SIGMA

Stimulating Innovation for Global Monitoring of Agriculture

PROTOCOL FOR LAND COVER VALIDATION

Reference: ***SIGMA_D33.2_Protocol for land cover validation***

Author(s): EFTAS & FAO

Version: 2.0

Date: 30/04/2015

DOCUMENT CONTROL

Signatures

Author(s): Carsten Haub, Luca Kleinewillinghöfer & Virginia Garcia Millan (EFTAS)

Antonio Di Gregorio (FAO)

Reviewer(s): John Latham, , Edi Kollozaj & Renato Cumani (FAO)

Elisabetta Carfagna (University Bologna)

Javier Gallego & Olivier Leo (EC JRC)

Qinghan Dong (VITO)

Approver(s): Sven Gilliams (VITO)

Issuing authority: VITO

Change record

Release	Date	Description	Editor(s)/Reviewer(s)
0.0	21/08/2014	Draft concept	Haub, Kleinewillinghöfer (EFTAS)
0.1	09/12/2014	Internal draft data collection and validation protocol	Haub, Kleinewillinghöfer (EFTAS)
1.0	31/03/2015	Revised data collection and validation protocol	Haub, Kleinewillinghöfer (EFTAS), Latham, Di Gregorio, Kollozaj, Cumani (FAO), Carfagna (University Bologna), Gallego & Leo (JRC)
2.0	22/06/2015	Revised data collection and validation protocol	Haub, Kleinewillinghöfer, Garcia Millan (EFTAS) & Di Gregorio (FAO)

TABLE OF CONTENT

Executive summary

In the last decades the interest in application of remote sensing technology to model the status of land cover and monitoring its changes has been growing continuously. In addition the application of these technologies has been promoted as a powerful tool for agricultural statistics. The fundamental step is the generation of updated and reliable datasets as a valid baseline. Even if the importance of this is recognised, many users of land cover information lack access to valid, updated and reliable land cover data. In this light it is of the uttermost importance to define proper validation routines to assess the accuracy of these remote sensing products.

To remind Foody (2002) the state on how classical validation techniques are recently employed opens certain deficits between research and operational applications. There is a need to overcome these deficits. Considering this, the aim of this document is to provide a synthesis of the common land cover map validation approaches using reference data. The protocol defines requirements for a validation process and picks up shortcomings of the currently applied traditional approaches.

The protocol is structured in three sections:

1. The first part of this describes a basic validation protocol as result of the most common practices and which could be understood as the “standard” requirements for validation.
2. In the second chapter the major shortcoming of that “standard” validation procedures are critical analysed. This part is aimed to be a brief and propaedeutic to the third section.
3. In the third chapter possible requirements and approaches are discussed to overcome the shortcomings of the traditional validation approach.

1. Synthesis of a traditional field data collection and validation approach

In order to ensure credibility of earth observation derived products, assessing the accuracy of the classification results, i.e. land cover maps, should be considered as a mandatory step in geospatial map production. In this respect the most appropriate and doubtless approach is to validate the map result using collected ground truth data which are considered to be correct[1], and need to be upmost independent from the datasets used for map making. Furthermore these ground data should be as far as possible collected during the same period of image registration, from which a land cover use map is derived. This “temporal compliance“ is further discussed in the section 1.4. The quantitative measurement for the accuracy of the map is the level of agreement or correspondence between the classes from the map and the independent ground truth data observed in the field. Ground truth data can be collected by different means, such as ground survey or using VHR image interpretation, which we call “pseudo-truth” data. Whereas it needs to be considered that the data obtained from an image interpretation can include errors and a ground survey is always preferable over remote sensing.

In addition to proper preplanning of field surveys and clear observation rules, methodological shortcoming appear in practice through inappropriate selections of the survey type. In an assessment of ground truthing approaches in 8 of out of 10 cases surveys are executed following a ‘windscreen survey approach’ along roads (SIGMA, 2014). This method, although very fast, bears the danger of creating errors through not precisely identifying or mapping the sample units while sitting in the car. Further, the representativeness of samples only along roads is questionable in terms of distribution of the different land cover classes over the test site and well balanced sample selection.

This chapter collates a simplified overview upon the key components of traditional land cover map validations by means of ground truthing methods, while emphasizing on **six key components** of a well-defined approach. It is based on most commonly applied steps, while addressing the challenges of dedicated ground truthing initiatives under operational conditions. It is particularly addressing the need of predefining and planning of the concepts of all six steps and transparently documenting how far these plans could have been executed. These six points are the minimum mandatory variables ensuring the appropriateness and credibility of accuracy assessments of remote sensing earth observation products:

I. Legend and cartographic standard
II. Sampling unit
III. Sampling design
IV. Survey guidelines
V. Data processing and documentation
VI. Analysis of the data

[1] See Foody, 2002 for a critical overview of accuracy assessment and its implications.

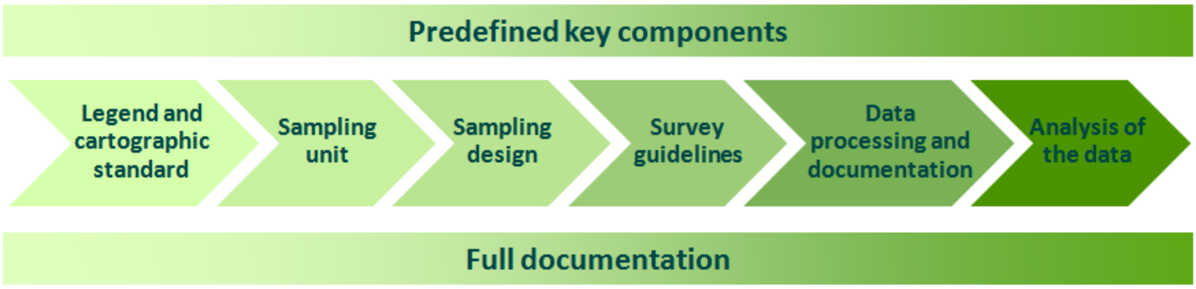


Figure 1: Key components for data collection and validation

This is being embedded into the Land Product Validation (LPV) sub-group of the CEOS Working Group on Calibration and Validation (WGCV). LPV has developed a global validation hierarchy (table 1). It is ranging from accuracy assessment with a small set of reference data to systematic, quantifiable and statistically robust assessments which are regularly updated. Although the hierarchy is intended for global validation it can be used as a benchmark for local or regional levels as well. The proposed data collection and sampling protocol shall contribute to a minimum 'stage 2 validation' but provides the principle guidelines to advance to stages 3 or 4.

Table 1: CEOS WGCV Land Product Validation Hierarchy[2]

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

Principal strategy behind the above mentioned 'six points' should be the **pre-definition** of the chosen parameters with a clear reasoning behind, which can be matched with the applied mapping methodology and the full and **transparent documentation** of the data collection at all steps, in order to being able to trace discrepancies, eventually necessary adoptions through unforeseen events or data loss. This ability is a crucial aspect to ensure credibility and acceptance of the data and to being

[2] Baret et al. 2009

able to assess potential impact on the validation results through misrepresentation of samples or classes.

The document is structured according to the principle requirements of the validation. In the next chapter the requirements for the six key components for field data collection are described in detail with selected practical examples to better visualize the relevant concepts behind and to strengthen the definition of new individual field work and validation initiatives.

1.1. Legend and cartographic standards

A map is an attempt to represent the "real world" features through a process of "generalization". This process can be done in many different ways, however two aspects are the most important, both for the map production itself and for a correct and transparent analysis of its accuracy:

- The thematic content of a map, expressed in his legend
- The cartographic standards used to represent the geographic extend of the "real world" with cartographic rules.

The legend is a powerful method where the thematic content of a map is instantiated and shared between different users. The legend reflects how the semantic "generalization" of a specific geographic area has been conceived. The real world can be considered a "continuum" with a huge granularity of different information. The process of categorization, is therefore, a process to minimize this complexity. Class creation is by its nature an arbitrary process, and therefore it can be the source of a large amount of uncertainties and vagueness that will deeply affect the whole validation process. It is important that the "map producer" will assure the minimum requirements for a correct utilization of a legend:

- Clear and unambiguous class definition
- No class semantic fuzziness. Class (semantic) boundary not overlapping with other legend classes

The cartographic standards are the information and the rules that make it possible to represent the geographic portion of the earth surface into a map. Some of them are very familiar to any map users, other which are similarly important are often neglected or not clearly stated by the map producer:

- Map coordinates is an obvious parameter present in all the modern maps, however it would be useful to document the RSM error of the satellite images used to create the map itself.
- Scale is another obvious parameter everybody knows. It is important to clarify, however, that in modern maps instead of "scale" the "Minimum Mapable Unit" (MMU) should be considered. The MMU defines "*the smallest surface of the ground represented in the map".* The MMU is, therefore, not only a much more precise parameter to represent the cartographic limits of a map, but also influence the thematic details of it.
- Linked with the concept of MMU, but generally often neglected is, how a map deals with the representation of "heterogeneous areas". Heterogeneous areas are geographic areas that at a certain scale (or image resolution) results composed by different patches of different

legend classes. Very often those situation are represented as “mixed classes”. It is a useful method to overpass the limitation of the MMU, however it must be clearly stated at least the following:

- A rough proportion between the different features forming the *mixed class*
- The minimum % a certain feature must have to be represented in a *mixed class*

In section 3 of this document the representation of heterogeneous areas will be more exhaustively discussed.

The prerequisite of any validation process is that the data collected in the field must be comparable to the data from the map, spatially and thematically. For this reason it is the responsibility of the map producer to secure a map with an adequate legend and cartographic standards. In absence of these minimum requirements the validation cannot be executed or is done with parameters artificially created by the validation team.

An example for legend and cartographic standard is described in “JECAM Guidelines for cropland and crop type definition and field data collection”[3]. It provides a hierarchical nomenclature for agricultural mapping, defines annual cropland in up to four thematic level of detail and a minimum mapping unit.

Another example is the Land Cover Classification System (LCCS)[4] or the ISO certified Land Cover Meta Language (LCML) developed by FAO. It provides a consistent and objective classification scheme with a defined set of classifiers attributes and objects to represent any land cover.

A translation of the SIGMA cropland definitions in LCML is in preparation allowing an interoperability of the LCCS and the JECAM nomenclature.

Legend and cartographic standards

Questions to be raised:

- ⇨ Is there a clearly pre-defined legend, which ensures:
  - Full thematic match between map and reference data?
  - Clear rules for survey observations?
- ⇨ Are the cartographic standards of the map clearly defined:
  - Map coordinates and the parameters of the satellite input data?
  - Scale and MMU of the map?
  - Are the rules and definitions for mixed classes clearly defined?

Check box 1

1.2. Sampling units

The second key component for data collection is the definition of a sampling unit. The sampling unit is the area unit which is observed in the field. It can be a field or land pattern (in the sense of a polygon), an artificially defined segment (in the sense of squares, rectangles or from image pre-processing such as eCognition or alike), transects, points or others. The size and shape of the

[3] JECAM Guidelines for cropland and crop type definition and field data collection (version 1 2014-10-24)

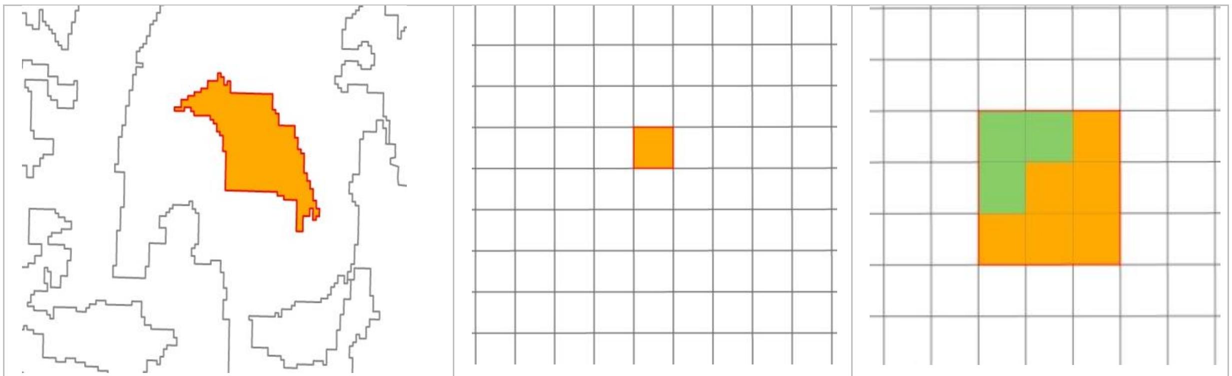
[4] Di Gregorio 2005

sampling unit has to be defined according to the type and resolution of the map as well as the input remote sensing data.

The sampling unit is the reference unit to link the spatial location on the ground with the corresponding location in the map. The shape and size of the sampling unit has to be defined in a way that the observation in the field can be spatially linked and compared to the ground resolution of the corresponding spatial unit in the land cover map which are either classified pixels or polygons. There are different examples for sampling units with advantages and disadvantages.[5]

A straightforward approach is to use the map's units as sampling units (pixels or polygons). Practically a sample of the map units is selected from the map, allocated and observed in the field. This requires that the map is already available for the survey planning and that the map units can be precisely located in the field. For example pixels or group of pixels might be located in the field by their corner or centre coordinates using GPS device, whereas the positional accuracy of GPS and map has to be considered (see 1.4).Figure 2 shows examples of map units as sample units.

An option to sample polygons is to (automatically) generate observation points or areas inside the selected polygons. The coordinates of the points or the corner coordinates are used to allocate the observation areas inside the selected polygon. This technique can be adapted according to the size of the polygon or the land cover class, for example assigning more than one observation area to large polygons. The sample unit is still the polygon, but the use of multiple observation areas allows to observe polygons which are too big to be observed with a single observation point (see chapter 1.6.1).



Polygon (e.g, from segmentation)	Single pixel	Group of pixels

Figure 2: Examples for sampling units from the map units.

Another approach is to use sample units which are independent from the map units for example transects, points, square segments or segments with natural boundaries as sampling units. This in general includes setting up a sampling frame from which the units are selected (see 1.3.2). Figure 3 shows some examples of map independent sample units.

[5] See also Stehman and Czaplewski 1998

Sample units can also be used in a two stage sampling approach for example points are used to select land cover plots which are then delineated and observed as segments[6]. Sampling units can also be adjusted according to the land cover class encountered at the unit. For example the observation radius around a sample unit can be extended if it falls in a certain land cover. This allows some flexibility for example to properly capture heterogeneous land cover classes, like open forest classes.

Figure 3: Examples for sampling units independent from the map (segments, polygons, points).

The choice of the sample unit should be guided by the following considerations:

- The map resolution and size and shape of map units: The sample units should be spatially comparable to the map unit considering the minimum mapping unit (MMU) and/or the pixel resolution of the remote sensing input data in order to get a pure reflectance of one sample per pixel.
- Land cover characteristics of the test site: The sample unit should allow to adequately capturing the land cover (e.g. an open forest consists of patches of trees and grassland; the sampling unit should be large enough to observe the land cover as open forest).
- On the practical side it shall be easy to set up, handle and process the sampling units. For the surveyor it shall be easy to locate and observe the sample unit in the field. For logistical aspects one surveyor should be able to complete minimum one sampling unit within one working day.

A crucial aspect is if the land cover map to be validated is already available for the survey planning, or more precise, if the map units are already available, e.g. from early segmentation of the satellite data. For the validation of maps where seasonal land cover information (e.g. crops maps) is collected, the field data collection has to take place at an adequate date during the season in order to well represent the target land cover. In most cases this is the same period where the remote sensing data was captured which is used to generate the map. In that case it is unlikely that the map is already available and that the map units (e.g. land cover polygons) can be used as sample units. This needs to be considered in the planning of the ground truthing process and alternative, map independent, units have to be used. This problem is picked up again in the following chapters.
Whatever sampling unit is used for the validation they have to be precisely defined with a spatial position (geographic coordinates) and a unique ID. The entirety of the sampling units must represent the map area (population). This means that the sampling units must either cover the entire area of

[6] Kerdiles et al. 2014

the map directly or a sampling frame is used which represents the entire area. This ensures that the selected sample is representative for the entire map.

Sampling units
Questions to be raised:

- ⇨ Is there a clear definition of the sampling unit which adequately suites the map input data?
- ⇨ Is this applicable for a field data collection in terms of costs, accessibility feasibility in terms of work load?
- ⇨ Is it open to also suite other input data in terms of data gaps?

Check box 2

1.3. Sampling design

The third key component for field data collection is the **sampling design**. It describes the protocol by which the sampling units are selected for field observation. The aim of the sampling design is to define a valid and transparent selection process which produces a representative sample of the map and efficiently allocates available resources.

Selection of the best sampling design depends on many factors and no specific method can be recommended without knowing the details of the map to be validated, the landscape characteristics, the aim and background of the validation approach and the available resources for the ground truthing. In this part only some examples shall be given and the common requirements for defining a sampling design are described. Shortcomings of traditional approaches and specific methods shall be picked up in chapter 2 and 3 of this document.

In general there are two different types of sampling designs, **probability sampling** and **non-probability sampling**.

Probability sampling provides that each sampling unit has a known chance of being selected from the population and that all units have a non-zero chance of being selected. Non probability sampling includes approaches where the chance that a sampling unit is selected is unknown and can be zero.

Examples of non-probability sampling are generally used for collecting training data for the map classification. An example is '*along the road survey'* where the surveyor observes "suitable" samples of land cover and their position along the road. This data can be suitable for training, provided the exact position of the sample is ensured, but would not provide the quantifiable inclusion probability to generalize from the sample to the entire map. The recommended sampling design for a valid accuracy assessment is a probability sampling.[7] The calculable selection probabilities of each sample provide the statistical foundation to estimate the accuracy for the entire map. It also allows further

[7] Stehman and Czaplewski 1998

analysis and application of the field data[8]. There are several probability sampling designs with disadvantages and advantages.[9]

Practically, the map units, for example polygons, have to be selected following an objective and defined probability sampling approach. This could be for example a stratified random selection process where map polygons are selected randomly from each land cover class. Other approaches are for example systematic sampling which uses a sampling frame to select the samples.

For the selection process two cases can be considered. (i) The map to be validated is ready and the map polygons or classified pixels can be directly selected for ground truthing as described above. (ii) The map is not yet available and therefore alternative sampling units have to be defined. In this case the setup of a sampling frame is recommended. The sampling frame represents the entire population (map) through defined spatial units and boundaries.[10]

The sample frame is defined by the size of the sample units and possibly the distance between individual sample units. It shall adequately consider the mapping units, the mapping scale and the specific characteristics of the test site (see 1.2). Examples of sampling frames are regular grids with points or square cells as sample units. Other approaches use segments with physical boundaries derived for example from a segmentation of satellite imagery. Figure 4 shows examples for different sample frames.

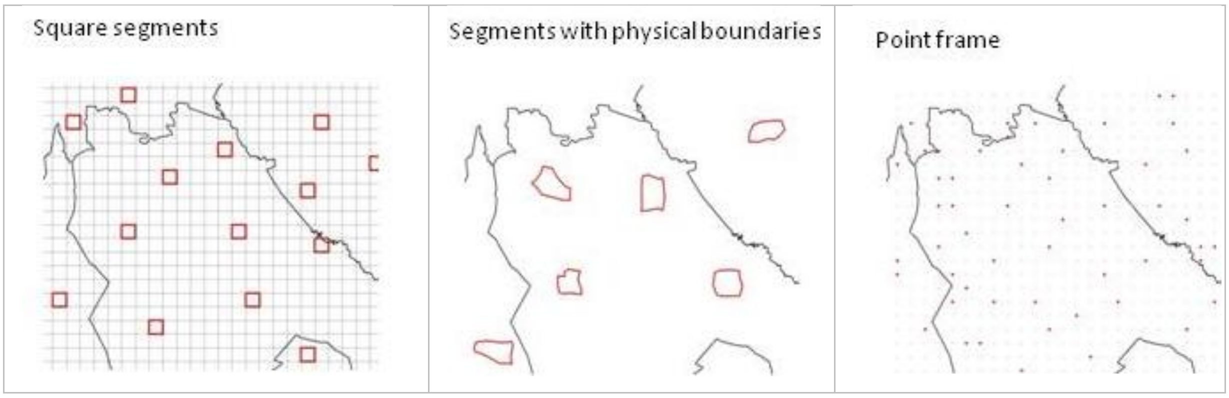


Figure 4: Examples for sampling frames with different units

The advantage of a sampling frame with a regular grid of sample units is that it ensures the selection of a representative and spatially well distributed sample with a transparent selection process. This ensures understanding and acceptance of mapusers but might not be the most cost efficient way. Further it is very easy to set up and clearly independent from a specific land cover map and its map units and extent. This does not limit the usage of the collected data for one purpose and one map alone.

[8] Gallego et al. 2010
[9] Congalton and Green 2009, Stehman and Czaplewski 1998
[10] Stehman and Czaplewski 1998

If a new regular grid is set up the exact starting point of the frame has to be chosen randomly. It is further recommended that the frame does not only cover the test site but the entire administrative unit or country to allow future extension.

1.3.1. Stratification

Stratification is commonly used in probability designs to increase the efficiency of the sampling and improve the allocation of the samples. Stratification uses prior information of the population (map or map units) to divide it into different strata and select samples from each stratum separately. Stratification can be based on land cover, agro ecological zones, complexity of the map, altitude, inclination, administrative units, or any attributes defining areas where separate sampling rules shall apply. Stratification can also use the properties of the sample units, e.g. size or heterogeneity of the map polygons. Important is that each sampling unit is assigned to only a single stratum and that the sampling probability for each sample is calculable.[11]

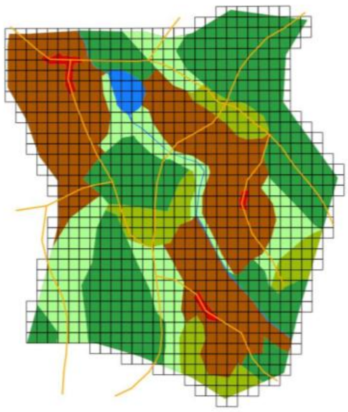


Figure 5: Example of a Land Cover map with overlaid square grid frame.

Stratifying according to the land cover classes of the map to be validated allows controlling how many samples per map class are selected and ensures that all classes are well represented in the analysis [12]. In case if the land cover classes of the map to be validated is not yet available it is recommended to use other available ancillary data, preferably land cover data, which can be used for stratification (see figure 5).

In areas with difficult terrain or poor road network access to the sampling units can become very expensive and time consuming. Stratifying according to zones of accessibility (e.g. distance to roads) and selecting a higher proportion of samples from the "easy to access" zones is a compromise to reduce travel expenses but keeping the probability sampling approach (see figure 6).[13]

If appropriate knowledge or data is available it is recommended to exclude areas which are not accessible for data collection. This can be no access areas like military areas, national parks, steep or mountainous areas etc. The samples falling in these areas are excluded from the ground observation, accordingly the accuracy assessment does not cover this areas and no statement on the map accuracy for this areas is possible.

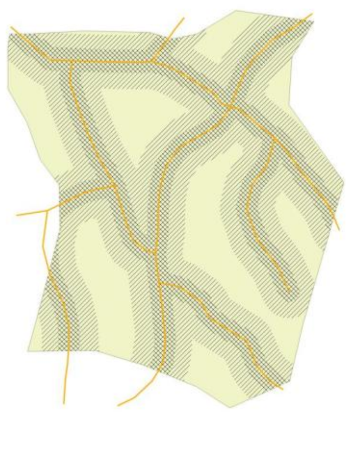


Figure 6: Example for zones of accessibility (distance to road). Each sample is assigned the accessibility zone it falls inside

[11] See also Strahler et al. 2006
[12] See also Olofsson et al. 2012
[13] Stehman and Czaplewski 1998

Alternatively the samples in those areas could be photo interpreted using for example recent VHR imagery, provided that the land cover information can be doubtless identified.

A further optimisation of the sampling can be the use of pseudo-truth for classes which can be doubtless identified using for example VHR imagery. This depends on the land cover classes to be assessed and the available input data. More systematic rules are yet to be defined. The optimisation should reduce the cost and time required for data collection in the field but should not provide an excuse to avoid field survey.

Practically the layer with the sample units is overlaid with the spatial information used for stratification (e.g. distance layer) and each sampling unit is assigned to the strata it falls inside or where most of it falls inside. From each stratum the samples to be observed are selected e.g. randomly or systematically and based on the desired sample size per stratum. These steps can be easily done in a GIS environment.

If more than one attribute shall be used to create a strata e.g. the analysis of the data becomes more complex because the number of strata increases as a combination of the attributes and the different inclusion probabilities need to be considered for the analysis. For example to account for the different sizes of land cover polygons an approach is to stratify the polygons by land cover class and by size of the polygon[14].

Whatever stratification is applied it is important that the proportion of each stratum from the population (map area or entirety of sampling units) is known so that the sample probabilities can be calculated for each sampling unit in each stratum.

1.3.2. Sample selection

One of the critical questions for the sampling design is how many samples are required in total and how many samples to allocate to each stratum (map class) to get a satisfactory estimation of the map's accuracy?

In general the more samples are collected the higher is the reliability of the estimated map accuracy result, but at a certain number of samples the level of reliability which is gained by additional samples is negligible. Field survey is expensive therefore it is desired to find the optimal number of samples as a compromise between the required level of reliability and available resources for ground truthing.

Often the number of samples proposed for validation derives from non statistical analysis or empirical values or just experience. For example the approach proposed by Congalton[15] who defines a number of samples per map class or strata according to the number of classes and the size of the test site. This "rule of thumb" shall give an orientation to estimate the appropriate number of total required samples.

[14] Carfagna & Marzialetti 2009

[15] Congalton and Green 2009 p. 75 and Congalton 1991

Table 2: Sample sizes per class according to Congalton[16]

Land cover map classes and area of test site	Samples per class
< 12 classes and < 1 million acres (≈ 4000 km²)	50
> 12 classes or > 1 million acres (≈ 4000 km²)	75-100

Following this example a land cover map with 8 classes and below 4000 km² should be sampled with around 8x50 = 400 samples.

The number of samples which are selected from each stratum can be the same for each stratum or based on different criteria, e.g. proportional to the number of samples per strata or proportional to area of a stratum. Other criteria can be based on specific test site conditions and project requirements.

A scientifically sounded and more advanced procedure to estimate the optimal number of samples should include clearly defined accuracy targets for the validation process expressed through levels of confidence. An example is the use of the multinomial distribution which requires some estimated and known map parameters. [17] In any case the advanced procedures for sample size calculation should involve the help of a statistician.

Examples how to calculate the appropriate sample size is provided in Congalton and Green 2009, Cochran 1977 and Olofson et al. 2014 who refers to Cochran and provides some examples how to allocate the samples per class.

Whatever the selection process, for a valid and scientifically defendable accuracy assessment it is important that it follows a concrete and transparent protocol considering the following fundamental elements[18]:

i. **Predefined sample selection:** The selection of sampling units for validation should always be done before the field survey and any subjective choice in the selection process must be avoided.
ii. **Probability sampling**: The probability of being selected should be calculable for each sampling unit and should be non-zero.
iii. **Good spatial distribution of the samples:** This is provided for example by a regular sampling frame.
iv. **Appropriate sample size:** Enough samples per land cover class for valid assessment of class accuracies.
v. **Samples covering all land cover classes of the map**: Stratification according to land cover class supports that samples can be selected from all land cover classes.
vi. **Well representation of the different types, dates and categories of the input data.**

[16] Congalton and Green 2009 p. 75

[18] Congalton and Green 2009, Foody 2002, Olofsson et al. 2014, Stehman and Czaplewski 1998, Strahler et al. 2006

The final sampling units are identified by their geographic coordinates and their unique sample ID. The samples should be processed in an appropriate GIS file and as GPS waypoints to be uploaded to a GPS device.

Sampling design
Questions to be raised:

- ⇨ Is there a clear predefined sampling design, which
    - o Ensures representative and sufficient allocation of samples?
    - o Allows to verify the representativeness of the validation results?
    - o Documents a sound validation planning?
    - o Helps guiding the surveyors itinerary planning?
    - o Ensures sound extrapolation and calculation of significant validation results?

Check box 3

1.4. Survey guidelines

The fourth key component for data collection is survey guidelines. The survey guidelines define what data is collected in the field and how this data has to be collected in a proper way. Rules and instructions are required for the surveyors to ensure the consistency and quality of the collected data and allow a smooth and error free integration into automated GIS routines for analysis and further processing.

The ground truth or reference data collection is the fundamental step of the validation. The accuracy assessment is based on the assumption that the collected ground truth data is correct and a representative but independent sample of the land cover map. Any failure or mistake during data collection will produce misleading results. The methodology for data collection has to fulfil among others the following requirements[19]:

i. **Positional correctness:** The sample must be precisely located on the ground and on the map. Even with GPS devices and high resolution imagery correct co-registration between the sample in the field and the corresponding pixels or polygon in the map can be challenging. The accuracy of a GPS device might be in the range of 5-10m and the accuracy of a RS image in the range of 0.5-1 pixel which can be 15-30 meters for 30 meter resolution. Using a GPS device and a high resolution field map showing the position of the sample relative to its surroundings should be the standard equipment to locate a sample in the field.

ii. **Thematic compliance:** The land cover class parameters assigned to the sampling units in the field and in the map must be comparable. In general this is guaranteed if the same classification scheme is used for the data collection and for the map. For specific cases it could be necessary to observe different object types as class parameters e. g. to split classes into subclasses such as summer crop and winter crop = cropland. If different schemes are used the classes have to be associated (see chapter 1.6.1). In any case it is recommended to define the land cover classification scheme in a specific survey nomenclature to be used by the surveyors in the field.

[19] Congalton and Green 2009, see also Büttner et al. 2006

iii. **Consistency:** The data collection itself has to be consistent, for each sampling unit the same observation procedures have to be applied. This is supported by the use of standardised field forms, survey /observation instructions, training and digital data processing interface.

iv. **Temporal compliance:** Timing of the survey is crucial especially for agricultural crops mapping. Ground truth data should be collected in the same growing season as the corresponding land cover map's RS input data. The survey itself should be according to the crop calendar to allow crop identification in the field.

The data collection has to be transparent and consistent and well documented to avoid any manipulation or the impression that the assessment was inconsistent. To fulfil the above stated requirements the following aspects of data collection are important issues:

- Survey variables: Definition of the variables to be observed in the field.
- Observation guidelines: Rules and instructions how the observation of the samples has to be executed
- Field work execution: Description of the data collection process and its components
- Digital data entry: Guidelines and requirements for entering the observed data into a digital database for analysis

1.4.1. Survey variables

Precondition for data collection is that the data collected in the field can be linked to the sample in a database. This can be done in a geodatabase using the unique ID of the sampling unit to establish a link between the different data types (photos, GPS waypoint, and observation data).

The variables to be collected at the sample have to be adapted to the specific requirements of the survey and local land cover characteristics of the test site. Apart from the land cover observation the survey has to include variables which document the circumstances of the survey execution itself. The specific description of the variables and how to observe them shall be documented in the survey guidelines. For the data collection it is recommended to use a standardized field form. In general the following variables are recommended:

- General sample and observation parameters
- Land cover parameters at the sample
- Geographical position of the point of observation (e.g. GPS waypoint)
- Digital photos to support the documentation of the land cover at the sample

General sample and observation parameters: The observation of each sample has to be described with some parameters documenting the execution of the observation itself. The ID of the sample, the name/number of the surveyor, survey time and date are mandatory variables. The position of the surveyor relative to the sample should be documented in case the sample cannot be reached or only observed from a certain distance e.g. when a sample falls inside a field with high crops or inside a fenced area. In general any irregularity in the observation procedure has to be documented including: Problems in the exact location of the sample (lack of adequate landmarks for orientation, loss of the GPS signal etc.); reason for not reaching the sample; problems in the coding of land cover (e.g. crop recognition, problems with nomenclature, etc).

Land cover parameters: The land cover variables have to be in compliance with the maps classification scheme whereas the thematic detail of the field data should always be higher than the map's detail. The land cover variables have to be adapted according to the map's classification scheme, desired level of thematic detail and the regional land cover specifics. The definition of the land cover variables and how to observe them in the field shall be specified in the survey nomenclature.

GPS waypoint: For each sample a GPS waypoint has to be taken at the position from where the observation takes place. It provides a proof that the surveyor has reached the sampling unit, supports clarification in case of doubts about the survey and allows calculating the distance between the position of observation and the sample. To link the observed data to the GPS waypoint the number / ID of the waypoint shall be recorded along with the other observations e.g. in the field form.

Photos: Taking photos at the sample documents the sample's position relative to its surroundings. The photos document the land cover at the time of the survey and can illustrate the character of a landscape and its special features better than words. Changes can be visually documented in case the photos are taken from the same position in different years.

It is mandatory that the photos can be linked with the sample observation. A possible approach is to link the photos through the unique sample ID. For example recording the filename of the photos in the field form of the sample or renaming the photo with a name that includes the sample ID of the corresponding sampling unit. Another option is the documentation of the precise spatial position and view direction of the photo to link it with the corresponding land cover unit. Apps for data entry in the field and/or taking photos can be applied if appropriate[20] .

To ensure consistent quality of the photos some technical and layout requirements should be defined for example[21]:

- minimum resolution
- landscape format
- focal length at wide angle (no zoom)
- no use of flashlight
- avoid persons and cars (number plates) on the photos
- horizon shall be at around 2/3 of the photo (2/3 ground, 1/3 sky)

In the example below photos were taken at sample points. Photos of the point, land cover and landscape photos in each cardinal direction were made. The surveyors recorded the filename of the photo in the field form of the corresponding sample point.

[20] See for example "Geo-wiki pictures" by IIASA (www.geo.wiki.org/mobile-apps/)

[21] The guidelines for the photos are adopted from the European LUCAS survey.

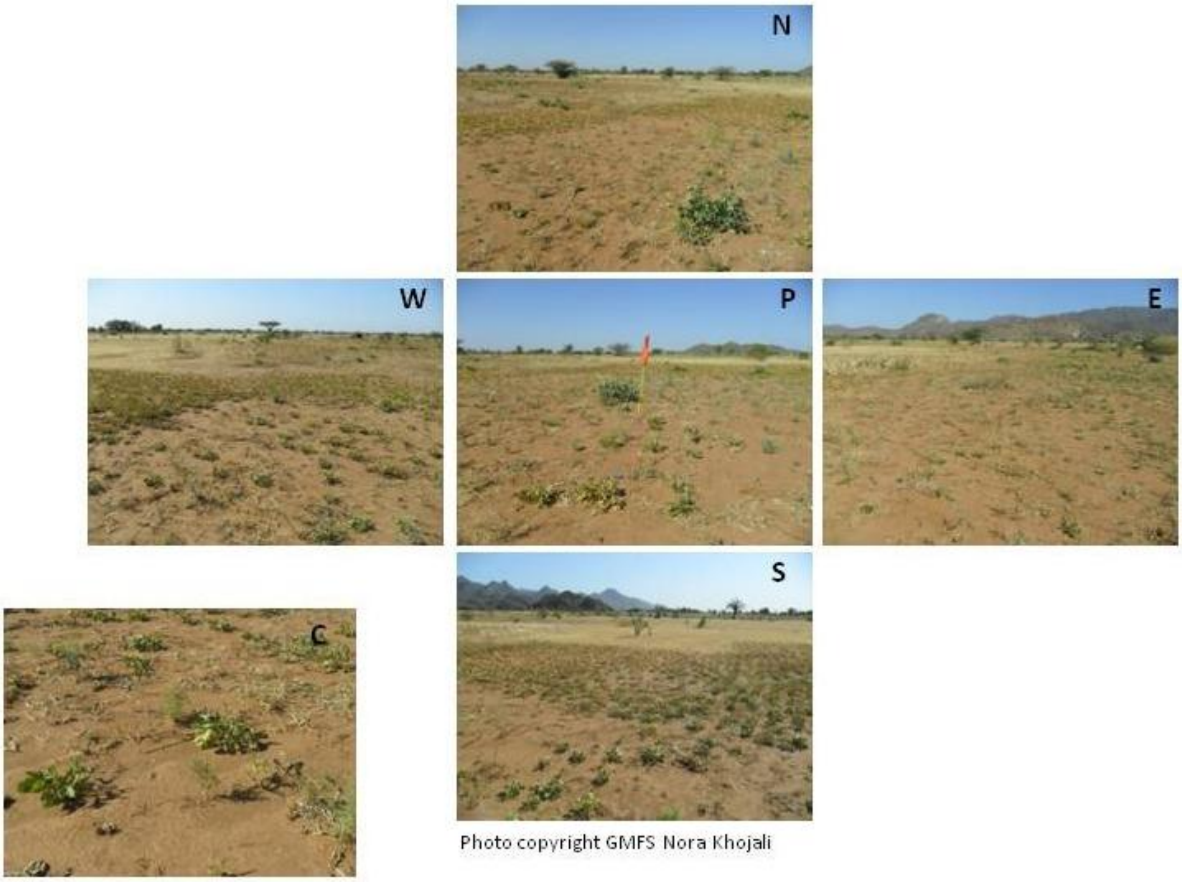


Figure 7: Example for point (P), landscape (N, E, S, W) and crop (C) photos taken at a validation sample point in Sudan.

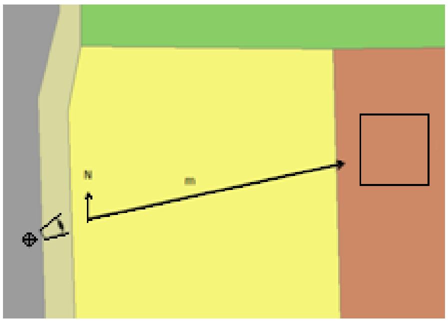
1.4.2. Observation rules

Observation guidelines shall define the rules and instructions how to execute the survey and how to handle special circumstances in the field.

As discussed in chapter 1.2, the sampling units shall be predefined and unambiguous identifiable in the field by their geographic position through GPS coordinates. The surveyors shall be additionally equipped with field maps showing the position of the sampling unit. The surveyor has to reach the sampling unit or observation area or get as close as possible to its position. To locate the sampling unit in the field the GPS and the field map should both confirm the position.

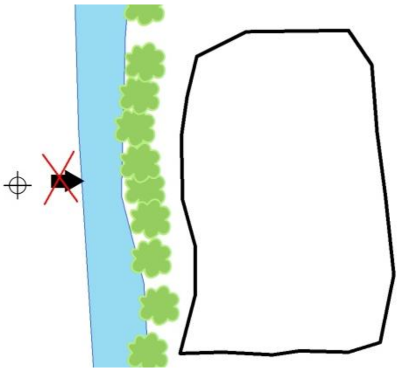
In some cases samples are not accessible due to several reasons: restricted area, private property, fenced area, impassable river, wetland, thick forest, dense crops, security issues, blocked roads etc. or it is not possible to precisely locate the sampling unit because it falls on a border of different land covers or other difficulties. For those cases clear guidelines need to be defined to have a clear and transparent documentation of the survey execution. In any case parameters such as the position of the surveyor to the sampling unit should be documented for each sample (see 1.4.1.).

Some examples for special cases are provided below:

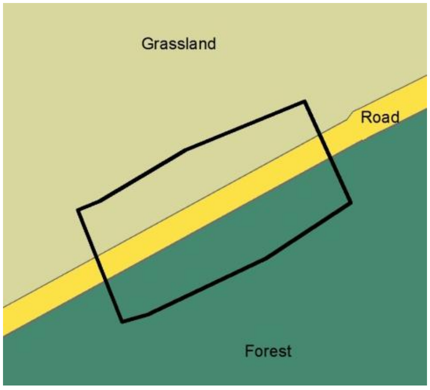


The sampling unit is observed from a distance. The relative position of the surveyor to the sample has to be recorded. E.g. with a GPS waypoint, the distance to the sample and the direction of observation. The same applies to photos to avoid misregistration. It might also help to define a maximum allowable observation distance.

The sample cannot be reached. The surveyor cannot reach the sampling unit and cannot observe it. The reason for not observing the sample (river), the position of the surveyor and distance to the sample should be documented.



The sample unit cannot be precisely allocated. The sample falls on a border between two land cover or on a linear element (road). Rules need to be defined to clarify this cases (e.g. shifting the sample towards the northern plot).



1.4.3. Field work execution

The recommended method for ground truth data collection is by field survey with surveyors working in teams or individuals. Practically the proposed survey methodology itself consists of three tasks:

1. **Daily preparation of the survey**:
    - Route planning, material preparation
2. **Field work:**
    - Approach to the sampling units using maps and GPS device
    - Sample observation and data collection at the sampling unit
        - Land cover observation

  - Taking GPS waypoint
  - Taking digital photos

3. **Data entry and backups:**
   - Digital entry of the data
   - Quality control and backup

For consistent data collection some minimum equipment should be considered for the survey.

For example:

- Maps for route planning and navigation to the sampling units
- Field Maps of the sampling units on basis of recent HR / VHR satellite image showing the allocation of the sampling units
- Handheld GPS device to navigate to the sampling unit and take GPS waypoints
- Blank field forms or any device to enter the field observations in a structured way
- Digital camera
- Notebook/device and software for data entry

Survey guidelines
Questions to be raised:

⇨ Are there clear observation rules and a defined field form and nomenclature which allow to:
   - Collect precise and complete field observations?
   - Transparently document the particular observation at the sample unit's level?

⇨ Are the logistical aspects of the survey execution considered:
   - Transport to reach the sample units?
   - Equipment to perform the observation?

Check box 3

1.5. Data processing and documentation

The fifth key component for data collection is the data processing and documentation. It encompasses how the collected data is entered into a digital database, the quality control of the data and the documentation of these steps. The entry of the field data is based on the applied legend and survey guidelines and provides the automated processing of the data with GIS routines (see 1.6.). Each step shall be documented including any deviations from the protocol.

Analysis and management of the field data requires that the data can be used to:

- spatially analyse the field data and compare it to the land cover map
- link photos with the observation data for LC verification
- derive descriptive statistics and basic analyses

Common data formats are GIS shapefile or geodatabase for management and spatial analysis of the data and linking of the photos. MS access provides a good structure for non-spatial data analysis and data entry. Excel spreadsheet is sufficient to build the confusion matrix and calculate the accuracy parameters. A convenient way to share and display field data and attached photos is by means of Google earth with the formats kml and kmz.

Data entry should be as soon as possible after the observation to clarify any errors immediately while the memory of the surveyor is still fresh. Data entry can happen either directly in the field through a mobile device and appropriate software or recording the observation in the field in standardised field forms and entering the data after the survey.

It is recommended to use a software or application that enables data entering through customised data entry masks to minimize errors during data entry, especially typing errors. The software should be able to perform some defined plausibility checks.

The collected field data is considered to provide the true land cover at the sampling unit, any errors in the dataset will have a huge impact in the accuracy assessment. It is therefore of uttermost importance that errors in the reference dataset are minimised.

Each step of data processing shall be well documented.

It is important that any doubtful observations are excluded from the data base to have an unbiased ground truth dataset for analysis.

Consistency in the data collection is improved through personal training of the surveyors before the survey and application of strict and objective observation procedures.

Data processing and documentation

Questions to be raised:

- ⇨ Is there a dedicated digital data entry and digitization interface which allows to completely collate the collected field results in a digital format for further processing and back up?
- ⇨ Is there a clear data flow mechanism in order to ensure completeness of the data transmission?
- ⇨ Is it ensured to explore data loss at any processing level?

Check box 5

1.6. Analysis of the data

The sixth key component is the analysis of the data. The collected data is compared with the map data in a possibly automated way. The analysis shall be, considering component 1.5, repeatable and comprehensible.

In the analysis the land cover information observed from the surveyor, which is considered to be accurate (true), is compared to the land cover of the corresponding map location. The level of agreement between both data is the measurement of the accuracy of the land cover map. The most applied way is to analyse the data using a confusion matrix (also called error matrix), a contingency table which represents the per class confusion and agreement between the map classification data and the reference ground truth data. The confusion matrix is the central element of the analysis and allows deriving a number of overall and per class accuracy parameters[22]. All previously defined 5 components are defined in order to ensure sound match of the independent reference data versus the map result. Any constraint at an earlier stage leads to possible problems here.

Precondition for the analysis with a traditional confusion matrix is that each map unit and each sampling unit is assigned to a single class and that they represent the same spatial extend and are perfectly registered.[23] This may be more or less difficult to achieve depending on multiple factors like the landscape structure, pixel resolution, minimum mapping unit, sampling unit and positional accuracy of the map and the field sample. For example in a homogenous landscape with huge land cover units and predominantly "pure" pixels it is unlikely that positional errors e.g. some meters shift between the sample position in the field and on the map, or problems with classification will have an impact on the accuracy assessment. But they may have an impact in areas with a heterogeneous fragmented landscape where mixed pixels are common and where a slight change in the spatial position will shift the sample to a different land cover class.

These implications are initially considered in the sampling design especially in the selection of an appropriate sampling unit (e.g. observation radius) and classification nomenclature. In case land cover has been classified differently in the map than in the field e.g. different level of details have been used, a protocol how to associate the land cover classes should be defined as described below.

Different approaches how to spatially register the sampling units to map units are also presented below.

1.6.1. Preparation of the data

In general the thematic compliance between map and field data is provided by using the same classification scheme for field observation and map data classification. Nevertheless it is useful to define a protocol and document, e.g. in an association table, how different cases are handled. For example:

- The field data is observed with a higher level of thematic detail (as recommended) than the map. The field data has to be collapsed to the maps level of detail.
- More than one land cover class has been observed at the sampling unit. The dominant land cover needs to be defined (e.g. biggest coverage). Or the sampling unit is excluded.
- Mixed land cover, e.g.: intercropping of more than one crop type, vertical layers of different crops, forest cultivation (trees and crops), ...

The rules how to associate the sample and map land cover classes should be included in the accuracy report e.g. as a table24. An example of an association between map class and observed land cover with different level of detail is given below.

[22] Congalton and Green 2009, Congalton 1991, Foody 2002, Olofson et al. 2014, Strahler et al. 2006
[23] Strahler et al. 2006
[24] GMFS3 C05 Validation Protocol, Büttner et al., 2006

Table 3: Example for a map association table

Map land cover class	Field data land cover
AG - Agriculture in terrestrial and aquatic/regularly flooded land	Main crop – maize
	Main Crop – sugarcane
	Main Crop - millet
TCO - Trees closed-to-sparse in terrestrial and aquatic/regularly flooded land	Main layer trees
SCO - Shrubs closed-to-sparse in terrestrial and aquatic/regularly flooded land	Main layer shrubs
HCO - Herbaceous closed-to-sparse in terrestrial and aquatic/regularly flooded land	Main layer herbaceous
BS - Bare rocks and soil and/or other unconsolidated material(s)	Bare rock
	Bare soil
	Dunes
URB - Urban areas	Built up

Spatial registration of the sampling unit with the map unit

Apart from the thematic registration it needs to be defined how to spatially register the sampling units to the map units. In case the map units have also used as sample units in the field they can be directly compared. In case map independent units have been used a method how to register them has to be defined. Whatever the sampling unit it should be documented how the registration between the sample and the map unit is performed.

As an example three methods with point and polygon sampling units are described.

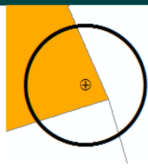
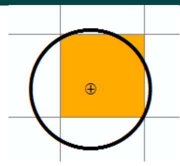
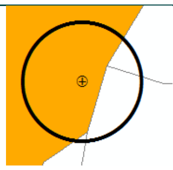
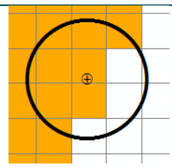
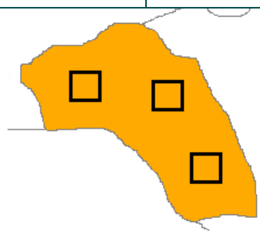
The sample points consist of two spatial units, the point defined by the coordinate and an extended observation area defined by the radius around the point. For the accuracy assessment both information can be applied. Two possible approaches are described here (see table 4):

- Comparing the sample point to the map unit it falls inside. This is a straightforward and easy to apply approach, but it requires that each sample point can be doubtless assigned to precisely the corresponding map unit. For heterogeneous areas or for maps with small map units (pixel sizes) this approach might lead to mis-registration through positional errors. It is more appropriate for homogenous landscapes where positional errors will not result in the shift of the sample to a different land cover class.
- The main land cover from the extended observation radius of the sample is compared to the dominant land cover class of the corresponding map area. Practically the digital land cover map is overlaid with the sample points in a GIS environment and a buffer with the same size as the observation radius is applied around the points. The dominant land cover falling inside the buffer is calculated and compared to the dominant land cover from the field data.

An alternative approach could be to reject mixed samples for the assessment and use only “pure” samples where only one land cover falls inside the radius.

The third example is a case were the map consists of polygons which are used as sampling units. To avoid that the entire polygon has to be observed in the field, a number of small observation areas are generated inside the polygon. Those observation areas are observed and the results are compared with the land cover from the map polygon.

Table 4: Methods for spatial correlation of field samples with map units

Method	Field sample	Map classification	Polygon	Pixel
Point to map unit	Land cover at the point (1.5m radius)	Land cover of the pixel / polygon the point falls inside		
Dominant class	Main Land cover in the extended radius around the point (20m)	Dominant land cover (coverage) inside a radius around the point (20m)		
Polygon sub sampling	Land cover in the observation areas inside the selected polygon.	Land cover of the selected map polygon		

Whatever approach is used it should be clearly and transparent documented in the accuracy reporting.

After the preparation of the data each sample shall have a land cover class assigned from the field and from the map. Those are compared in a confusion matrix.

1.6.2. The confusion matrix and accuracy parameters

A confusion matrix is a contingency table between the land cover class observed in the field (ground truth data) and the land cover class of the map at the sample sites. The columns represent the land cover classes observed in the field and the rows the corresponding values from the map. The diagonal entries represent the correct classifications and the off-diagonal values the misclassification.

Table 5: Confusion matrix for a classification with k classes[25]

	Ground truth data				
Map classification	**a**	**b**	…	**k**	Classification total
a	naa	nab	…	nak	na+
b	nba	nbb	…	nbk	nb+
⋮	⋮	⋮	…	⋮	⋮
k	nka	nkb	…	nkk	nk+
Reference total	n+a	n+b	…	n+k	

[25] Stehman and Czaplewski 1998 (modified)

The confusion matrix provides the key information of the correspondence between map and field data and allows estimating accuracy parameters:

Overall accuracy: Proportion of correctly classified land cover. Sum of the correct classifications (diagonal element) divided by the number of samples n,

$$OA=n_{aa}+n_{bb}...+n_{kk} / n$$

User´s accuracy: Per class probability that a sample unit in the map belongs to the same class on the ground. The diagonal element divided by the classification total, e.g. for class a:

$$U_k=n_{aa}/n_{a+}$$

Error of commission: The complement to user´s accuracy, providing the per class probability that a unit in the map does not belong to the same class on the ground. The sum of the off diagonal cells divided by the classification total or *1- user's accuracy*, e.g. for class a: 1-(n_{aa}/n_{a+})

Producer´s accuracy: Per class probability that a unit on the ground is classified in the same class in the map. The diagonal element divided by the reference total, e.g. for class a:

$$P_k=n_{aa}/n_{+a}$$

Error of omission: The complement to producer´s accuracy, providing the per class probability that a sample on the ground is classified into a different class in the map. The sum of the off diagonal cells divided by the reference total or *1- producer's accuracy*, e.g. for class a: $1-(n_{aa}/n_{+a})$

Kappa coefficient is another overall accuracy parameter. It compensates for the chance agreement between classes, or in other word the chance that samples are mapped correctly by pure chance. It is calculated by subtracting the chance agreement from the overall accuracy. The result ranges between 0 for random agreement and 1 for perfect accuracy[26]

The accuracy measures should be accompanied by standard errors to quantify the variability of the estimate. The standard errors are the square root of the variance. For the overall accuracy the variance is estimated by[27]:

$$v(OA)=\sum \text{area proportion of class } k^2*(U_k)*(1-U_k)/(n_{k+}-1)$$

The standard error is the square root of the variance: $STD(OA)=\sqrt{v(OA)}$

The provided accuracy estimators are valid for a sampling design with equal selection probabilities for all samples, e.g. simple random sampling or a sampling proportional to the land cover strata. To compensate for different inclusion probabilities of a stratified sampling design the confusion matrix values should be expressed in proportions rather than in counts and incorporate the different inclusion probabilities. [28]

For sampling designs using equal sized sampling units (e.g. points or squares) where each sampling unit represents the same surface area, the calculation is as follows.

[26] The use of the Kappa coefficient as a standard measurement for map accuracy is discouraged by Olofson et al. 2014 and Strahler et al. 2006

[27] adopted from Olofson et al. 2014, the article also provides the formulas to calculate the variance of user's and producer's accuracy

[28] Olofson et al. 2014, Stehman and Czaplewski 1998, Strahler et al. 2006

For each cell count *n* the proportion *p* is calculated using the per class proportion and weighing with the weight of the strata. E.g. for the cell n_{aa}:

$$p_{aa} = n_{aa}/n_{a+} * \textit{proportion of stratum a from the total map}$$

An example how to calculate the accuracy measures for a fictive land cover map is described below.

Another accuracy measure are confidence level and confidence interval. They provide information about the precision and accuracy of the validation process itself, in particular if the number of observed samples was enough to provide a reliable estimate of the map accuracy. The confidence interval, also called margin of error, describes the possible range of the result (e.g. +/- 5%). The confidence level describes how certain you can be about the result e.g. 95%. Both values are directly linked, the higher the confidence level the larger the confidence interval. The confidence level should be defined before the validation process to estimate how many samples are required for a certain confidence level, e.g. 95%, and a corresponding confidence interval, of e.g. +/- 5% (see chapter 1.3.2). After the survey the same calculation can be made again to estimate the confidence interval and level which was actually reached with the number of actual observed samples.

Analysis of the data

Questions to be raised:

- ⇨ Is there a clear analysis functionality which is defined on basis of the previously defined parameters?
- ⇨ Are the results and calculation steps transparently documented?
- ⇨ Is it possible to provide access to the both data, field data and mapping product in order to allow independent assessments?

Check box 6

Example calculation

The example below describes the accuracy assessment for a fictive raster land cover map with 7 classes. The land cover proportions of the land cover map are provided in table 6. It is assumed that field data has been collected using a stratified sampling design with randomly selected 50 to 100 samples per class.

Table 6: Example of land cover / strata proportion of a land cover map

map strata	proportion	description
AG	0.25	Agriculture in terrestrial and aquatic/regularly flooded land
TCO	0.16	Trees closed-to-sparse in terrestrial and aquatic/regularly flooded land
SCO	0.16	Shrubs closed-to-sparse in terrestrial and aquatic/regularly flooded land
HCO	0.20	Herbaceous closed-to-sparse in terrestrial and aquatic/regularly flooded land
BS	0.10	Bare rocks and soil and/or other unconsolidated material(s)
URB	0.05	Urban areas
WAT	0.08	Seasonal/perennial, natural/artificial waterbodies

The confusion matrix is built by entering the number of class combinations in the correct cells of the matrix.

Table 7 provides the raw confusion matrix for the example data. The columns are the field data and the rows are the map classes. If for example at 4 sample points the map's land cover has been classified as AG but the surveyor has observed the class TCO the value 4 has to be entered in the corresponding cell of map class AG and reference class TCO. The main diagonal of the matrix shows the number of corresponding land covers. E.g. 65 for AG. The row total provides the number of samples per map land cover class. They should reflect the approximate number of samples which have been selected from the stratified grid, minus sample points which could not be observed. The column total provides the total of the observed land cover classes.

The raw confusion matrix already provides some good information about the map accuracy. E.g. 65 samples out of 97 in the class AG actually belong to class AG (this proportion would express the user's accuracy). Or from 94 samples belonging to class AG 65 have been classified correctly. Also the intra class confusion can be analysed, for example 80 sample points have been observed in the class Bare soil (BS), but only 52 of them actually belong to this class. The majority of misclassified samples belong to Urban (15) and to Agriculture (8).

Table 7: Example of a sample confusion matrix for 7 land cover classes and 515 sample points

	Reference data - field data observed on the ground							
map classes	AG	TCO	SCO	HCO	BS	URB	WAT	Row total
AG	65	4	1	8	9	10	0	97
TCO	2	55	13	4	0	3	0	77
SCO	2	22	42	6	1	1	0	74
HCO	14	5	6	59	8	1	0	93
BS	8	0	2	2	52	15	1	80
URB	2	2	0	1	4	37	0	46
WAT	1	0	0	0	1	0	46	48
Column total	94	88	64	80	75	67	47	515

Before estimating the accuracy parameters the counts of the raw confusion matrix are transferred to proportions. Each cell value is divided by the row total and multiplied with the mapped proportion of its class. E.g. for the class AG: 65/97*0.25=0.168 for the next cell 4/97*0.25=0.010, and so on. See table 9. From this confusion matrix the recommended accuracy parameters can be calculated.

Table 8: Example for a confusion matrix with proportions and different accuracy parameters

	Reference data - field data observed on the ground									
map classes	AG	TCO	SCO	HCO	BS	URB	WAT	Row total	User's accuracy	Error of commission
AG	0.168	0.010	0.003	0.021	0.023	0.026	0.000	0.25	0.67	0.33
TCO	0.004	0.114	0.027	0.008	0.000	0.006	0.000	0.16	0.71	0.29
SCO	0.004	0.048	0.091	0.013	0.002	0.002	0.000	0.16	0.57	0.43
HCO	0.030	0.011	0.013	0.127	0.017	0.002	0.000	0.20	0.63	0.37
BS	0.010	0.000	0.003	0.003	0.065	0.019	0.001	0.10	0.65	0.35
URB	0.002	0.002	0.000	0.001	0.004	0.040	0.000	0.05	0.80	0.20
WAT	0.002	0.000	0.000	0.000	0.002	0.000	0.077	0.08	0.96	0.04
Column total	0.22	0.19	0.14	0.17	0.11	0.10	0.08	1		
Producer's accuracy	0.76	0.62	0.67	0.74	0.57	0.42	0.98		OA	0.68
Error of omission	0.24	0.38	0.33	0.26	0.43	0.58	0.02		Kappa	0.62

Overall accuracy: Sum of the "correct" proportions of the main diagonal.

0.168+0.114+0.091+0.127+0.065+0.040+0.077=0.68

68% probability that a random point is mapped correctly.

The variance and standard error of the overall accuracy are measures for the variability of the overall accuracy estimate.

The variance of the overall accuracy is the sum of each class's:

*(mapped area proportion²*user's accuracy*(1-user's accuracy) / (total samples per class -1))*

$0.25^2*0.67*0.33/(97-1)$

$+ 0.16^2*0.71*0.29/(77-1)$

$+0.16^2*0.57*043/(74-1)$

.....

=0.00044321

The standard error of the estimated overall accuracy is the square root of the calculated variance:

√0.00044321=0.0210523 or 2.11%

User´s accuracy: For each class the row total divided by the correct classification value. E.g. for AG: 0.168/0.25=0.67

67% probability that a random point from class AG actually belongs to this class.

Error of commission is the complement, 0.33 or 33%.

Producer's accuracy: For each class the column total divided by the correct classification value. E.g. for AG: 0.168/0.22=0.76.

76% probability that a random point belonging to class AG is correctly mapped as AG.

Error of omission is the complement, 0.24 or 24%.

Kappa coefficient: (Overall Accuracy – chance agreement) / (1-chance agreement).

The chance agreement is the sum of row total multiplied by column total from each class.

Chance agreement =

(0.22*0.25)+(0.19*0.16)+(0.14*0.16)+(0.17*0.20)+(0.11*0.10)+(0.10*0.05)+(0.08*0.08)= 0.16.

Kappa= (0.68-0.16)/(1-0.16)=0.62 or 62%

Accuracy reporting should always include the raw confusion matrix[29], the normalised confusion matrix and the map strata proportions. This provides the sufficient base to calculate the most common accuracy parameters.

2. Shortcomings of a traditional field data collection and validation approach

Based on critical literature review and real project outputs this chapter briefly lists the problems affecting the validation of Land Cover information. Those shortcomings are ranging over various levels such as given in the next sections.

2.1. Logistical constraints

Before discussing technical or scientific aspects, prevailing shortcomings at a logistical level, such as (i) limited resources for field work, (ii) overall feasibility and high costs of surveys and (iii) interdependencies between field data collection and remote sensing applications, should be addressed here.

In a brief survey of selected institutions and mapping initiatives [30] and through numerous applied experiences in science, administration and professional service provisions it became obvious, that in a significant part of the land cover mapping projects adequate attention and respective resources for appropriate field work activities are planned only insufficiently in advance.

Even if field data collection has been considered beforehand it is linked to enormous costs for transportation and fuel and is to be executed under time constraints in dependence to the prevailing vegetation period or crop season as well as to the respective acquisition dates and acquisition windows of remote sensing data. This issue is present in any country and irrespective of the thematic focus of the land cover mapping project.

There is a need of a good understanding of the overall project approach in order to properly address ground truthing initiatives. But on the other hand side the fact that in general the institution, which is doing the land cover mapping is also carrying out the ground truth data collection, creates situations of interdependences. This can possibly result in conflict of interests, particularly in view of data validation.

In essence available resources for field work are in general limited often due to neglecting the need for proper accuracy assessment of map products during the project planning. On the other hand there is often a deficit in the validation methods to address the cost factor of ground truthing. The

[29] Strahler et al. 2006
[30] SIGMA 2014

logistical constraints and related costs should be properly addressed to apply the in general limited resources to fully employ the impact field surveys can contribute to the land cover mapping tasks.

2.2. Vagueness of class meaning and its impact on the validation results

As described in chapter 1.1 above the observed field parameters must match the final outcomes of respective map legends and class nomenclatures to adequately compare classified map results with the in-situ reference data (see chapter 1.6). Although that is an obvious criterion it is practical not always easy to achieve. The map result is generally based on interpretation and classification of remote sensing, the in-situ data observed in the field. While the map makers look from above the surveyor is at the side. Parameters which are obviously detectable form the remote sensing perspective might be difficult or assess in the field and vice versa. This might require different class descriptions and guide lines for each activity, one for the map making and one for the ground truthing. An example is the handling of heterogeneous areas or land cover polygons which consist of a combination of land cover elements, e.g. grassland and forest. In the traditional validation approach using a confusion matrix and a “hard” comparison between in-situ and map values the problem of heterogeneous polygons is insufficient addressed. The definition of association tables (1.6.1) is picking up the problem but not fully addressing it.

It becomes even more problematic when the nomenclature and class definitions are not are comprehensively documented beforehand. In the above mentioned brief survey [31] 10 out of 10 mapping initiatives are making use of field data for validation purposes, whereas none having standardizations in terms of field survey approach or nomenclature.

As this is addressing the traditional approach it is not addressing at all on how to handle interactive land cover data bases or object based classification approaches, where strict class definitions are substituted through various class elements, which in turn can appear in numerous combinations, as for instance given with FAOs LCML[32].

2.3. Lack of standardized procedures to evaluate/asses polygon land cover maps

The classical validation approach in general considers the comparison at pixel level[33] which depends on the chosen image source and sensor type. The above collated classical approach (chapter 1) has not been developed in a period of very high resolution satellite images (VHRSI) but in an epoch of 20 – 30 m ground resolution of one pixel. Validating at pixel level in this context may have been more appropriate, but is lacking behind the today’s depth and detail of VHRSI and various new means such as UAV and drone borne images. Based on Herold et al. (2006) and Möller et al. (2013) new ways to establish mapping accuracy objectives towards thematic and position criteria should be explored in

[31] SIGMA 2014

[32] Di Gregorio 2005

[33] Congalton, 1991 and Congalton & Green 2009

adequate relation to the related geographical scale and source information beyond the purely pixel based classical approach.

The traditional comparison at pixel level does not consider the intended geographical scale or the dimension of minimum mapping units applied in recent land cover maps which use polygons as map units.

The validation of map polygons, especially large aggregated areas requires procedures to assess the accuracy of the entire polygon. Apart from the assessment of the correctly mapped land cover classes the validation of land cover maps with polygons (e.g. based on segmentation and visual interpretation) has to take into account the positional accuracy of the borders of the polygons. This is an aspect which is so far not much addressed in the traditional remote sensing approaches which focuses on raster map based approaches. There is a need for standard methods to assess the positional border accuracy of polygons and there is a need for standards to quantify and express the polygon accuracy.

Polygons or segments e.g. from an image segmentation in general represent visible land cover borders and should be easy to allocate in the field. Nevertheless depending on the landscape and the mapping scale it can be very difficult to allocate the precise outline of the unit.[34] A further difficulty is that map polygons can be very large and complex and it becomes impracticable to observe the entire polygon in the field. To validate the positional accuracy of the boundaries of land cover polygons the use of remote sensing data sets (pseudo-truth) seems to be the most practical approach. This could be a comparison of the selected polygon outlines with best available data e.g. Google earth, Bing maps or others. The positional accuracy could be described by the deviation of the polygon border using defined tolerance values.

Accuracy assessment considering the geometric quality of a classification, which could also be addressed in thematic map accuracies, have been explored by Möller et al. (2013).

2.4. Insufficient application of statistical procedures both for the sample strategy and for the final analysis of the assessment

Generally speaking, a sound accuracy assessment needs a precisely defined sampling design (e.g. sampling unit sizes distribution etc.) as one of the core components of accuracy assessment along with a proper documentation of the applied methods and data. For example information on positional accuracy of the image data applied, or the time of field data collection are rarely stated in accuracy assessment and thus not allowing to assess sampling bias [35]. Standards for accuracy assessment reporting could help to overcome this problem.

The core procedure in the above collated classical approach is the collection of a given number of samples allocated over the target areas, which is brought in confusion with the classification results. Statistical rules to define the sufficient number and the appropriate allocation are derived through

[34] Lea and Curtis 2010

[35] see Hammond & Verbyla 1996

various sources, whereas in the light of the given logistical and budgetary constraints, mostly rules of thumb are applied as described in chapter 1.3 - if any rational decision is being laid down at all. The sample selection should consider accuracy targets, already defined during the validation planning, to include measurements for the significance of the validation itself. This could be confidence intervals and confidence levels (see chapter 1.6.2).

Other approaches address the variance of selected samples to define the optimal sample size. This sequential sampling methods as explored by Carfagna et al. (2012), Torregiani et al. (2012), or Carfagna & Marzialetti (2009).

The selection of samples per class is directly impacting the commonly applied confusion matrix (chapter 1.6) which is often reduced to the value of the 'Overall Accuracy' only. As this value summarizes the degree of agreement over all classes, it has only limited degree of information without the full information including omission and commission error at class level.

Advanced statistical means to assess the confidence of the given results are currently addressed too short in the mostly applied classical approach. Foody (2002) showed that the use of the confusion matrix is often inappropriate, and he identified eight problem areas in accuracy assessment, such as the choice of accuracy measure (e.g. overall accuracy, kappa coefficient to name just two), the sampling design (e.g. units size and distribution) and other types of errors are neglected (e.g. caused by mis-registration of images, or mis-registration between images and field data).

2.5. Lack of standardized procedures to collate "pseudo truth"

One major limitation of the validation by means of in-situ data is the costs and constraints of accessibility of entire target areas, due to difficult terrain conditions, security issues or simply lacking road network infrastructures. One of the advantages of the remote sensing is to overcome these impediments. The original need for reference data sets are to be independent, objective and as precise as possible but certainly more detailed as the land cover map to be validated. Pseudo truth such as very high resolution satellite images, or airborne images or even derived from drones can be used as reference data under certain requirements. This may open possible application to generate reference data allowing the validation of land cover maps at a suitable scale, i.e. more coarse than the input reference data. Nevertheless the method itself can be subject for processing errors and is to be applied carefully, given that inaccuracies of the pseudo-truth data will mislead the final accuracy results. Measures to quantify those input inconsistencies as well as rules defining the adequate relation and technical specifications on how to do and apply pseudo-truth are not yet applied at an operational level.

3. Synthesis of some major components for an advanced field data collection and validation approach

This part is a complementary synthesis of the previous two sections; it briefly reviews some critical components of "traditional" approaches and proposes aspects still to be addressed in order to come to an innovative advanced validation approach.

Specific components of this section are the following below.

3.1. Map legend and cartographic standards

The map legend is where the map producer instantiate the informative content of the data base, in other words is a means with which he share a certain type of knowledge of a specific geographic area with a community of end users. It is a critical theme because a lot of different information obtained through the use of different remote sensing techniques and field observations must be correctly and efficiently shared with a vast and heterogeneous end user community. Unfortunately this process is far from perfect and usually is a source of vagueness and ambiguity that not only affect the map readability but inevitably influence a correct and objective assessment of its accuracy. Categorization has always been a useful method to minimize the complexity of the real world and it is familiar to many aspects of our life. The process of categorization, however, has a lot of limitations that are partly inherent with its intrinsic nature of grouping the real world phenomena in a certain number of artificial categories and partly related to the methods with which its "formalization of the meaning" is ratified. Reality is by its nature a "continuum", and any partition of this virtual continuum into categories is intrinsically arbitrary.

In addition very little effort is generally put to the way how we explain the thematic content ("formalization of the meaning") of each class. Formalization of the meaning is the way how the author of a classification system (legend) make official and manifest the ontology (intend as "meaning "or "significance" of the things) of the categories and explain their relationship. The persistent use of simplified text description (a class name with class description) exacerbates the inherent problems of any "generalization" process. It introduces supplementary constraints that increase the fuzziness of the data and therefore create huge problems when an objective and transparent accuracy assessment must be put in place.

Class definitions can be often imprecise, ambiguous or, some time, absent. The build up of the definition in the form of a narrative text is unsystematic (many diagnostic criteria forming the system are not always applied in a consistent way) and in any case do not always reflect the full extent of the information, the amount and type of information is not logically structured. The accuracy assessment work often faces several limitations related to the legend semantic content. Some of them are listed below:

- Semantic understanding of the legend relaying only on class name, without any text description.
- Classes with an ambiguous meaning or referring to a "process" rather than to objective parameters and therefore impossible to be assessed in the field (e.g. "Transitional Woodland-Shrub").

- Classes with no clear threshold parameters (e.g. "Land principally occupied by agriculture with *significant* areas of natural vegetation").

Even when a text class explanation start with a rather strict description, in the prosecution of the text a large amount of exceptions "vaporise" the initial class meaning in a large labyrinth of exemptions and allowances that results in an increasing vicious circle of vagueness and ambiguity (see note 1). Often some vagueness in the class definition is artificially included by the map producer to hide some "technical problems" when mapping a certain land feature.

The graded and fuzzy nature of common Land Cover categories derived from traditional classifications/legends has been recognized for a long time by the remote sensing community, however limited debate has been done to analyse the effect of this problem on accuracy assessment outputs and even lesser efforts have been proposed to sort out semantic uncertainty in land cover accuracy assessment studies. It must be clearly stated that the efficiency, transparency and exactness of an accuracy assessment procedures starts from a clear and unambiguous understanding of the land features to be observed in the field. Therefore a legend must at least exhibit the following principles:

- Use of consistent, unique and systematically applied classificatory principles.
- The classes are all unique, mutually exclusive and unambiguous.

These are, however, just the minimum requirements, more advanced methods for building up of thematic legends exist, and their use will have a cascade of advantages in the assessment and analysis of map accuracy. One example is the FAO "parametric approach" using the proprietary software LCCS v.1-2. Recently this original idea have been further developed in a fully "object oriented" Land Cover Meta Language (LCML) that has become in 2012 an ISO standard.

The Land Cover Meta Language (LCML) (and its derived application tool LCCS v.3) provides a consistent and objective classification scheme with a defined set of objects and relative attributes to represent any land cover. The idea is that is deemed more important to standardize and make manifest the attribute terminology rather than relay in a class name (and/or text class description).

The basic concept is that a specific land feature is defined by a set of distinct diagnostic biotic and abiotic objects and related attributes. The advantage of this system for the field work is that land cover observation remains very objective by describing a set of defined objects (trees, shrubs etc.) without the need for the surveyor to classify the land cover class in its complexity.

The LCCS concepts and a detailed description are provided in the LCCS version 2 manual by FAO. The manual for the LCCS 3 or Land Cover Meta Language (LCML) is in preparation. The software package for LCCS is available through the FAO GLCN homepage.

In addition to the problems of correct understanding of the class meaning as above explained, the surveyors face another important problem the different perspective a specific land feature can have if observed from remote sensing or from the ground. Land features can have different appearance when observed from the ground and the separation from other contiguous land features that is easy from remote sensing can became difficult in the field. Therefore is extremely important to define which elements (and related parameters like %cover, height, management practices etc.) are constitutive of a certain land features and therefore are key elements to be observed in the field. If the legend has been generated with a modern "object oriented" system (like FAO LCML/LCCS) this process is already done because the "parametrization" is inherent to the definition of a class. In the other cases the original legend classes must be translated in a set of clear and unambiguous elements and parameters to be observed in the field. It should not be the task of the surveyor to assess the presence or absence of a certain feature in the field, the surveyor's task is to systematically describe the selected sample in the field. The correlation field sample vs. map class will be done later and not necessarily by the surveyor who has done the field work.

A map can be defined as the representation of a certain geographic area in a bi-dimensional form. This complex representation of the "real world" passes through a composite process of "generalization". Mainly this is instantiated through the way how we express the "semantic content" of a legend as discussed previously. There are, however, other methods with which we simplify the reality. They can be grouped and named as "cartographic standards". They are those types of generalizations we use to overpass the constrain of scale and other related problems when we create a map. Their influence can be crucial in the procedure of accuracy assessment. A brief list and relative effect on the accuracy assessment is listed below:

- Scale or MMU (minimum Mapable Unit): scale is the obvious parameter we use to understand the relationship of a map and the geographic area it want to represent. In modern digital maps, however, scale is replaced by the concept of MMU, defined as "*the minimum ground surface explicitly represented in a map unit"*. It is imperative any map product shows these information and a validation procedure must take these parameters into account because they influence the size of the sampling units (if sample units have to be defined). Moreover the MMU concept (in absence of other cartographic standards, see next point below) may affect the expected level of generalization of a map and therefore the assessment of its accuracy.
- Mixed units: linked with the concept of MMU, there is the approach with which a map producer addresses the representation of the so called "heterogeneous areas" (complex areas formed by a combination of small patches of different land features each of one smaller than the MMU). In modern map this problem is addressed by creating a specific class (not really an efficient method) or coding the related polygon with multiple coding that reflect the codes of the different land features (assuming they are listed in the legend). This representation of complex areas must be supported by clear and unambiguous parameters (for instance the % range of occurrence of each land feature, or the minimum % of occurrence of a feature to be represented as mixed unit). For instance in the FAO AFRICOVER cartographic standards a given class can be represented as mixed unit if it occupy at least 10 % (in specific cases 20 %) of the polygon surface, otherwise its presence is ignored and the polygon is represented by the code of the dominant class only. The consequences of these rules are important as they can determine the attribution of a "correct" or "wrong" assessment. This type of advanced understanding of the "cartographic rules" is however depending both from the detailed definition of those standards (often neglected) and from the sampling method adopted.

Room for innovation
• *The map thematic content and its formalization of the meaning* • *Defining the legend classes to be checked* • *Class parameterization (defining the rules and conditions to clearly define the thematic meaning of a class)* • *Defining the applied cartographic standards and their implication in the accuracy assessment* • *Defining class frequency, area and geographic distribution for a correct sampling design*

3.2. Sampling design

Defining the objective of the accuracy assessment: One of the distinctive differences between a map and a conventional tabular statistic is, that a map not only assesses the total surface area of each class but in addition is able to geo-locate the different land features. This peculiarity is often an important and decisive justification of the time and resources needed to conduct a mapping activity. It is therefore unexpected that this specific aspect is not normally properly evaluated in a conventional accuracy assessment. Unfortunately an assessment of the ability of a map to correctly geo-locate the land features (Mapping accuracy versus Area accuracy) is complicated and time consuming. In case of an assessment of the "mapping accuracy" it is mandatory the sample unit must be the "map polygons" and procedures must be adopted to evaluate if a given polygon correctly represent a specific land feature. This implies a correct understanding of the "cartographic standards" of the map and possibly a massive use of pseudo truth (VHR images). Diffuse availability of VHR images (Google Earth, Bing etc.) could make this type of assessment more feasible than in the past. It is, however, needed substantive work to make this type of assessment systematized in order to be commonly applied.

Sampling methods: The concepts of sequential sample selection procedures to select and define appropriate number of samples (Carfagna et al., 2012) could be opportunities for innovative sampling methods in the sphere of map validation frameworks. The Adaptive sequential procedure (ASPRN) particularly takes into account the scale and legend of the data (in this case, maps to validate) in order to select a representative number of samples per class and scale. The first step of the method is stratification, which is a key step to assess the optimal number and distribution of samples. The following steps measure the variance of randomly selected samples of each strata[36]. Further explorations should be addressed towards the potentials and limitations to adequately integrate 'pseudo-truth' methods into this context.

In any case robust guidelines on sampling design will provide the opportunities to define and measure confidence of accuracies at class level. A next focus should be to extent this to the level of class elements.

Combined use of "pseudo truth" and field observation: Use of VHR images as "pseudo truth" is a rapidly expanding practice that will allow a quicker and cheaper accuracy assessment of maps. In many cases VHR images are a complementary tool to field observations. Therefore more effective procedures should be established to enhance this powerful combination of information. For instance, while the ground truth is able to effectively and undoubtedly assess certain types of information, the supplementary combined use of VHR images can, at certain level, help to evaluate the level of "homogeneity" of the surrounding area of the observed sample zone. This would be particularly useful if, for instance, the totality of a polygon must be evaluated. The use of an "object oriented" method to formalize the meaning of a certain class is an added value to this procedure. Assuming we need to check a class named "Apple orchard". The "object oriented" classification system will represent the class using a main object called trees. The main object trees will be further

[36] Carfagna & Marzialetti 2009

enriched in its semantic meaning by a series of different attributes as: regular disposition of the trees, cover, leaf type, leaf phenology and finally the floristic aspect of the object tree that is in this case Apple sp. Only a field observation can assess if the observed trees in the sample area are Apple sp. this specific information, however, can be combined with VHR images that are able to assess all the other characteristics peculiar of this class (trees, regular disposition, cover, leaf type and leaf phenology). If the polygon is homogeneous for a series of attributes (trees, regular disposition etc.) derived from the use of VHR images and the field observation confirm the trees in the sample area are Apple sp., there is a convergence of evidences that make very high the level of confidence on the assumption that the whole polygon is correctly coded and the land feature correctly represented in the map.

Room for innovation
• *Defining the objective of the accuracy (Area versus Mapping accuracy, position and thematic accuracy)* • *Sampling methods* • *Stratification criteria* • *Assessment total observations per class and definition of expected level of confidence of accuracy results* • *Combined use of "pseudo truth" and field observation* • *Sample selection*

3.3. Survey guidelines

Beyond the development of survey instructions for any single mapping activity (chapter 1.4) it is felt more relevant to employ adequate rules and guidelines on how to plan, set up and execute surveys for established mapping frameworks, such as FAOs LCCS[37] or the recent LCML in order to contribute and lead a standards development through consolidated concepts (chapter 2.2). That is currently lacking. As part of the forthcoming LCML instructions 3.0 it is recommended to implement basic principles for classical validation approaches but also to develop approaches to adequately assess class elements rather than aggregated classes. These survey guidelines do then need to consider as well aspects of innovative sampling designs (chapter 3.2) and statistical innovations (chapter 3.1).

This should also include well consolidated approaches and rules to implement 'pseudo-truth' observations in order to reduce the logistical burdens (chapter 2.1) and to increase the acceptance of validation standards.

Room for innovation
• *Survey planning* • *Observation rules and guideline on the field and for "pseudo truth"*

[37] Di Gregorio 2005

3.4. Data analysis

As explored in chapter 2.4 above there is still a need to adequately apply the principles of the classical confusion matrix for fixed class definitions. Even the documentation of entire confusion matrices is not yet a standard for land cover mapping initiatives. Investigations might help to understand the main obstacles which might explain this prevailing deficit.

The increasing utilization of “object oriented” classification systems to generate a map legend can potentially open new perspectives in the assessment of land cover maps. If the thematic content of a class, instead that with a simple name is formalized by the combination of a series of objects and their relative attributes, then it is theoretically possible not to limit the assessment only to the class as a whole but to detail it for each one of the objects (and attributes) representing the class itself. The potentiality of this method can have large implications for the flexibility with which the end user community can utilize the results of an accuracy assessment.

The end user community is large and variegate. A map can be utilized for a variety of different applications that do not necessarily need to use the totality of the information a specific class represents, in this case the accuracy just by class in its totality is reductive and rigid. In figure 8 an example of this type of "flexible" assessment is shown. To apply this method, in addition to an "object oriented" classification, specific computational efforts are necessary.

FAO has developed a software (MAP) able to calculate the accuracy of the single elements forming a class, this software, however works only with an older version of LCCS (LCCS v.2). An updated version for LCCS v.3 is forecast to better explore the advantages of this approach.

Name	code	GIS code	LCCS elements	% accuracy by element
Enset (false banana)	1En	11217-12626-S13Zs1	A2 = Shrub Crops	90
			B2 = Small Sized Field(s)	90
			C2 = Intercropped (Second Crop)	90
			D1 = Rainfed Cultivation	90
			D9 = Permanently Cropped Area	90
			C3 = One Additional Crop	90
			C7 = Herbaceous Terrestrial Crop (Additional Crop)	90
			C17 = With Simultaneous Period (Second Crop)	90
			ZS1 = User defined code:Enset	60

Figure 8: The figure shows how an LCCS class is structured, in the first four columns there is the class name, the map code, the LCCS GIS code and the list of the all elements used to build up the class itself. If the accuracy is calculated in the traditional way by the entire class the resulting accuracy in the example will be very low (60%), however if the accuracy it is calculated separately for each of the elements forming the class, a new flexibility on the use of data materializes. The user can, for instance, utilize the class information (organized in LCCS elements) with a high accuracy and remove the ones with a lower one. This option gives a more flexibility on the use of final data.

Room for innovation
• *The confusion matrix* • *Accuracy assessment by class and by class elements.*

3.5. Transparency and credibility

The overall aim of a land cover mapping product is to serve as geospatial information source which is widely used and applied. Confidence and credibility are therefore an utmost objective, which still receives too little attention. Although not really a new invention, transparent planning, process documentation and access to the validation results, calculation and data helps to overcome doubts, and to strengthen the awareness of appropriate map validations. The above shown shortcomings of the traditional approach are certainly valid (chapter 2), but no reason not to start at least with minimum measures as collated in chapter 1. An ultimate goal should be, that land cover maps without a comprehensive documentation and access to validation results is to be handled with questions. An example on a sound validation framework could be implemented was demonstrated through the GMFS3 project of the European Space Agency[38].

Room for innovation
• *Definition of survey parameters ahead of the mapping activities* • *Documentation of all steps including full access to calculations after the mapping activities*

[38] GMFS3 2014

Reference list

Baret, F., Nightingale, J., Garrigues, S., Justice, C., 2009. Report on the CEOS Land Product Validation Sub-group Meeting. *The Earth Observer* 21, 26–30.

Büttner, G., Maucha, G., *European Environment Agency, 2006. The thematic accuracy of Corine land cover 2000: assessment using* LUCAS (land use cover area frame statistical survey). EEA, Copenhagen.

Carfagna, E., Gallego, F.J., 2005. Using Remote Sensing for Agricultural Statistics. *International Statistical Review* 73, 389–404.

Carfagna E. & Marzialetti J. (2009) Sequential Design in Quality Control and Validation of Land Cover Data Bases, *Journal of Applied Stochastic Models in Business and Industry (ASMBI)*, Vol. 25, N. 2, 2009, 11 pp. 195-205.

Carfagna E. , Tassinari, P., Zagoraiou M., Benni S., Torreggiani D. (2012) Efficient statistical sample designs in a GIS for monitoring the landscape changes, in *Advanced Statistical Methods for the Analysis of Large Data-Sets*, Di Ciaccio A., Coli M. and Angulo Ibanez J.M. (editors) Springer.

Cochran, W.G., 1977. *Sampling techniques*, 3rd ed. John Wiley & Sons, New York.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of Environment* 37, 35–46.

Congalton, R.G., Green, K., 2009. *Assessing the accuracy of remotely sensed data: principles and practices*, 2nd ed. ed. CRC Press/Taylor & Francis, Boca Raton.

Di Gregorio, A., 2005. *Land cover classification system: classification concepts and user manual : software version 2* : based on original software 1 prepared by Antonio Di Gregorio and Louisa J.M. Jansen. Food and Agriculture Organization of the United Nations, Rome.

Eurostat, 2012. *LUCAS 2012 Technical Reference Document: C1-Instructions for Surveyors*. [Online] Available from http://epp.eurostat.ec.europa.eu/portal/page/portal/lucas/documents/LUCAS2012_C1-InstructionsRevised_20130110a.pdf (03.08.2014)

Foody, G.M., 2002. Status of land cover classification accuracy assessment. *Remote sensing of environment* 80, 185–201.

Gallego, J., Carfagna, E., Baruth, B., 2010. Accuracy, objectivity and efficiency of remote sensing for agricultural statistics, in: Benedetti, R., Bee, M., Espa, G., Piersimoni, F. (Eds.), *Agricultural Survey Methods*. John Wiley & Sons, Chichester, U.K., pp. 193–211.

Gallego, J., Delincé, J., 2010. The European land use and cover area-frame statistical survey, in: Benedetti, R., Bee, M., Espa, G., Piersimoni, F. (Eds.), *Agricultural Survey Methods*. John Wiley & Sons, Chichester, U.K., pp. 151–168.

GMFS3 *Validation protocol*. [Online] Available from http://www.gmfs.info/uk/publications/gmfs3_docs/GMFS3_C05_ValidationProtocol_v3%201.pdf (17.07.2014).

GMFS3 *Service Validation Report Agricultural Mapping Sudan 2011*. [Online] Available from http://www.gmfs.info/uk/publications/gmfs3_docs/GMFS3_C06_AM_SDN_v2.3.pdf (17.07.2014).

Hammond, T.O., Verbyla, D.L., 1996. Optimistic bias in classification accuracy assessment. *International Journal of Remote Sensing* 17, 1261–1266.

Herold, M., Woodcock, C.E., Di Gregorio, A. di, Mayaux, P., Belward, A.S., Latham, J., Schmullius, C.C., 2006. A joint initiative for harmonization and validation of land cover datasets. *IEEE Transactions on Geoscience and Remote Sensing* 44, 1719–1727.

Kerdiles, H., Dong, Q., Spyratos, S., Gallego, F.J., 2014. Use of high resolution imagery and ground survey data for estimating crop areas in Mengcheng county, China. *IOP Conference Series: Earth and Environmental Science* 17, 012057.

Lea, C., Curtis, A.C., 2010. *Thematic Accuracy Assessment Procedures: National Park Service Vegetation Inventory*, version 2.0. (No. NPS/2010/NRR-2010/2014), Natural Resource Report. National Park Service, Fort Collins, CO.

Möller, M., Birger, J., Gidudu, A., Gläßer, C., 2013. A framework for the geometric accuracy assessment of classified objects. *International Journal of Remote Sensing* 34, 8685–8698. doi:10.1080/01431161.2013.845319

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment* 148, 42–57.

Olofsson, P., Stehman, S.V., Woodcock, C.E., Sulla-Menashe, D., Sibley, A.M., Newell, J.D., Friedl, M.A., Herold, M., 2012. A global land-cover validation data set, part I: fundamental design principles. *International Journal of Remote Sensing* 33, 5768–5788.

Stehman, S.V., Czaplewski, R.L., 1998. Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing of Environment* 64, 331–344.

Strahler, A.H., Boschetti, L., Foody, G.M., Friedl, M.A., Hansen, M.C., Herold, M., Mayaux, P., Morisette, J.T., Stehman, S.V., Woodcock, C.E., 2006. *Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps* (No. 25).

SIGMA, 2014. Outcomes of a questionnaire survey on applied field data collection and validation approachs in SIGMA Annual project meeting – presentations of WP 3.3. Beijing 22. -25. 10. 2014.

Torreggiani D.; Benni S.; Zagoraiou M, Tassinari P. Carfagna E. (2012) A GIS for Rural Landscape Sampling, in *International Conference of Agricultural Engineering, CIGR-AgEng2012*, Agriculture & Engineering for a Healthier Life. Valencia, CIGR AgEng, 2012, pp. 1-6.