the selection of samples and the analysis and evaluation of results the strata sizes need to be known. The selection of individual pixels (per stratum) would require costly handling of reference data. Insufficient accuracies, especially one-sided bias, of the image is reflected by the determination of strata sizes and would introduce the risk to over-or underestimate true strata sizes. In addition the analysis procedures become rather complex with post-stratification and unequal selection probabilities in the individual strata. The approach selected allows for more routine in the sample selection and reduces the cost for handling of reference data.

Figure 7: Sampling frame and sampling units

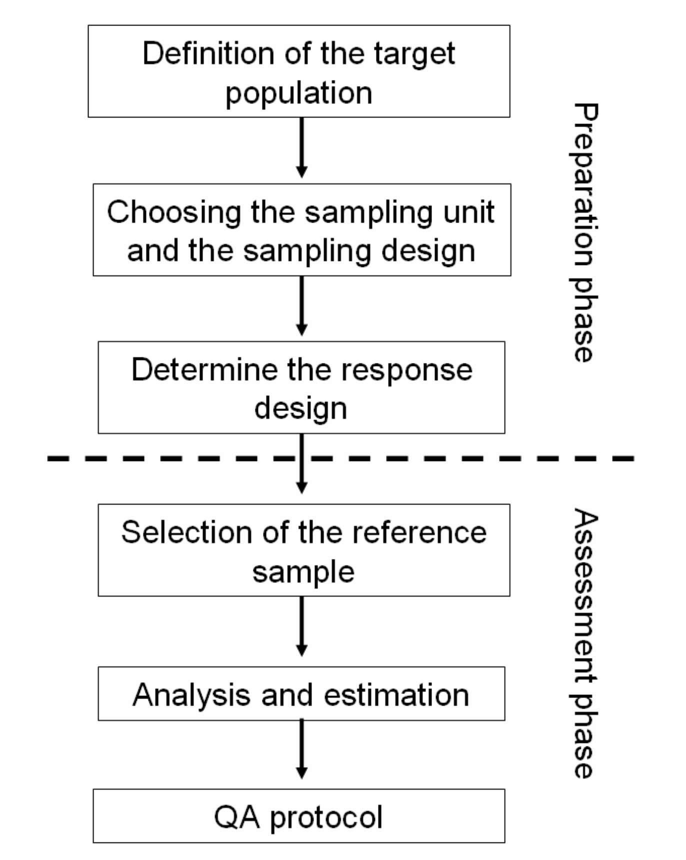


The protocol by which units are selected into the sample is called the ***sampling design***. A scientifically defensible accuracy assessment requires a probability sampling design, which is defined in terms of inclusion probabilities. Inclusion probabilities represent the probability of including a particular sampling unit in the sample. Probability sampling requires that all inclusion probabilities be greater than zero and must be known for those units selected in the sample. If some sampling units have a zero section probability, the assessment does not represent the entire target population. Simple random sampling, stratified sampling, systematic sampling, and cluster sampling are all probability sampling designs. When using such designs the inclusion probabilities do not have to be computed as they are already included in the estimation procedures.

In order to improve the efficiency of the sampling design applied for accuracy assessment one-stage ***cluster sampling*** was chosen as sampling design. The ability to sample several reference sites in close proximity allows generating a larger sample size for fixed cost compared to selecting individual pixels.

The thematic accuracy assessment begins with the definition of the ***target population***, which is the area or region presented by the land-cover map. The individual units (or elements) of the population depend on the map representation and are either pixels or polygons. A sample of this population is selected from the population for the thematic accuracy assessment. The target population, i.e. the population for which the thematic accuracy of the map has to be presented, and the sampling population, i.e. the population from which the sample is taken, need to be identical. It is important to verify this constraint. Differences may occur by several reasons, e.g. limited spatial cover of reference data, inaccessible areas that do not allow for field checks, or different assessment periods of the data used for map production and data used as reference that allow for land cover changes.

Figure 8: Thematic accuracy assessment – concept



Choosing the sampling unit and the sampling design are two major decisions required when planning the sampling protocol and determine the analysis and estimation procedures utilized for the accuracy assessment.

The ***response design*** includes procedures to collect information pertaining to the reference land-cover determination, and rules for assigning classes to each sampling unit. The system of nomenclature for different classes has been designed in an earlier stage of the map production, but need to be adjusted to the specific sources of the reference data. For example, where aerial photography is to be used as a reference the interpretation rules have to be adjusted so that the same classes are obtained as with digital image classification.

Once the target population has been defined, the sampling unit and the sampling design are chosen and the response design is laid down the preparation phase of the thematic accuracy assessment is finalised. In the assessment phase the reference sample is selected according to the chosen sampling design and for each sampling unit a class is assigned according to the response design. The map to be verified and the reference sample are merged in the analysis and estimation procedures. The final output of the assessment phase is the Thematic accuracy protocol that contains a confusion matrix and Kappa coefficients as major components.

The selected sampling design and sampling units have already been defined in the previous section. The analysis and evaluation procedures and the preparation of the Thematic accuracy protocol are described in the following section.

Analysis and Evaluation

The following parameters proved helpful for accuracy assessments and will be provided:

- Overall proportion of pixels classified correctly, P_c
- User's accuracy
- Producer's accuracy
- Kappa coefficient of agreement, κ
- Conditional Kappa
- Commission error (conditional probability that a pixel classified as cover-type I is actually cover-type j)
- Omission error (conditional probability that a pixel that is actually cover-type j is classified as cover-type i)

The estimation formulas under cluster sampling are the same as those for random sampling. However, as the presented parameters are estimates obtained from a sample, they are subject to error. The reliability of the parameters can be specified by the associated sampling error. The standard errors for cluster sampling and for simple random sampling are different. The equations presented below assume random distribution of the clusters within the population. In order to cover the entire population and to avoid the risk that sample locations are (randomly) clustered and do not cover the entire population, a systematic allocation of clusters is chosen. The equations shown below result in a slight underestimation of the true variances.

Under cluster sampling the error matrix for an individual cluster is summarized in the following table, when the error matrices of all m cluster are combined the usual sample error matrix can be derived (next table)

Table 6: Error matrix for Cluster u

		Reference				Row total
		1	2		q	
Classified	1	a_{u1}	b_{u12}	......	b_{u1q}	r_{u1}
	2	b_{u21}	a_{u2}	......	b_{u2q}	r_{u2}
	:	:	:		:	:
	:	:	:		:	:
	q	b_{uq1}	b_{uq2}	....	a_{uq}	r_{uq}
Column total		c_{u1}	c_{u2}	......	c_{uq}	t=number of pixels in cluster

Table 7: Error Matrix Combining Data from all m Sample Clusters

		Reference				Row total
		1	2		q	
Classified	1	n_{11}	n_{12}	......	n_{1q}	n_{1+}
	2	n_{21}	n_{2}	......	n_{2q}	n_{2+}
	:	:	:		:	:
	:	:	:		:	:
	q	n_{q1}	n_{q2}	....	n_{q}	n_{q+}
Column total		n_{+1}	n_{+2}	......	n_{+q}	n=mt

From the previous table accuracy parameters such as Pc, κ, or producer's and user's accuracy can easily be obtained; but for calculating sampling errors this table is not sufficient. Data need to be rearranged as shown in the following table.

Table 8: Data structure needed to calculate Sampling Errors under a Cluster Sampling Design

Sample cluster	d_u	Row Totals			Column totals			Number on diagonal			Off diagonal
		r_{u1}	…	r_{uq}	c_{u1}	…	c_{uq}	a_{u1}	…	a_{uq}	b_{uij}
1	d_1	r_{11}	…	r_{1q}	c_{11}	…	c_{1q}	a_{11}	…	a_{1q}	b_{1ij}
2	d_2	r_{21}	…	r_{2q}	c_{21}	…	c_{2q}	a_{21}	…	a_{2q}	b_{2ij}
3	d_3	r_{31}	….	r_{3q}	c_{31}	….	c_{3q}	a_{31}	….	a_{3q}	b_{3ij}
:	:	:		:	:		:	:		:	:
:	:	:		:	:		:	:		:	:
:	:	:		:	:		:	:		:	:
m	d_m	r_{m1}	…	r_{mq}	c_{m1}	…	c_{mq}	a_{m1}	…	a_{mq}	b_{mij}
Sum	$\sum_{k=1}^{q} n_{kk}$	n_{1+}	…	n_{q+}	n_{+1}	…	n_{+q}	n_{11}	…	n_{qq}	n_{ij}

The following notation is used for the equations needed for the accuracy assessment.

q = number of land-cover classes

u = cluster identity

t = number of pixels in each cluster (e.g. t=625 for a 25x25 cluster of MMUs)

M = number of clusters in entire map

N = number of pixels in entire map

m = number of clusters in sample

n = number of pixels in sample (n=mt)

f = m/M = n/N = sampling fraction (can be ignored is f is sufficiently close to 0)

d_u = number of pixels classified correctly from cluster u (regardless of land-cover class)

r_{ui} = pixel total of row i in cluster u

c_{uj} = pixel total of column j in cluster u

a_{ui} = number of pixels on the diagonal in row i of the error matrix for cluster u

b_{uij} = number of pixels in row I, column j of the error matrix for cluster u

A “^” is used to denote an estimator of a parameter obtained from the sample. Given the data structure summarized under the previous table the parameters shown below can be calculated. As a user does usually not require the same level of accuracy for each individual class, parameters are presented that provide the accuracy for individual classes. In combination with the respective variance of the parameter a confidence interval can be constructed that gives the likely range within which the true parameter is located with a defined (say 95%) probability. Where the confidence interval is too wide for a meaningful decision additional efforts are required.

Equation 1: *Overall proportion of pixels correctly classified, P_c*

$$\hat{P}_c = \sum_{u=1}^{m} d_u / n = \sum_{k=1}^{q} n_{kk} / n$$

$$\operatorname{var}(\hat{P}_c) = \frac{1}{t}\left(\frac{N-n}{nN}\right) s_d^2$$

where

$$s_d^2 = \frac{\sum_{u=1}^{m} d_u^2 - m\bar{d}^2}{m-1} \; ; \; \bar{d} = \frac{\sum_{u=1}^{m} d_u}{m}$$

Equation 2: *Producers’s accuracy for cover-type (column) j, P_{Aj}*

$$\hat{P}_{Aj} = \frac{\text{number on diagonal in column j}}{\text{total of column j}} = \frac{\sum_{u=1}^{m} a_{uj}}{\sum_{u=1}^{m} c_{uj}} = \frac{n_{jj}}{n_{+j}}$$

$$\operatorname{var}(\hat{P}_{Aj}) = \frac{1-f}{m\bar{c}_j^2} \frac{\sum_{u=1}^{m} \left(a_{uj} - \hat{P}_{Aj} c_{uj}\right)^2}{m-1}$$

where

$$\bar{c}_j = \frac{\sum_{u=1}^{m} c_{uj}}{m}$$

Equation 3: *Users's accuracy for cover-type (row) i, P_{ui}*

$$\hat{P}_{Ui} = \frac{\text{number on diagonal in column i}}{\text{total of row i}} = \frac{\sum_{u=1}^{m} a_{ui}}{\sum_{u=1}^{m} r_{ui}} = \frac{n_{ii}}{n_{i+}}$$

$$\operatorname{var}\left(\hat{P}_{Ui}\right) = \frac{1-f}{m\bar{r}_i^2} \frac{\sum_{u=1}^{m} \left(a_{ui} - \hat{P}_{Ui} r_{ui}\right)^2}{m-1}$$

where

$$\bar{r}_i = \frac{\sum_{u=1}^{m} r_{ui}}{m}$$

Equation 4: *Kappa coefficient*, κ

$$\hat{K} = \frac{n \sum_{k=1}^{q} n_{kk} - \sum_{k=1}^{q} n_{k+} n_{+k}}{n^2 - \sum_{k=1}^{q} n_{k+} n_{+k}}$$

$$\operatorname{var}\left(\hat{K}\right) = M^2 (1-f) s_w^2 / m$$

$$s_w^2 = \frac{\sum_{u=1}^{m} \left(w_u - w\right)^2}{m-1}$$

$$w_u = \frac{n^3}{ND^2} \left[\frac{Dd_u}{n} + \left(\hat{P}_c - 1\right) \sum_{k=1}^{q} \left(n_{+k} r_{uk} + n_{k+} c_{uk}\right) \right]$$

$$D = n^2 - \sum_{k=1}^{q} n_{k+} n_{+k}$$

Equation 5: *Conditional kappa for row i, κ_i*

$$\hat{K}_i = \frac{n n_{ii} - n_{i+} n_{+i}}{n n_{i+} - n_{i+} n_{+i}}$$

$$\operatorname{var}\left(\hat{K}_i\right) = M^2 (1-f) s_v^2 / m$$

$$s_v^2 = \frac{\sum_{u=1}^{m} \left(v_u - \bar{v}\right)^2}{m-1}$$

$$v_u = \frac{n^2}{N n_{i+} \left(n - n_{+i}\right)} \left[a_{uj} - \frac{n_{ii}}{n_{i+}} r_{ui} - \frac{n_{i+} - n_{ii}}{n - n_{+i}} c_{uj} \right]$$

Equation 6: *Conditional kappa for column j, κ_j*

$$\hat{K}_j = \frac{n n_{jj} - n_{j+} n_{+j}}{n n_{+j} - n_{j+} n_{+j}}$$

$$var\left(\hat{K}_j\right) = M^2 (1-f)\, s_v^2 / m$$

$$s_v^2 = \frac{\sum_{u=1}^{m} (v_u - \overline{v})^2}{m-1}$$

$$v_u = \frac{n^2}{N n_{+j}\left(n - n_{j+}\right)}\left[a_{uj} - \frac{n_{jj}}{n_{+j}} c_{uj} - \frac{n_{+j} - n_{jj}}{n - n_{j+}} r_{uj}\right]$$

Equation 7: *Commission error for row i, column j, C_{ij}*

The commission error for row i, column j gives the conditional probability that a pixel classified as cover-type I is actually cover-type j.

$$\hat{C}_{ij} = \frac{\text{number in row i and column j}}{\text{total of row i}} = \frac{\sum_{u=1}^{m} b_{uij}}{\sum_{u=1}^{m} r_{uj}} = \frac{n_{ij}}{n_{i+}}$$

$$var\left(\hat{C}_{ij}\right) = \frac{1-f}{m \overline{r}_i^2} \frac{\sum_{u=1}^{m} \left(b_{uij} - \hat{C}_{ij} r_{ui}\right)^2}{m-1}$$

Equation 8: *Omission error for column j, row i, O_{ij}*

The omission error for column j, row i gives the conditional probability that a pixel that is actually cover-type j is classified as cover-type i.

$$\hat{O}_{ij} = \frac{\text{number in row i and column j}}{\text{total of column j}} = \frac{\sum_{u=1}^{m} b_{uij}}{\sum_{u=1}^{m} c_{uj}} = \frac{n_{ij}}{n_{+j}}$$

$$var\left(\hat{O}_{ij}\right) = \frac{1-f}{m \overline{c}_j^2} \frac{\sum_{u=1}^{m} \left(b_{uij} - \hat{O}_{ij} c_{uj}\right)^2}{m-1}$$

Thematic accuracy protocol

The Thematic accuracy protocol summarizes the results of the analysis and evaluation step. It provides three major overviews:

- The error matrix
- Estimators and estimated standard errors for parameters related to the accuracy of the entire classification
- Class specific estimators

The output of the Thematic accuracy protocol is given below. The shaded areas present the results for an example with 3 land-cover types.

The error matrix presents the number of observations for all q*q combinations of the reference and image classes. The diagonal shows the correctly classified pixels.

Table 9: *Error Matrix*

		Reference			Row total
		A	B	C	
Image	A	156	51	24	231
	B	67	72	10	149
	C	16	33	71	120
Column total		239	156	105	500

The following table presents estimators and standard errors, which are related to the entire classification. These are the number of pixels correctly classified and the Kappa coefficient. As both parameters are derived from a sample it is possible to calculate their standard errors and to construct the resulting 95%-confidence interval. The confidence interval gives the range within which in 95 out of 100 samples the true parameter is located.

Table 10: *Estimators and Estimated Standard Errors*

Estimator	**Estimate**	**Standard Error**	**Confidence Interval**
Overall proportion of pixels correctly classified, P_c	0.598	0.0451	$0.508<P_c<0.688$

Kappa coefficient, κ	0.368	0.0721	$0.224<\kappa<0.512$

The third table in the Thematic accuracy protocol provides information on individual categories. As in the latter table standard errors and confidence intervals facilitate interpretation and allow to identify land-cover types for which the accuracy is not sufficient or the estimator is not precise enough for a final decision.

Table 11: *Class-Specific Estimators*

		Category		
Estimator		**A**	**B**	**C**
Producers's accuracy for cover-type (column) j	$\hat{P}_{Aj}$	0.653	0.462	0.676
	$\sqrt{var(\hat{P}_{Aj})}$	0.0403	0.0704	0.0486
	CI	$0.57<\hat{P}_{Aj}<0.73$		
Users's accuracy for cover-type (row) i	$\hat{P}_{ui}$	0.675		
	$\sqrt{var(\hat{P}_{ui})}$	0.0424		
	CI	$0.59<\hat{P}_{ui}<0.76$		
Conditional kappa for column j	$\hat{\kappa}_j$	0.378		
	$\sqrt{var(\hat{K}_j)}$	0.0795		
	CI			
Commission error for row i, column j	$\hat{C}_{ij}$ C_{ij}			
	$\sqrt{var(\hat{C}_{ij})}$			
	CI			
Omission error for column j, row i	$\hat{O}_{ij}$			
	$\sqrt{var(\hat{O}_{ij})}$			
	CI			

Where parameter estimates are given without information on their reliability (i.e. standard errors or confidence intervals), it is difficult to derive any decision about the appropriateness of the classification. The desired accuracies are depending on the user requirements and might be different for individual classes. The confidence intervals facilitate the interpretation of the accuracy assessment, as they allow for treating classes individually. The obtained accuracy for a class may either be accepted/ rejected, or – if the confidence interval is too large for the specific class – may require to assess additional reference points. Therefore, an objective accuracy assessment requires the a priori specification of both, the desired value for each parameter (by land-cover type) and the range within which a parameter can be accepted (e.g. 92% $\pm$ 2% of all pixels need to be classified correctly, or the conditional kappa for land-cover type “A” has to be larger than 0.45 $\pm$ 0.1). As the standard errors and confidence intervals can be improved by increasing the sample size, it may be necessary to increase the number of references for a specific land-cover type.

3.4.3.2 Quality control

The main objectives of the implementation of the thematic accuracy assessment are to provide (a) objective information on the map accuracy, and (b) a cost-efficient system for accuracy assessment. To find an optimum of the required number of references to achieve information on accuracy with desired reliability (i.e. precision, width of the confidence interval) is not an easy task, as many factors are influencing this decision, e.g. the desired level of accuracy the spatial fragmentation of land-cover classes, the length of boundaries between cover-types and the resulting probability of mixed pixels, the number of land-cover types, or the proportion of a distinct land-cover type within the target area. The a-priori calculation of necessary sample sizes is a complex task and requires a comprehensive list of input information.

A key requirement for the development of the sampling design was to select the validation areas *before* the end of the production, i.e. without knowledge about the actual distribution of land cover classes.

Implementation of the QA procedure

Due to the complexity of the scientific approach on the one hand and due to the need for a reliable and quick pragmatic solution, the following two-stage validation approach is proposed which is mainly based on cluster sampling and iterative cascades of independent samples.

Quality control of the mapping products should be performed using independent high-resolution information, ideally aerial photographs or very high resolution satellite images, like IKONOS or Quickbird will be used for this step. The QA team will request the availability of such reference information.

The validation process itself will be based on a “blind” re-interpretation of the land cover considering the available reference material as well as the original satellite images used by the service provider, without previous image pre-processing such as segmentation. This step will be performed by the QA technical team.

For each of the selected clusters (25x25 MMU) the technical team will, independently from the service provider interpretation, manually delineate land cover objects ("reference interpretation") following exactly the same interpretation guidelines for each respective product.

The area of each single cluster of the vector data set of the reference interpretation as well as the service provider interpretation will be rasterised (vector to raster conversion using centre point of the grid cell for class assignment) according to the respective MMU of the different classes[4]. Both raster data sets will be overlaid in a GIS and the number of correctly/incorrectly classified grid cells for each class within each cluster will be counted and the error matrices for each cluster and each product can be created. (see Table 9). By this error matrix the thematic as well as the geometric accuracy can be demonstrated.

Based on the 2.5*2.5 km grid (1.25*1.25 km for M1.1), the selection of an initial sample of 1% of all available clusters will be drawn. The clusters will be selected using random sampling.

After the initial samples have been selected and the analysis and validation has been carried out, an initial accuracy protocol is prepared. This accuracy protocol facilitates the selection of land-cover types with sufficient accuracy and the identification of land-cover types with insufficient (thematic) accuracy. The main objective of the initial QA protocol is to quantify the overall accuracy. When the specifications are met the QA process stops and the QA of the product can be approved.

The acceptance of the overall accuracy is based on the comparison of the desired accuracy and the estimated value for the overall accuracy. The desired accuracy is expressed as a range of ± 3% around the value specified (e.g. for a desired accuracy of 90% a range of 87% to 93%). This range is compared with the estimated confidence interval. Where the desired values of the accuracy parameters are within the confidence interval, but the confidence interval itself is found to be too wide (+/- 3%), additional samples are required for a final validation of the accuracy.

Table 12: Examples of accepted / rejected accuracies

Desired overall accuracy	± 3% -range around desired accuracy	Estimated confidence interval for overall accuracy	Decision
80%	77% - 83%	67% - 75%	Rejected
80%	77% - 83%	70% - 86%	Rejected
80%	77% - 83%	78% - 90%	Accepted
80%	77% - 83%	85% - 88%	Accepted

[4] Different minimum mapping units in one product will be rasterised at the resolution of the respective class. Accuracy parameters will be calculated for each class.

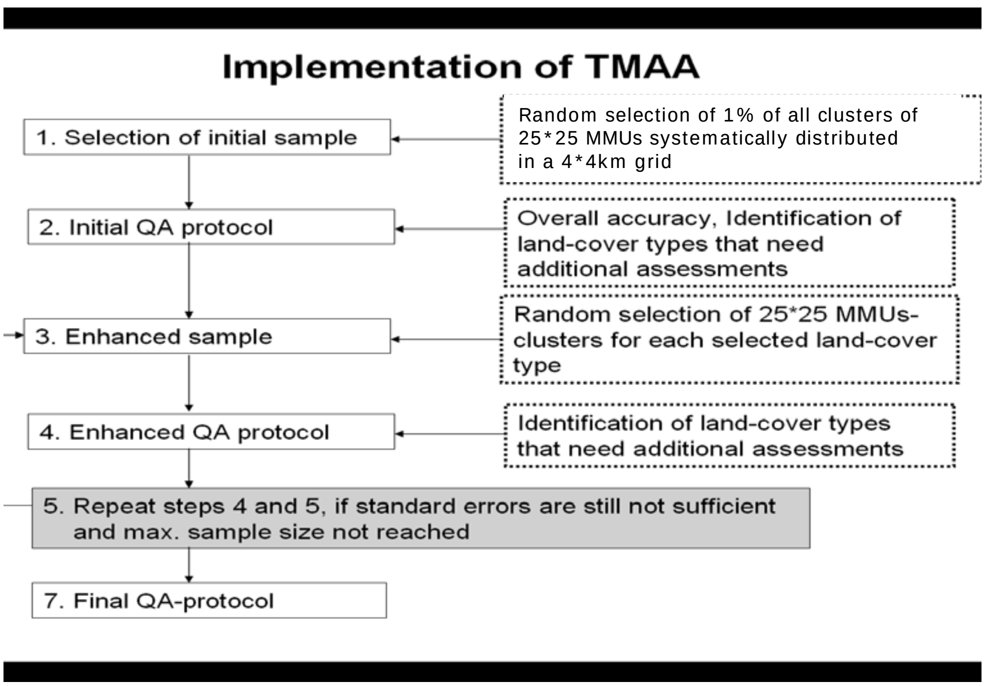
In addition rare land-cover types present in the classification but not included in the initial sample need to be included in the enhanced sample. To ensure that the real quality of the classification is assessed the enhanced sample will also contain typical confusion classes (classes that are confused often in a classification process) to assess errors of omission and errors of commission.

If the user agrees, rare land cover types (of minor importance for the end use of the product) can be left aside for the further evaluation.

For those classes that do not fully meet the accuracy requirements (e.g. too few samples, accuracy insufficient, confidence interval too wide) additional samples are drawn using a targeted sampling to ensure that the critical classes are contained in the selected clusters.

Where the quality assurance process makes reference to the overall accuracy *only* and the product is not rejected following the decision rules given in the example above it is sufficient to extract further samples from the entire set without taking into account the individual land cover types.

Figure 9: Thematic accuracy assessment – implementation



A new accuracy protocol is prepared for each land-cover type included in the enhanced sample. Where the accuracy protocol still does not allow a decision (i.e. confidence interval still too wide), another enhanced sample for the land-cover type in question is taken. The enhanced sample is terminated when the comparison of the confidence intervals and the range of desired precision allows a consistent decision about the compliance with the desired accuracy.

The following criteria will be used to abort additional sampling:

- the total sample size reaches 5% of the area, i.e. 5 times the initial sample size;
- the confidence interval does not change anymore significantly;
- the accuracies are far below the specifications and additional samples do not change this.

The findings for the initial and enhanced accuracy protocol are summarized in a final accuracy protocol. The protocol will include all accuracy measures mentioned above. For the user acceptance the overall accuracy of the product will be considered, the remaining measures are mostly for internal use by the service providers and will be forwarded to the end user only upon request.

Lessons learnt from phase 1

The first year of GSE Land has shown that the cluster method is rather work intensive and includes a certain subjectivity with respect to class delineation and identification. Especially in the case of the M1.1 product where external boundaries (TeleAtlas street network) are incorporated in the final result expert delineation of cluster boundaries are often not the same. This may lead to a rather high degree of disagreement.

Furthermore, the product specific accuracy specifications were defined based on a point sampling approach, not on the blind remapping of the cluster validation. Upon request by the service providers for some products both approaches (point and cluster) were applied to the data. In general the results were rather similar, even though in some specific cases the point approach met the accuracy specifications which were not fully met by the cluster approach.

It was agreed that for phase 2 and 3 both approaches can be requested by the service providers.

3.4.3.3 Specific remarks

Validation of changes

For the validation of changes (M2.4) no scientifically proven method exists today.

The approach implemented at this stage consists of a blind remapping of limited areas of the M2.4 product. For this limited area the errors of commission as well as errors of omission can be calculated. Nevertheless this approach is not fully representative as changes are not necessarily regularly distributed in the whole data set.

Currently a team of experts is reviewing this issue in the framework of the B4G project. The objective is to propose a more representative method for the validation of changes.

It was clarified that the M2.4 data set does not only contain change polygons, but that it is a full coverage map of the area of interest. This has an impact on the accuracy specification as even a 50% omission of changes (given a total amount of 2% of changes) would still allow to meet the overall 95% accuracy specification.

Urban Atlas

Due to the level of detail (level 4) and the land use aspect of the M1.1 product, the QA team proposed to validate (in phase 1) M1.1 only at level 3 of the nomenclature. Level 3 can be interpreted from the image data (including additional ortho-photos and topographic maps) while level 4 requires additional information such as city maps or cadastral information.

After intensive discussion with DG Regional Policy a new nomenclature for the Urban Atlas was agreed. The interpretation of the new nomenclature will be possible from EO data and limited ancillary information (TeleAtlas street network and topographic maps). This way the “new Urban Atlas” will represent a European minimum standard for city mapping.

The sealing levels in classes 1.1.2.1 – 3 will in the future be obtained from the FTS sealing layer.

Ortho-photos **must not** be used by service providers for the delineation of objects.

The QA team will validate the new nomenclature at the most detailed level.

Due to the specific nature of M1.1 (incorporation of external boundaries, land use character), the product will be validated by a point validation only.

- A regular grid with some 900 points (number might be adapted to actual size of mapping area) will be created covering the whole mapping area.
- The mapping (object delineation and interpretation) must always be done at the lowest level of the nomenclature. As the nomenclature is adapted to EO data and TeleAtlas information all existing classes and levels must be mapped.
- Validation will also be done at the lowest level.
- The identification of the point class code will consider the context and the MMU in which the point is located.
- Validation points which fall within a distance of less than 2.5m to an existing border will be removed from the reference interpretation. To compensate for this removal the reference interpretation will contain an oversampling of some 10%.

  The number of points removed due to this border conflict cannot exceed 15-20% of the total number of validation points.
- Overall accuracies for urban and non-urban classes will be calculated separately. The overall accuracy will be used for the benchmarking. Specs: 85% for urban classes (class 1) and 80% for classes 2 to 5.
- Individual class accuracies for all classes will be calculated. No class needs to achieve a minimum accuracy.
- Any material used by the service providers for the interpretation must be made available to the QA team.

- The QA team will not accept class codes different to their own interpretation if those codes can only be proven from additional undisclosed sources.
- Data delivered to the QA team must be in vector format and topologically correct (no unclosed polygons, no double labels). Data with incorrect topology will be returned only once to the service provider for correction.

Urban Atlas pre-qualification

DG Regional Policy and ESA have requested a benchmark for urban service providers. In this test 2 cities (Bremen and Badajoz) will be mapped by all interested service providers using the agreed European minimum standard nomenclature. Five independent experts were selected into the QA team to establish a reference interpretation based on a regular point grid (with different densities for urban and non-urban classes). The reference interpretation will then be compared to the service provider work.

The details of the process are described in the latest version of the mapping guide for the Urban Atlas.

Validation of automatic products

Due to the general lack of reference data sets for most of the automatic mapping products, the QA team proposes to entrust the validation of the products

- M2.3 Forest Parameters
- M2.5 Agricultural Land Use
- M3.1 Arable Acreages Map (medium resolution seasonal mapping)
- M3.2 Phenology (seasonal biophysical parameters)

to the end user. While for M2.5 a “simple” visual reinterpretation of the original satellite data could provide a means for determining the product accuracy, for the remaining products not even that seems a feasible option. The results of these products are so specific that they are impossible to validate centrally. If necessary validation is available, then this will be at the user premises.

4.SERVICE VALIDATION TEMPLATE

A. Product and service provider identification

PRODUCT NAME:	
SERVICE PROVIDER:	

B. Documentation (presence/absence)

Required document	**Present?**	**Comments**
Quality Assurance Plan (including processing chain, quality control milestones, and quality criteria)		
Interpretation guidelines (only for visual interpretation products)		
Nomenclature (only for visual interpretation products)		
Technical note on chief interpreters training (only for visual interpretation products)		
Technical note on interpreters training (only for visual interpretation products)		
Technical note on processing (including reports on quality controls)		
Service validation document		

C. Quality Assurance Plan (from the Service Provider)

Question	**Yes or no**	**Details (what's missing, comments)**
Are the quality control milestones defined?		
Are the required minimum quality control milestones included?		
Is there any quality criteria defined?		
Do the defined quality criteria fit in the minimum quality criteria defined in the		

GSE Land QA concept?		
Does the QA plan include a detailed processing chain?		
Is the agreed production chain followed correctly?		
Are quality control breakpoints observed and results documented?		

D. Technical note on chief interpreters training (only applies for visual interpretation products)

Question	Yes or no	Details (what's missing, comments)
Has the chief(s) interpreter(s) received the required training?		

E. Technical note on interpreters training (only applies for visual interpretation products)

Question	Yes or no	Details (what's missing, comments)
Have the interpreters received the required training from the chief interpreter(s)?		

F. Technical note on processing

Question	Yes or no	Details (what's missing, comments)
Has the service provider made the quality controls defined in the quality assurance plan?		
Are the quality criteria met after each quality control breakpoint?		
Has the service provider produced accordingly to the processing chain defined in the QA plan?		

G. Service validation document

Question	Yes or no	Details (what's missing,

		comments)
Has the service been validated by the QA technical team?		
Has the service been validated by the users, according to the requisites they defined?		

