Product validation plan

Milestone	Product validation plan
Authors	Frédéric Achard, Sophie Bontemps, Pierre Defourny, Martin Herold, Philippe Mayaux
Distribution	ESA : Olivier Arino, Vasileios Kalogirou

Summary

This document describes the Product Validation Plan of the CCI Land Cover project. This first version of this document (to be updated during the project) provides the design of the validation plan and describes the methods that will be applied during the validation exercise.

This validation plan is based on the lessons learned previously and is aimed at fulfilling the CEOS Working Group on Calibration and Validation stage-3 validation requirements, i.e. (i) uncertainties in the product will be quantified from comparison with suitable reference data and (ii) spatial and temporal consistency of the product will be evaluated. The validation results are expected to develop further international confidence on these global land cover products.

This Product Validation Plan deals primarily with the thematic accuracy of the CCI land cover products. Information on the positional accuracy assessment will be available in the document related to geometric correction. The thematic accuracy of the three global land cover maps for "epochs" 2000, 2005 and 2010 will be assessed using reference datasets acquired at similar time windows (few-years length) compared to the input data time acquisitions of the products. The validation of the products will be implemented independently from the production phase.

The overall validation process of the product relies on 3 complementary pillars: (i) the confidence-building procedure, (ii) the statistical accuracy assessment, (iii) the comparison with other global land cover products and (iv) the temporal consistency assessment the between the CCI land cover products.

Confidence-building procedure (section 2 of the document)

A confidence-building procedure will be performed, which will consist in a systematic quality control of the CCI land cover products. This procedure is intended to meet two main objectives: the elimination of macroscopic errors and an increase in the overall acceptance of the land cover product by users. Systematic quality control is also a way of assessing if the remotely sensed data have been correctly classified, i.e. if the errors are due to limitations of data quality rather than to poor classification procedures.

The quality control is based on a systematic descriptive protocol in which the map is segmented into regular cells, each cell is visually examined and the cells accuracy is documented in terms of type of error, landscape pattern, reference material used, etc. Ascertaining the nature of the errors occurring in the cell is of primary importance. Indeed, statistical accuracy assessment merges in the category "error" many different cases that quality control can easily document. Such information can be profitably used for improving the map during the updating phase. Once all the cells have been visited, the results will be presented in a tabular manner or on a map. It will then be possible to investigate the influence of the parameters (heterogeneity, dominant class) on the quality of the land cover map.

This systematic quality control will be integrated into the classification procedure, with the results of the analysis employed for removing errors and improving the map.

Statistical accuracy assessment (section 3 of the document)

Building a validation database (section 3.1)

The project will optimize the use existing reference datasets, in particular by relying on the following datasets:

1. The Landsat Global Land Survey (GLS) database. Nearly complete global coverage from the Landsat satellites is now available at no cost from USGS. The GLS products at 28.5 m × 28.5 m resolution were created from the epochs circa 2000 and mid-2000s. GLS-2010 is under creation.

2. Data derived from other moderate or higher resolution satellite imagery: "TropForest" dataset covering sample units over South America and South East Asia for epoch 2010, and complementary SPOT imagery to be acquired through Third Parties Missions. Google Earth imagery and multi-temporal NDVI profiles will also be used during the interpretation phase.
3. The GLC2000 and GlobCover reference validation databases. These datasets will be used for the comparison with other global products.

Sampling strategy (section 3.2)

As data and resources availability must be optimized, the selection of the reference data to be interpreted and used to validate the CCI land cover products needs to be selected in a statistically valid manner. For this purpose, the systematic sampling of the JRC TREES dataset (based on a lat/long geographical grid) combined with a two-stage stratified clustered sampling has been considered as the most efficient sampling.

A stratified random sampling (two strata) allows the selection of primary sampling units (PSU) from the entire population of potential primary units (i.e. sample units selected from the systematic grid with equal area probability), while secondary sampling units (SSU) are systematically distributed within the PSU sample units. The resources dedicated to the validation exercise constrain the number of SSUs to be interpreted. For this PVP, 2600 PSUs are initially selected as target number to allow achieving a pre-required accuracy precision.

Only a limited part of each medium resolution satellite scene will be analyzed to produce a PSU interpretation. It is planned to interpret five SSUs per 20 km × 20 km box, which will result in 5 times more SSUs than PSUs. The SSUs will be located at (i) the centre of each 20 km × 20 km box and (ii) at a distance of 5 km × 5 km from the centre of each box.

Image interpretation procedure (section 3.3)

The use of remote sensing specialists with local expertise has proven to be the most efficient strategy for the interpretation of medium resolution imagery. The application of image pre-processing to the reference satellite imagery is intended to allow for a more consistent identification of land cover objects. The moderate and/or high resolution satellite imagery will be pre-processed with a few automatic steps: orthorectification based on orbital parameters, extraction of PSU boxes and radiometric calibration. An automatic object-oriented approach is also envisaged with a minimum mapping unit (MMU) of 5 hectares for segmentation of the satellite imagery. A few (five) land-cover classes can be produced from the automatic operational land cover processing chain of the JRC over the tropics (to be used as initial land cover information to be refined). The moderate and/or high resolution dataset will be used in combination with other satellite imagery such as Google Earth imagery and multi-temporal NDVI and NDWI profiles (to display seasonal variations of vegetation) available from year 2000 through UCL.

Experts from previous validation networks will be contacted for interpretation of the Landsat-type satellite imagery. Such experts will be invited to visit the premises of one of the CCI Land Cover team members with a total duration of one week work with each expert. All necessary datasets and infrastructure will be put at the disposal of the experts. The experts will use a graphical interface to interpret Landsat-type imagery over the SSUs. The interpretation legend will be developed according to the LCCS classifiers; the interface for the interpretation will be developed from the existing GlobCover interface and will combine satellite imagery and ancillary information, including existing maps. A manual will be distributed to the experts with the description of the method to be used for the labeling of the SSUs. During the phase of interpretation by experts, quality control procedures will be introduced including repetition of SSU interpretations, comparison between interpretations from different experts and analysis of very fine spatial resolution for a sub-sample. The analysis of very fine spatial resolution (e.g. 10 m × 10 m) should allow quantifying the absolute accuracy of the products, but will be limited by availability of such imagery.

Reporting of results (section 3.4)

At the end of the project a validation report will describe the implementation of the validation procedures and their results. This validation report will include the various parameters describing the accuracy of the map: contingency matrix, user's and producer's accuracy, Kappa statistics, and area statistics.

Accuracies will be derived by comparing CCI land cover products with the results of the independent interpretations over SSUs. Accuracies of individual categories will be reported through the user's and the producer's accuracy. A number of different user perspectives can be envisaged for land cover (such as the carbon content and net primary productivity) for the use of the products into different climate models. For each of these potential uses, a matrix of similarity between classes will be constructed to derive specific accuracies.

Results are expected to be published in peer-reviewed scientific journals.

Comparison with other products (section 4 of the document)

Comparison with other land cover products will be made to develop confidence in the CCI land cover products. The comparison will be performed for the most recent global land cover products and the CCI land cover products, including the GlobCover products and the MODIS derived land cover products. The comparison approach used will consider on the principles for harmonizing land cover information based on LCCS.

Temporal consistency assessment (section 5 of the document)

Consistency assessment will be achieved between the three CCI-LC products (derived from the three different "epochs" of 2000, 2005 and 2010). It will compare them, one to one, on a per-class basis to identify discrepancies. These discrepancies will be identified either as land cover change or as temporal instabilities, using the outcomes of the other 3 previous validation steps.

Table of recorded changes

Issue Record Table

ISSUE	DATE	REASON
1.0	January 30, 2011	Completion of the first version
1.1	March 11, 2011	Update of the first version, based on RIDs received from ESA and on comments made during PM3
1.2	June 10, 2011	Update of the previous version, based on comments received from ESA during PM4
1.3	July 05, 2011	Update of the previous version, based on RIDs received from ESA

Detailed Change Record Table

ISSUE	RID No.	SECTION	CHANGE
1.1	Progress	1.2	A more complete introduction of the validation plan has been added, including the accuracy and stability targets, a description of the CCI products to be validated and a general and concrete presentation of the validation process
1.1	Progress	2	A new section has been added, to present the confidence-building procedure that will be performed before the statistical validation
1.1	Progress	3	The structure of the section 3 has been updated
1.1	Progress	4	The comparison with other existing land cover products is presented in a separate section
1.1	Progress	5	A new section has been added, to present the temporal consistency assessment as a fourth validation step
1.1	PVP/1.	1.2.2	This information has been added as a section in the introduction
1.1	PVP/2.	3.1.2	A detailed description of the validation dataset is provided in section 3.1.2 and a table summarizes the information in section 3.1.2.3
1.1	PVP/3.	3.1.1.1	The TropForest dataset has been further detailed (including KOMPSAT-2)
1.1	PVP/5.	3.1.1	The list is not yet determined, but rationale for the expert selection has been included in the

			document
1.1	PVP/7.	3.4.1	The discussion on the calculation of overall accuracy and Kappa statistics by region (continent) and by global biomes was expanded. Details about geographic stratifications were provided through 2 figures
1.1	PVP/8.	3.4.2	Section 3.4.2 has been expanded to better showcase how user-related accuracy reporting will be performed. A weighting matrix is shown from an example of the URD.
1.1	PVP/9.	3.4	A summary of the accuracy measures was added at the end of Section 3.4.1. Examples of aggregation rules based on PFT are provided in Figure 17.
1.2	PM4-7	1.2.2	Table 2 has been updated to show consistency with the PSD (Table 7 of this deliverable): OA target is set to 80% and temporal stability target is set to 80-85%
1.2	PM4-7	3.2.4.2	Figures have been added to include spatial representation of the validation data acquisition
1.2	PM4-7	5	The temporal consistency assessment will be achieved on a per-class basis
1.3	PVP_1.2/10	Reference	GlobCover 2009 validation report has been added
1.3	PVP_1.2/14.	3.2.4.2	Table has been added to provide an estimation of PSU coverage for each validation dataset for the three epochs.
1.3	PVP_1.2/15.	3.3.6	Information about the availability of the different dataset has been added in a new section
1.3	PVP_1.2/16	3.2.4.2	Captions of Figure 13 and Figure 14 have been clarified
1.3	PVP_1.2/17	4	The comparison with other existing products will be achieved on a per-class basis

Symbols and acronyms

ALOS	: Advanced Land Observing Satellite
ATBD	: Algorithm Theoretical Basis Document
AVNIR	: Advanced Visible and Near-Infrared Radiometer
CCI	: Climate Change Initiative
CCI-LC	: Climate Change Initiative – Land Cover
CEOS-WGCV	: Committee on Earth Observing Satellites Working Group on Calibration and Validation
CMC	: Climate Modelling Community
CMUG	: Climate Modelling User Group
DARD	: Data Access Requirement Document
ECV	: Essential Climate Variable
EDC	: EROS Data Center
EO	: Earth Observation
EROS	: Earth Resources Observation Systems
ESA	: European Space Agency
ETM	: Enhanced Thematic Mapper
FAO	: Food and Agriculture Organisation of the United Nations
FR	: Full Resolution
FRA	: Forest Resource Assessment
GCOS	: Global Climate Observing System
GLC	: Global Land Cover
GLS	: Global Land Survey
GOFC-GOLD	: Global Observation of Forest Cover – Global Observation of Land Dynamics
GTOS	: Global Terrestrial Observing System
IGBP	: International Geosphere Biosphere Program
JRC	: Joint Research Centre
KOMPSAT	: Korea Multi-Purpose Satellite
Landsat	: Land Remote Sensing Satellite
LCCS	: Land Cover Classification System
LPVS	: Land Product Validation Subgroup
MERIS	: Medium Resolution Imaging Spectrometer
MODIS	: Moderate Resolution Imaging Spectroradiometer
MMU	: Minimum Mapping Unit
NASA	: National Aeronautics and Space Administration
NDVI	: Normalized Difference Vegetation Index
NDWI	: Normalized Difference Water Index
PFT	: Plant Functional Type
PSD	: Product Specification Document
PSU	: Primary Sampling Unit
PVP	: Product Validation Plan
RR	: Reduced Resolution
SDSU	: South Dakota State University
SPOT	: Système pour l’Observation de la Terre
SPOT-VGT	: SPOT-VEGETATION
SSU	: Secondary Sampling Unit
TM	: Thematic Mapper
TPM	: Third Parties Mission

TREES : Tropical Ecosystem Environment Observations by Satellite
UCL : Université catholique de Louvain
URD : User Requirement Document
USGS : United States Geological Survey

References

Applicable documents

[AP-1] ESA, 2010, ESRIN/Contract No. 40000101774/10/I-LG - Programme Phase 1 of the ESA Climate Change Initiative – Land_Cover_cci

[AP-2] ESA, 2009, CCI Statement of Work version1.4. Climate Change Initiative Phase 1 - Scientific User Consultation and Detailed Specification – Statement of Work. Ref: EOP-SEP/SOW/0031-09/SP , 09/11/2009

[AP-3] Defourny P. et al., 2010, ECV Land Cover Part 3 - Technical proposal for the European Space Agency. 05/03/2010

[AP-4] Bojkov B. et al., ESA Climate Change Initiative - Validation of CCI Products. 1st CCI collocation meeting, ESA/ESRIN, 15/10/2010

[AP-5] ESA Land Cover CCI Project, User Requirement Document (URD), v. 2.2, 22/02/2011

[AP-6] ESA Land Cover CCI Project, Data Access Requirement Document (DARD), v. 1.6, 02/03/2011

[AP-7] ESA Land Cover CCI Project, Products Specification Document (PSD), v. 1.1, 15/03/2011

[AP-8] ESA Land Cover CCI Project, Algorithm Theoretical Basis Document (ATBD) – Pre-processing, v. 1.0, 31/01/2011

[AP-9] ESA Land Cover CCI Project, Algorithm Theoretical Basis Document (ATBD) – Classification, v. 1.0, 31/01/2011

Reference documents

Beuchle R, Eva H., Stibig H-J et al. A Satellite Data set for Tropical Forest Change Assessment, *International Journal of Remote Sensing,* accepted

Bicheron, P., Defourny, P., Brockmann, C., Schouten, L., Vancutsem, C., Huc, M., Bontemps, S., Leroy, M., Achard, F., Herold, M., Ranera, F. and Arino, O. GlobCover: products description and validation report, ESA GlobCover project (2008). Available at: http://ionia1.esrin.esa.int/

Bontemps, S., Defourny, P., Van Bogaert, E., Kalogirou, V. and Arino, O., GlobCover 2009 - Products Description and Validation Report (2010). Available at: http://ionia1.esrin.esa.int/

Defourny, P., Schouten, L., Bartalev, S., Bontemps, S., Caccetta, P. , de Wit, A.J.W., Di Bella, C., Gérard, B., Giri, C., Gond, V., Hazeu, G.W. , Heinimann, A. , Herold, M. , Knoops J., Jaffrain, G., Latifovic, R., Lin, H., Mayaux, P., Mücher, C.A. , Nonguierma, A., Stibig, H.-J., Van Bogaert, E., Vancutsem, C. , Bicheron, P., Leroy, M. and Arino, O. Accuracy Assessment of a 300 m Global Land Cover Map: The GlobCover Experience. *Proceedings of the 33rd International Symposium on Remote Sensing of Environment*, Stresa, Italy (2009)

Di Gregorio, A. & Jansen, L. J. M. Land cover classification system (LCCS): Classification concepts and user manual. GCP/RAF/287/ITA Africover—East Africa Project and Soil Resources, Management and Conservation Service, Food and Agriculture Organization, Rome, Italy (2000)

FAO, JRC, SDSU, UCL. The 2010 global forest resources assessment remote sensing survey: an outline of the objectives, data, methods and approach. Food and Agriculture Organization of the UN, Rome (2009)

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A. and Huang, X. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, *Remote Sensing of Environment*, 114, 168-182 (2010)

Herold, M., Woodcock, C., Mayaux, P., Baccini, A. and Schmullius, C. Some challenges in global land cover mapping: an assessment of agreement and accuracy in existing 1 km datasets, *Remote Sensing of Environment*, 112, 2538-2556 (2008).

Herold, M., Woodcock, C., Cihlar, J., Wulder, M., Arino, O., Achard, F., Hansen, M., Olsson, H., Schmulllius, C., Brady, M., Di Gregorio, A., Latham, J. and Sessa, R. Assessment of the status of the development of the standards for the Terrestrial Essential Climate Variables: T9 Land Cover, Global Terrestrial Observing System Report 64, FAO, Rome, Italy (2009). Available at: http://www.fao.org/gtos/pubs.html

Loveland, T., Zhu, Z., Ohlen, D.O., Brown, J.F., Reed, C. and Yang, L. An analysis of the IGBP global land-cover characterization process, *Photogrammetric Engineering and Remote Sensing*, 1021-1032 (1999)

Mayaux, P. & Lambin, E.F. Estimation of tropical forest area from coarse spatial resolution data: A two-step correction function for proportional errors due to spatial aggregation *Remote Sensing of Environment*, 53 (1), pp. 1-15. (1995)

Mayaux, P., Holmgren, P., Achard, F., Eva, H., Stibig, J.P. and Branthomme, A. Tropical forest cover change in the 1990s and options for future monitoring, *Philosophical Transactions of the Royal Society*, 360, 373-384 (2005).

Mayaux, P., Eva, H., Gallego, J., Strahler, A., Herold, M., Shefali, A., Naumov, S., de Miranda, E., di Bella, C., Johansson, D., Ordoyne, C., Kopin, I. and Belward, A. Validation of the Global Land Cover 2000 Map, *IEEE Transactions on Geoscience and Remote Sensing* 44(7-1), 1728-1739 (2006)

Scepan, J. Thematic validation of high-resolution global land-cover data sets, *Photogrammetric Engineering & Remote Sens*ing, 65, 9, 1051-1060 (1999)

Schouten L., Defourny P. and Vancutsem C. GlobCover Validation Plan, MEDIAS-France report to ESA (2006)

Stadelman, M., Curtis, A., Vaughan, R., Bailey, M., Convis, Ch., Goodchild, M., Davis, F., Li, X., Goodin, K. and Grossman, D. Accuracy assessment procedures, final draft; NBS/NPS Vegetation Mapping Program, ESRI, NCGIA and the Nature Conservancy, Redlands (1994)

Stehman S. Sampling designs for accuracy assessment of land cover, *International Journal of Remote Sensing*, 30, 20, 5243-5272 (2009)

Strahler, A.H., Boschetti, L., Foody, G.M., Friedl, M.A., Hansen, M., Mayaux, P., Morisette, J.T., Stehman, S.V. and Woodcock, C.E. Global Land Cover Validation: Recommendations for evaluation and accuracy assessment of global land cover maps, Office for Official publications of the European Communities, Luxembourg (2006)

Table of content

1. Introduction and objectives

The European Space Agency (ESA) Climate Change Initiative (CCI) projects will deliver the next generation of satellite derived geophysical parameters, with quantified uncertainties that will allow each parameter to be assessed against requirements from the Global Climate Observing System (GCOS) for Essential Climate Variables (ECV) and the Climate Modelling Community (CMC), represented within the CCI program by the Climate Modelling User Group (CMUG). A critical step in the acceptance of the CCI products by the GCOS and CMC communities is providing confidence in the quality of each CCI product and its uncertainties through validation against independent data such as ground based reference measurements or alternate estimates from other projects and sensors.

The objective of the document is to describe the Validation Plan of the CCI Land Cover (CCI-LC) project. This plan will be updated during the course of the project. The current version of the validation plan gives a description of the methods and designs that will be applied in the validation of the CCI-LC products.

1.1. Validation: definition and standard protocols

The first CCI co-location on 15th September 2010 at ESA/ESRIN addressed a few main issues relating to validation which are summarized hereafter with recommended guidelines. Consideration was only given to validation of "Level 4" products (i.e. final results from analyses of lower level data (i.e. Level 0 or 1 input data)).

1.1.1. Definition of "validation"

There are several definitions of validation available from various agencies, and it was agreed that the Committee on Earth Observing Satellites Working Group on Calibration and Validation (CEOS-WGCV) definition would be adopted within the CCI program, which defines validation as:

> *"The process of assessing, by independent means, the quality of the data products derived from the system outputs".*

It is assumed that the term "data products" in the above definition refers to both the geophysical parameter and its uncertainties, so it is vital that all available information on data uncertainty is used and validated.

1.1.2. Independence of validation process

The CCI project will produce a set of output products that require validation, including in particular, any associated quality indicators and uncertainties. Ideally the validation process should follow clearly defined protocols and should be independent from the production process. The independence of the validation process should follow three requirements:

1. CCI project shall use, for validation, in situ or other suitable reference datasets that have not been used during the production of their CCI products.
2. CCI project teams shall consider the independence of the geophysical process and ensure that if a particular auxiliary dataset is used in the production of their CCI products then, the same dataset is not used in the validation.

3. CCI project teams shall ensure that the validation is carried out by staff not involved in the final algorithm selection; ideally the validation of the CCI products should be carried out by external parties, i.e. by staff / institutions not involved in the production of the CCI products.

The present CCI-LC project Product Validation Plan (PVP) shall adhere to the above three requirements regarding independence.

1.1.3. Protocols

It is as important to mention that only a transparent traceable validation procedure will be accepted by the user community and is mandatory for the CCI-LC product. This is especially important for validation procedures which rely on statistical quantities. Here below, the CCI-LC consortium uses established, community accepted, traceable validation protocols and follows terminology approved by the CEOS Land Product Validation Subgroup (LPVS) (http://lpvs.gsfc.nasa.gov/).

1.1.4. Selection of datasets for validation

Independent reference data for validation is selected to ensure complete coverage of the various epochs of the CCI-LC products. Therefore, the selection of validation datasets follow different levels of rigour depending on the level of quality of each dataset, thus making sure that some level of confidence can be given to every output product. Each CCI-LC product should contain an indication of the level (or confidence) in the data quality resulting from the validation process.

Possible levels may include validation with:

1. Independent in situ data (the “true” reference dataset);
2. Other in situ data;
3. Reference data interpretations from high-resolution satellite data;
4. Large scale comparisons with other satellite datasets;
5. Large scale comparisons with historic datasets;
6. Impact studies using other CCI products.

This approach (levels of validation) is adopted by the CEOS Land Product Validation Subgroup with a four level approach to validation that depends on the temporal and spatial coverage of available reference data, thus providing a confidence estimate in each product even where little if any in situ data exists (Table 1). The product validation proposed here is aiming to fill the CEOS WGCV stage 3 validation requirements; noting that for a full ECV monitoring implementation, a stage 4 validation will be required.

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 characterized 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.

Table 1. Four levels of validation adopted by the CEOS Land Product Validation

1.1.5. When is a CCI product considered "validated"?

The validation is an ongoing process that shall take into account requirements and responses from users. The validation process should use approved community protocols where they exist and must be fully traceable and subject to scrutiny by peer-review by an independent international board of experts. A CCI product will be deemed to be validated once all steps of the validation process documented in the PVP have been completed and documented accordingly.

1.2. Objective and design of the product validation plan

The objective of this CCI-LC project PVP is to describe the strategies and methods which have been selected prior to and during the actual validation process. The description of these strategies and methods are required in order to show to both the user community (i.e. the CMC but also the land cover community) and ESA that the ultimate figures of accuracy have been derived in a sound and reproducible way. The plan not only describes the selected strategies and methods but explains also the rationale of the methodological choices.

The main objective of the validation is to allow a potential user to determine the "map's fitness for use" for his / her application. This concern is addressed in section 1.2.1, where the user requirements in terms of validation are reminded. Section 1.2.2 presents the components of the CCI-LC project that are validated. The PVP concepts and process are detailed in sections 1.2.3 and 1.2.4.

1.2.1. Validation requirements from the users' point of view

During the first three months of the CCI-LC project, a user requirement analysis has been conducted to derive the specifications for a new global land cover product to address the needs of key-users from the CMC. The complete analysis can be found in the User Requirement Document (URD) [AP-5].

In terms of validation, the findings of the user requirement analysis highlighted that:

- **Consistency** among the different model parameters is often more important than accuracy of individual datasets, and it is important to understand the **relationship between land cover classifiers with the parameters** and the relative importance of different land cover classes;
- The relative importance of **different class accuracies** varies significantly depending on which surface parameter is estimated and the need for stability in accuracy should be reflected in implementing a multi-date accuracy assessment;
- Quality of land cover products need to be **transparent** by using quality flags and controls, and including information on the probability for the land cover class or anticipated second class or even the probability distribution function for each class (coming from the classification algorithm).

The user requirements assessment also resulted in quantitative outputs. At least two levels of requirement were identified:

- The **threshold requirement**, standing for the limit at which the observation becomes ineffectual and is not of use for the climate-related application;
- The **target requirement**, which is the maximum performance limit for the observation, beyond which no significant improvement would result for climate applications.

These two levels of requirements are detailed in Figure 1.

	Threshold requirement	Target requirement
	Coverage and sampling	
Geographic Coverage	*Global*	*Global with regional and local specific products*
Temporal sampling	*Best / stable map and regular updates*	*Monthly data on vegetation dynamics and change*
Temporal extent	*1-2 years, most recent*	*1990 (or earlier)-present*
	Resolution	
Horizontal Resolution	*1000 m*	*30 m*
Vertical Resolution	-	-
	Error / Uncertainty	
Precision	*Thematic land cover detail sufficient to meet current modeling user needs*	*Thematic land cover detail sufficient to meet future model needs*
Accuracy	*Higher accuracy than existing datasets*	*Errors of 5-10% either per class or as overall accuracy*
Stability	*Higher stability than existing datasets*	*Errors of 5-10% either per class or as overall accuracy*
Error Characteristics	*Independent one-time accuracy assessment*	*Operational and independent multi-date validation*

Figure 1. Threshold and target requirements for land cover products, resulting from the URD

1.2.2. Components of the CCI-LC product to be validated

The validation process only concerns the land cover maps which will be generated by the CCI-LC project. State-of-the-art radiometric calibration is applied to deliver the best possible surface reflectance products but the absolute radiometric accuracy will not be assessed by any independent sources of reflectance data. The precision and the relative radiometric accuracy in space and over time are indeed considered to be much more critical than the absolute accuracy performance since any classification algorithm dealing with reflectance time series relies on temporal consistency and proceeds by relative statistical comparison or similarity analysis in space. In other terms, the most important is the possibility to work with surface reflectance products described by specific quality indicators.

This phase of quality control will take place in the development phase for the daily and composite surface reflectance products. It will be based on the following indicators: (i) reflectance dynamics based on the overall spectral reflectance distribution, (ii) temporal variance at the pixel level for the various spectral reflectance values and (iii) local variance for the various spectral reflectance values within a LC class and across LC classes. These indicators will allow quantifying the discrimination potential of a given time series. The obtained values will be compared with reflectance products available from other sensors and other projects (e.g. the GlobCover project).

In addition, this PVP only focuses on the quality assessment of the land cover “state” products. The land cover “condition” will not be validated by independent reference dataset at this stage. The definition of the land cover “state” and “condition” concepts are provided in the Product Specification Document (PSD) [AP-7]. By sake of clarity, the “land cover state products” will be referred simply to as “CCI-LC products” in the rest of the document.

The quality of a product like the CCI-LC product one can be a property of various aspects of the product. It can relate to:

- The positional / coordinate accuracies of the geospatial products. The quality will be determined by establishing its overall positional accuracy and by determining the internal positional consistency of the product;
- The precision and the accuracy of the land cover state that describes a geographic feature. The quality will be assessed through comparison between the land cover state identified by the product and the actual state of the feature as determined by suitable reference datasets. It is also affected by the positional accuracy;
- The stability of the information over a given period of time is important to characterize in order to assess the consistency of the product as required by the CMUG. The quality of CCI-LC products, corresponding to a few temporal periods, will be assessed using similar-time reference datasets.

Information on the positional accuracy assessment can be found in the Algorithm Theoretical Basis Document (ATBD) related to the pre-processing steps [AP-8], and more particularly in the section related to the geometric corrections. This document presents the tool developed for the step of automated pattern matching of optical Earth observation (EO) data, which will also be further used to assess the geo-location accuracy of the final products. A set of Land Remote Sensing Satellite (Landsat) Thematic Mapper (TM) images will be used as reference for this purpose. It is however important to realize that a low positional accuracy would have consequences for the thematic accuracy.

This PVP primarily deals with the thematic accuracy of the land cover state of the CCI-LC product. The concepts and methodologies used to assess this thematic accuracy are presented in sections 1.2.3 and 1.2.4 and will be addressed in more details in the rest of the document.

The issue of temporal stability will be addressed in a more general sense. The CCI-LC products to be assessed will be the three global land cover maps generated for the “epochs” around 2000 (based on SPOT-VEGETATION (SPOT-VGT) time series), 2005 and 2010 (both derived from the Medium Resolution Imaging Spectrometer (MERIS) Full and Reduced Resolution (FR and RR) and SPOT-VGT data). They will be validated using reference datasets acquired at similar epochs (time windows) compared to the input data time acquisitions of the products. Furthermore, an overall inter-comparison between the different CCI-LC products will allow quantifying the product stability per biome or strata.

Table 2 presents the CCI-LC project target in terms of overall accuracy and temporal stability. It shall be noted that the accuracy assessment only focuses on the LC state products, the LC condition ones being not validated by independent reference dataset. These figures are slightly slower than the target requirements expressed in the URD but meet the threshold requirements (i.e. the need to have higher accuracy and stability than existing datasets). They should allow making significant progress with regard to current global land cover products and therefore, taking a step forward to fulfil the 90-95% target requirements.

Overal accuracy	**Temporal stability**
80%	80-85%

Table 2. Targets in terms of overall accuracy and temporal stability for the CCI-LC products

1.2.3. Concept of the PVP

The validation is essential for providing a high quality product that is accepted and applied by the user community. Different steps of validation that jointly lead to the achievement of the validation objectives (Figure 2) are anticipated in the project. The internal validation procedures are described in the ATBD, both for pre-processing and classification steps ([AP-8 and AP-9]). This PVP describes the approach used to perform independent thematic product validation and to allow users to provide assessment and feedback.

The independent validation of global land cover products has been already applied operationally to three well-known products: International Geosphere Biosphere Program (IGBP) DISCover (Scepan et al., 1996), Global Land Cover (GLC) 2000 (Mayaux et al., 2006) and Globcover (Defourny et al., 2009), while the accuracy of the Moderate Resolution Imaging Spectroradiometer (MODIS) Global Land Cover was evaluated by a cross-validation analysis (Friedl et al., 2010). The overall accuracies of the IGBP-DISCover, GLC2000, GlobCover and MODIS Global Land Cover (through cross-validation analysis) products have been reported at 67%, 69%, 73% and 75% respectively.

The current validation exercise (CCI-LC) is based on the lessons learned from these previous projects, and it is intended to establish a precursor for operational land cover validation activities in collaboration with international scientific community dealing with this topic, in particular the Global Observation of Forest Cover – Global Observation of Land Dynamics (GOFC-GOLD) panel of the Global Terrestrial Observing System (GTOS) (http://www.fao.org/gtos/gofc-gold/index.html). The latest information and protocols produced by the GOFC-GOLD Land-Cover Implementation Team (http://www.gofc-gold.uni-jena.de/) and the CEOS Working Group of Calibration and Validation (Land cover validation subgroup) have been used for designing this PVP. In particular, the design and implementation of the validation plan and the creation of the reports follows the general recommendations of the GOFC-GOLD validation report (Strahler et al., 2006) and other scientific publications from these groups (Herold et al., 2009, Mayaux et al., 2006).

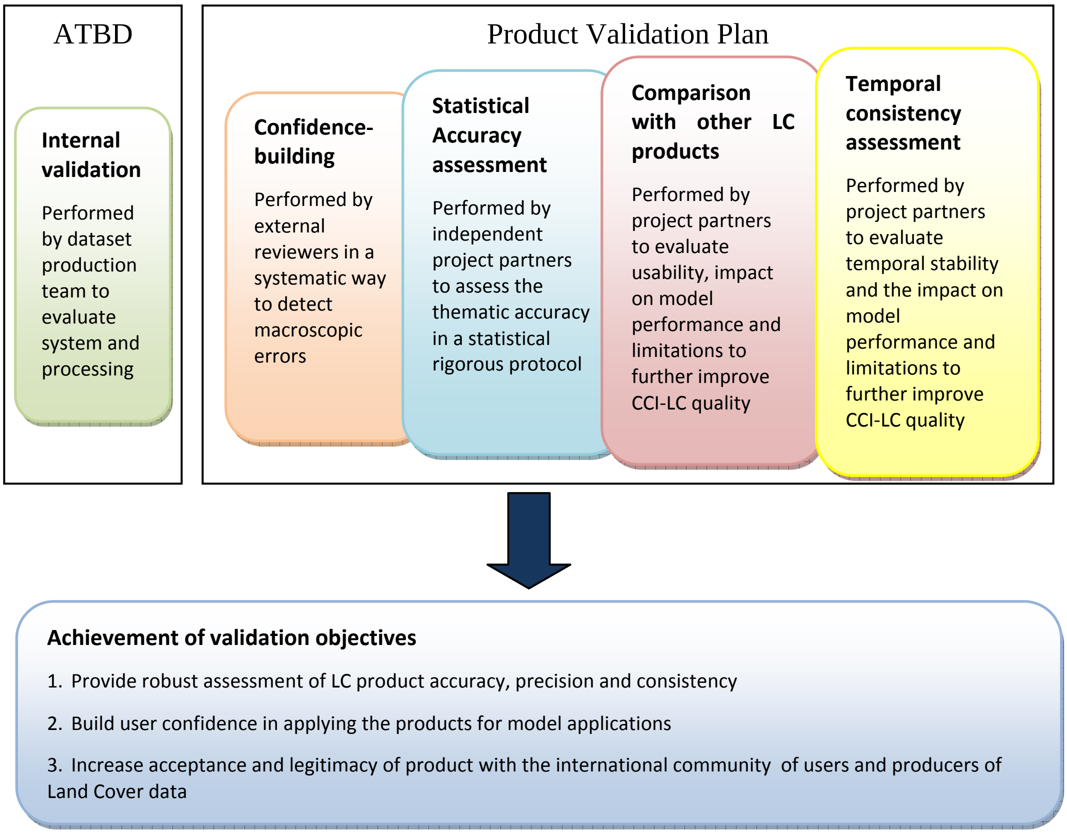

Figure 2. Overall organization of the validation and related user assessment activities

In the current document, the different elements of the validation activities are detailed, pointing out the challenges to face and the long-term perspectives. The validation activities will be implemented independently from the production. Different partners will be performing the accuracy assessment in close cooperation with the international community and various experts in order to provide a validation exercise accepted and useful beyond the CCI-LC project. The validation results are also expected to support the comparison with other products and to build international confidence on these global land cover products in relation to specific purposes.

Unlike previous global land cover exercise, the CCI-LC will deliver several products corresponding to different epochs (2000, 2005 and 2010). A one-shot effort for the thematic validation is therefore not appropriate and a longer-term data validation tool (allowing validating future land cover products but also historical ones) should be designed. However, the absence of decision on the CCI-LC option hampers the design and the development of such permanent and collaborative validation environment.

1.2.4. Overall validation process

As shown in Figure 2, the overall validation process of the product relies on 4 complementary pillars: (i) the confidence-building, (ii) the statistical accuracy assessment, (iii) the comparison with other global LC products and (iv) the temporal consistency assessment.

Prior to the independent and statistical quantitative assessment of the thematic accuracy of the CCI-LC products, a confidence-building procedure will be conducted in order to assess the quality of the map in a systematic manner. This step is aimed at reinforcing the overall acceptance of the land cover product by users. The results of such a qualitative systematic assessment will also allow investigating the influence of different parameters on the quality of the land cover map such as: heterogeneity, class dominance, etc.

A complementary comparison with other land cover products will also be performed between the most recent global land cover products and the CCI-LC maps. This comparison has the objective of building confidence in the CCI-LC products. The comparison is, to some extend, driven by the notion, of "best" available map. Therefore, there is a need to quantify advantages of the different maps (spatial resolution, temporal update, thematic accuracy, etc.) and seek opportunities to combine the most detailed and accurate areas of each dataset to a new global dataset.

At the end of the exercise, a temporal consistency assessment between the three CCI-LC products will be conducted. This assessment is totally driven by users' requirements. It will compare the three products, one to one, on a per-class basis with the twofold objective of identifying discrepancies and making the difference between land cover changes and temporal instabilities.

The general relationships between the different steps are shown in Figure 3.

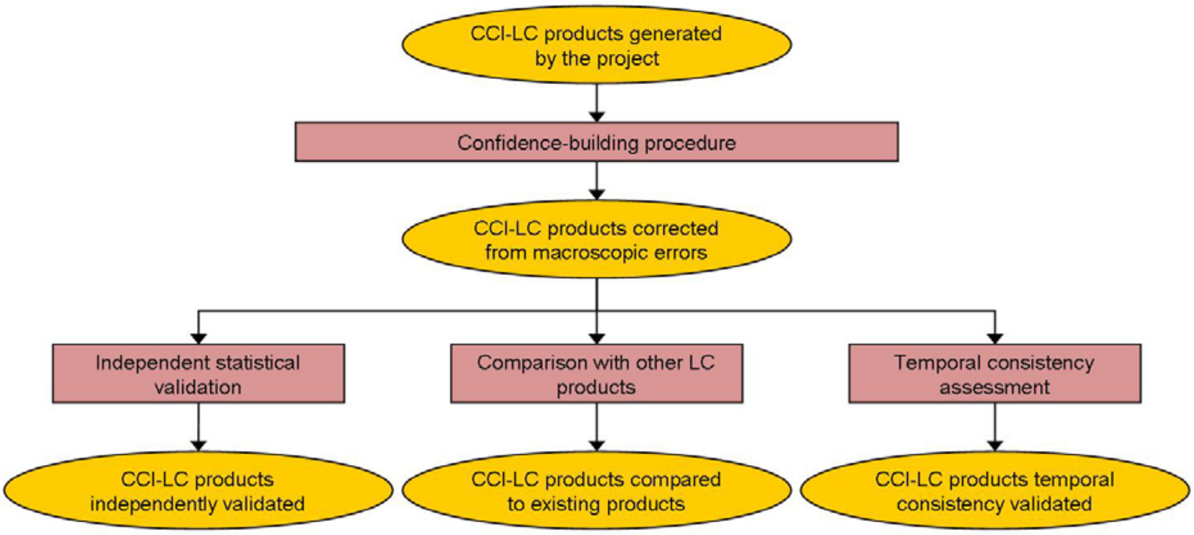

Figure 3. Organization and links between the different validation components

The structure of the PVP details the four validation components and for each one, presents the associated steps. Figure 4 summarizes the structure of the rest of the document.

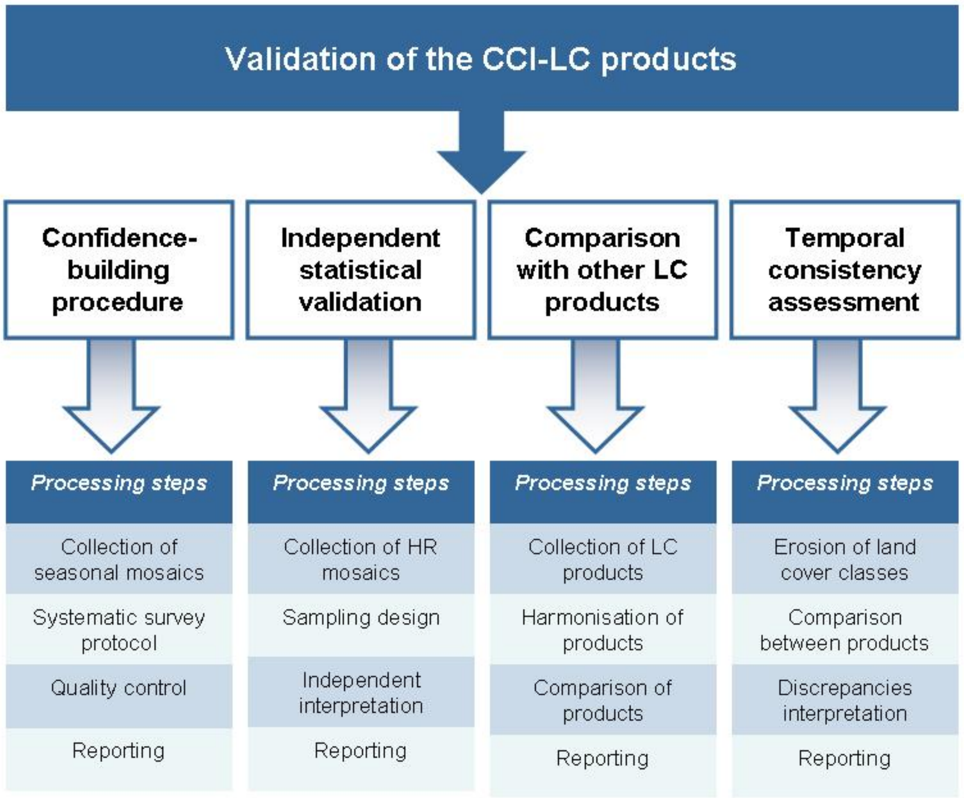

Figure 4. Detailed description of the different validation components

2. Confidence-building procedure

Systematic quality control arises because recent global land cover products, although of good overall quality, exhibit in some parts major errors that could be avoided by a careful review of the draft products. Such errors reduce the user's overall confidence in the products, even if the quantitative accuracy is high. Errors affecting accuracy of thematic maps can be caused by confusion between the land cover classes (wrong label, missing classes) or can be spatial errors (wrong position of the boundary between classes, disappearance of small patches). The identification of systematic biases affecting some land cover classes or some regions of the world can influence the quantitative validation (sampling strategy) and area estimates. For example, obvious errors in the Siberian part of GlobCover2009 have reduced the global acceptance of the product, while its main characteristics in terms of overall accuracy and legend made this product unique for specific users, like climate modellers.

Systematic quality control is intended to meet two main objectives: the elimination of macroscopic errors and an increase in the overall acceptance of the land cover product by users. Systematic quality control is also a way of assessing if the remotely sensed data have been correctly classified, i.e., if the errors are due to limitations of data quality rather than to poor classification procedures. Systematic quality control should be integrated into the classification procedure, with the results of the analysis employed for removing errors and improving the map.

2.1. Systematic protocol and comparison material

Qualitative validation is based on a systematic descriptive protocol, in which each cell of the map is visually examined and its accuracy documented in terms of type of error, landscape pattern, reference material used, etc. The grid size could be adapted to the characteristics of the landscape, the map, and the reference material. For example, in the central part of the Amazon Basin or in the heart of the Sahara, the grid cells can be much larger than in the complex landscapes of Western Europe. But for simplicity reasons, we suggest to keep a uniform cell size of 200 to 400 km on a side as a target for providing a good idea of the overall quality of a global product, keeping in mind that the goal of this exercise is a quick survey. As illustration, Figure 5 presents the grid that was used in the GLC2000 confidence-building exercise.

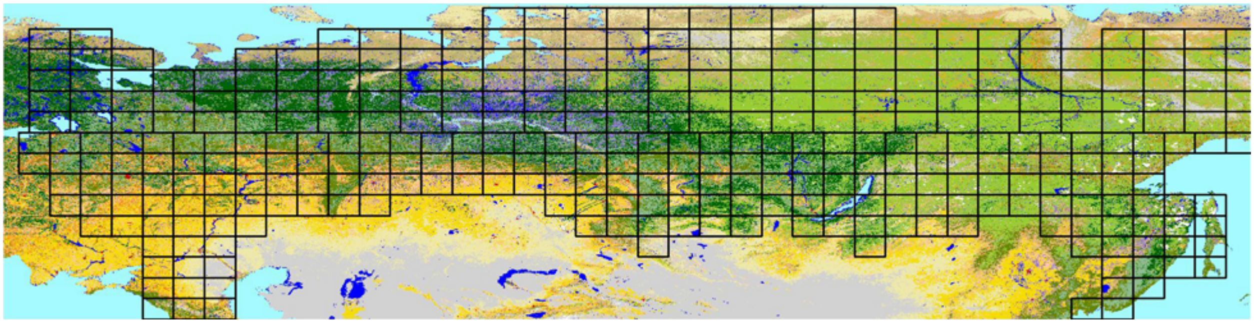

Figure 5. Grid used for the systematic survey of the initial products

For the systematic survey, different reference materials can be used, including single-date coarse resolution images, detailed thematic maps, and quick-look imagery derived from fine-resolution sensors. Pre-processing and classification procedures applied to multi-date imagery often lead to the loss of many spatial details that are clearly visible on original images or temporal synthesis. This loss of detail is particularly obvious when long time series of derived parameters such as vegetation or moisture indexes are used as input for the classification. For consistency purposes, we will use the best seasonal mosaics used for the classification for this visual examination. A careful examination of the phenological cycles should be conducted in order to choose the most characteristic period(s) of the year. For example, the winter and summer mosaics can be needed in Siberia, while an annual mosaic is sufficient in the evergreen part of the tropics.

2.2. Quality control

In a systematic quality control exercise, each cell examined during the quality control procedure is characterized in detail by a few parameters: the composition and the spatial pattern of the cell, its comparison with other existing global land cover products, the overall quality of the cell, and the nature of any problems.

The cell composition is a key factor affecting the precision of a map because some land cover classes (*e.g.,* evergreen forests, deserts, water bodies) are easier to discriminate than others (*e.g.,* deciduous forests or woodlands, grasslands, extensive agriculture). Information on the composition of the cell contributes to a better understanding of the errors and can help to stratify the population, as in design-based inference. On the other hand, some users focus on specific land cover classes and will be interested in a spatial representation of the errors for cells dominated by their class of interest.

It is widely recognized that the spatial pattern of the landscape influences the appearance or disappearance of land cover classes at varying resolution as well as the area estimates derived from coarse resolution maps (Mayaux and Lambin, 1995). Landscape heterogeneity can be expressed by means of qualitative definitions (*e.g.,* highly fragmented, moderately fragmented, little fragmented, not fragmented) or by quantitative metrics (*e.g.,* diversity, perimeter-area ratio, mean patch size). A catalog of qualitative fragmentation categories should be completed before starting the evaluation process in order to insure consistent categorization throughout the map. A basic quantitative estimator of the landscape complexity, like the Shannon entropy index, can and should be computed for every cell. Specific quantitative metrics of spatial pattern can be also applied. They should be selected on a case-by-case basis, since many indexes are class-specific and can be useful only if proper classes are identified. Computing heterogeneity indexes, as well as reporting the composition of each cell, can be systematically performed in a GIS.

The overall quality of each cell can, as a first approximation, be categorized in qualitative classes using a linguistic scale. Like GLC2000, we propose to use five classes: excellent, very good, good, moderate, unacceptable. As with qualitative labeling of heterogeneity, a catalogue of representative cases should be provided in order to ensure consistency. The labeling of overall quality, once performed for all the cells, allows for a synthetic spatial representation of the quality of the product.

This information is stored in a database, as illustrated in Figure 6.

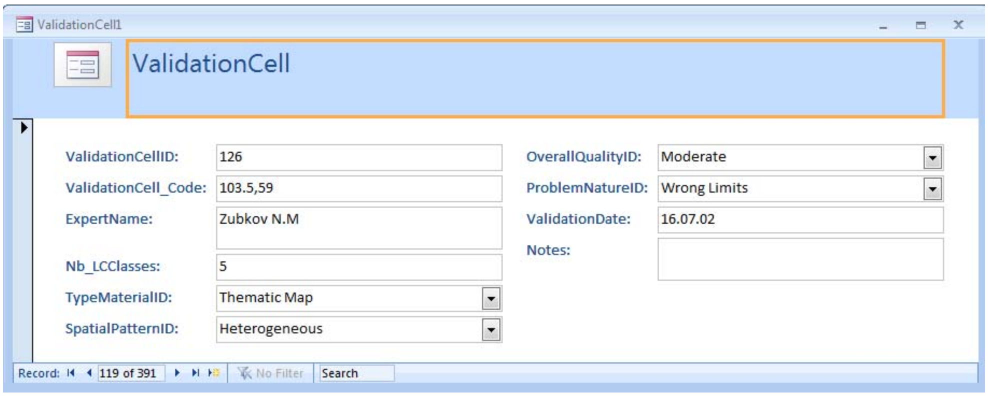

Figure 6. Form used for the systematic quality control

2.3. Nature of the problems

Ascertaining the nature of the errors occurring in the cell is of primary importance. Statistical accuracy assessment merges in the category “error” many different cases that quality control can easily document. Such information can be profitably used for improving the map during the updating phase. The main cases that can be found in global products are the following:

- The delineation of a land cover feature is accurate, but the label is wrong. In this case, the type of confusion must be specified in order to derive a thematic “distance” between the right and the wrong labels. It is, for example, generally more problematic to classify tropical forests as grasslands than to classify woodlands as savannas.
- The proportions of labels present in the cell are generally correct, but the delineation of the various features is wrong. If this case is the most frequent, it means that the spatial resolution (and eventually the pre-processing steps) precludes any accurate delineation of land cover features. The first global land cover products derived from AVHRR suffered from limitations, such as geolocation. The extreme case of this category occurs when no clear structures appear on the map. The land cover map then corresponds more to a climatic stratification.
- One important land cover feature is missing in the map or a feature is mapped while it is not present in the field. This is a particular case combining a wrong label and an inaccurate delineation of the land cover features. For example, it happens when specific features are derived from erroneous ancillary data, like planned infrastructures never actually built (dams).

2.4. Presentation of the results

Once all the cells have been visited, and the various fields stored in a database, the results can be presented in a tabular manner or on a map (Figure 7). It is possible to investigate the influence of the parameters (heterogeneity, dominant class) on the quality of the land cover map. Some of the interactions that can be investigated are:

- Map quality *vs*. land cover classes: Is the quality of the map uniform among the different land cover classes?
- Map quality *vs*. landscape diversity and fragmentation: Is the quality of the map the same in simple and in complex landscapes?

- Map quality *vs*. agreement with other global land cover maps: Are the errors mainly located in the areas of poor agreement with other maps?
- Land cover classes *vs*. type of error: Do land cover classes suffer always from the same type of error?

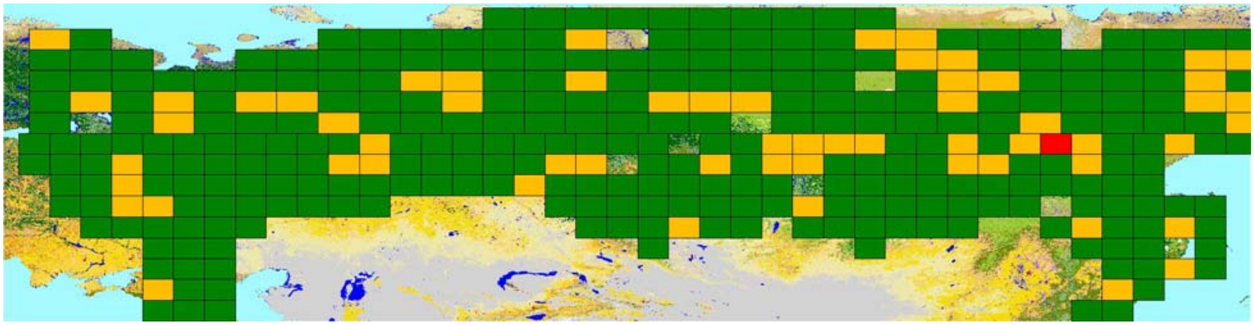

Figure 7. Map of errors (example taken from GLC2000 in Eurasia)

3. Independent statistical product validation

3.1. Building a validation database

The reference data collection can only rely on already existing expertise and reference data (including imagery) available over representative places of the world. One key element of the validation process is to build a reference dataset for the year 2010. It will be achieved through the collection of a fine spatial resolution reference dataset from the 2010 epoch. This reference dataset is then intended to be interpreted through an international network of experts in a standardized manner. This interpretation must be compatible with the CCI-LC legend (i.e. “translatable” in a transparent and objective way) in order to allow the comparison with the CCI-LC products and the computation of validation figures.

3.1.1. Experts network

Selecting appropriate experts is a key element of the validation process. The experts need to commit themselves to contribute to the quality of the CCI-LC products. Experience from previous projects (notably the GLC2000 and GlobCover projects) has shown that the most efficient way to obtain this commitment – and hence, the required information from these experts – is to invite the experts to visit the premises of one of the CCI-LC team members in Europe. This approach overcomes misunderstanding in the needs of the CCI-LC project, in particular the ambiguity in the interpretation of findings.

We expect a total duration of one week work with each expert. All necessary datasets and infrastructure (hardware / software) will be put at the disposal of the experts in order to comply with the requirements of the project.

The selection / involvement of experts will be based on the following criteria:

- ***Recognized expertise on land cover over large areas***

  It is clear that we seek people with some technical and scientific knowledge. Their ability to assess the land cover should be undisputed and if they assign a certain land cover type to a sample point we must be confident that it is trustworthy. Errors can never be completely avoided, but the experts must more or less guarantee the highest possible quality validation set that we can obtain within the context of this project. On top of this, they need to have this expertise over a substantial area.

- ***Familiarity with interpretation of remote sensing imagery***

  Some experts on land cover may have built their knowledge on the basis of extensive field surveys, but lack the experience of using remote sensing imagery. This could play a role in deciding which expert will be selected in the end. Ultimately the knowledge of the expert on the vegetation / land cover is the main criterion. Yet, we do realise that time can be a factor. If the experts need a substantial part of the available time to make themselves familiar with the data, the effectiveness of their contribution is affected. Therefore we prefer to choose experts who already know what kind of information they can expect to find in satellite images and who know how satellite data can be used.

- ***Commitment to perform the interpretation***

The experts need to commit themselves to the CCI-LC product in two different ways. The first one is obvious in the sense that their agreement to participate is not without obligations. If they agree to join the experts' network, then they will have to allocate some time for the validation at some stage of the project. It should be clear to everyone involved that this is a direct consequence of the agreement. The second commitment is less obvious but perhaps equally important. They will have to comply with the methods that the CCI-LC project team has designed for the validation, no matter what traditions and / or ideas they normally work with. In order to have a consistent dataset, this has to be a well-understood and accepted principle.

- ***Complementarities with the other experts***

The ultimate goal is to identify a group of experts that brings in complementary knowledge. For example, there is no point in looking for a 4th or 5th expert for South America, if we still lack experts for Africa or Asia. Some overlap in expertise can be beneficiary to the project as this may give the opportunity to check the consistency in interpretations between experts, but the main goal if to get a group with complementary expertise.

For the interpretation of 13000 plots (2600 primary sample units × 5 secondary sample units), the group / network should be optimally composed of at least 25 experts. The list of international experts that will compose the international network of the CCI-LC products will be defined during the course of the project as it is mainly based on willingness from the experts. The willingness of the experts to participate to the project can not be guarantied in anticipation as their benefits are rather limited (in particular compared to the expected benefits of ESA and of the consortium partners). However, the group of experts is expected to be based on networks previously or currently used by the Université catholique de Louvain (UCL) and the European Joint Research Centre (JRC) through their validation activities. UCL hosted a total of six 5-day validation workshops organized within the GlobCover 2005 project. The experts' network involved in this project (and / or in the following GlobCover 2009 project) is presented in Table 3.

Region	**Experts**	**Institution**
Africa	GÉRARD Bruno	ICRISAT
	KIBAMBE Jean-Paul	Université catholique de Louvain
	MAYAUX Philippe	Global Environment Monitoring unit – Joint Research Centre
	NONGUIERMA André	Centre Agrhymet – Niger / Economic Commission of Africa
Europe	DE WIT Allard	Alterra - (Pelcom) - Netherlands
	HAZEU Gerard	Alterra - (Pelcom) - Netherlands
	JAFFRAIN Gabriel	ETC-LUSI Technical Team
	MÜCHER Sander	Alterra - (Pelcom) – Netherlands
Russia	BARTALEV Sergey	Space Research Institute (IKI), Russian Academy of Science
Asia	HEINIMANN Andreas	National Centre of Competence in Research North-South Centre for Development and Environment (CDE)
	LIN Huang	Institute of Geographical Sciences and Natural Resources Research Chinese Academy of Sciences
	STIBIG Hans-Jürgen	Global Environment Monitoring unit – Joint Research Centre

North and Central America	GIRI Chandra	United States Geological Servey – EROS Data Center
	LATIFOVIC RASIM	CANADA CENTRE FOR REMOTE SENSING – OTTAWA – CANADA
South America	DI BELLA Carlos	Instituto Nacional de Tecnología Agropecuaria – Argentina
	GOND Valéry	CIRAD-Guyane – Université Laval
	SHIMABUKUO Yosio	INPE
Australia	CACETTA Peter	Commonwealth Scientific and Industrial Research Organisation – Australia

Table 3. Name and affiliation of the international land cover experts involved in the GlobCover projects

In addition, JRC scientists have collaborated with remote sensing experts from tropical countries in order to estimate forest cover changes at European and pan-tropical level. In this context, the JRC organized several workshops for the validation of the tree cover maps in 2009 and 2010. JRC scientists and Food and Agriculture Organization of the United Nations (FAO) officers collaborated with more than one hundred remote sensing forestry experts from tropical countries. If a need to complement the experts from the GlobCover networks is identified, the JRC can help to identify appropriate experts, in particular for the tropical continents. Complementary experts from Africa networks or TropForest project investigators could be contacted by JRC continental focal points for potential participation to the interpretation phase of the CCI-LC validation. The continental focal points of the JRC are indicated in Table 4 by continent.

Region or country	**Continental experts**	**Institution**
Africa	MAYAUX Philippe	Global Environment Monitoring unit – Joint Research Centre
Southeast Asia	STIBIG Hans-Jürgen	Global Environment Monitoring unit – Joint Research Centre
South and Central America	EVA Hugh	Strategy unit – Joint Research Centre

Table 4. Name and affiliation of the international land cover experts involved in the GlobCover projects

3.1.2. Reference data sources collection

The collection of ground information is considered as the best option to support the validation of remote sensing products in general and of the CCI-LC maps in particular. Normally, this is performed by carrying out field surveys. For global land cover products (such as the CCI-LC maps), a field survey over thousands of plots of large size (c. 100 ha) would be too costly due to the amount of man power and logistic effort needed to organize field visits to remote areas with difficult access. This makes the collection of ground truth data on the ground not feasible for a large number of plots distributed all over the globe.

However surrogate to “ground truth” can be obtained from existing “reference data sources” interpreted by experts. These reference data sources will be made of several types of datasets:

- Moderate and high spatial resolution imagery;
- Google Earth facilities;
- Multi-temporal Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI) profiles derived from SPOT-VGT time series.

These three kinds of dataset are described in more details in the following sections (3.1.2.1 and 3.1.2.2) and a summary of all data sources that will constitute the reference database in this validation exercise is provided in section 3.1.2.3.

3.1.2.1 Existing moderate and high spatial resolution imagery datasets

With regard to moderate and high spatial resolution imagery, the CCI-LC project intends to use existing datasets as much as possible, in particular to optimize the limited resources, as the CCI-LC project has dedicated very limited resources for the interpretation of reference satellite data, and as three epochs at global scale have to be considered (2000, 2005 and 2010). In that respect, the availability of historical data (2000 and 2005) is also a challenge for the project.

Establishing operational validation activities requires, as a prerequisite, the availability of a large amount of good quality reference imagery at moderate spatial resolution (20m x 20m or 30m x 30 m), such as Système pour l'Observation de la Terre (SPOT) and Landsat TM. The use of moderate resolution satellite imagery for large scale historical assessments has been boosted by the recent free availability of the Landsat Global Land Survey Database through the United States Geological Survey (USGS). Nearly complete global coverage from the Landsat satellites is now available at no cost from the Earth Resources Observation Systems (EROS) Data Center (EDC) of the USGS (http://glovis.usgs.gov).

This recent product, called the ***Global Land Survey*** (GLS), was derived by reprocessing the GeoCover data. The GeoCover dataset consisted of a selection of good quality, orthorectified and geodetically accurate global land dataset of Landsat TM (30 m × 30 m) and Enhanced Thematic Mapper (ETM+) (30 m × 30 m) satellite images with a global coverage. It was created from the epochs circa 1990, circa 2000 and mid-2000s at 28.5 m × 28.5 m resolution by the NASA. In addition, USGS has undertaken the effort of improving the geo-location for the best quality scenes of their extensive Landsat data archive. In this context, the GLS time indications stand, roughly, for the period of 1986 – 1993 for GLS-1990 and of 1999 – 2001 for GLS-2000. GLS-2010 is under creation. The CCI-LC project will make use of the GLS from 2000, 2005 and 2010 as main source of satellite imagery for the validation process. The GLS datasets present the advantages of standardized imagery which is commonly used for land cover mapping at fine scale, with a spatial resolution which is a good compromise between spatial precision (less than 0.1 ha) and allows easily aggregation at usual mapping scales including a SWIR channel which is important for the interpretation of a number of thematic classes.

A global systematic sampling scheme has been developed jointly by FAO (Forest Resource Assessment (FRA) 2010 Remote Sensing Survey) and the JRC (Tropical Ecosystem Environment Observations by Satellite (TREES) project) to estimate rates of deforestation at global or continental levels and at intervals of 5 to 10 years (FAO/JRC/SDSU/UCL, 2009). Time-series of Landsat TM or ETM+ data are attached to each sampling location through a quality-controlled, standardized and decentralized process. For the FAO's FRA2010 RSS exercise, the South Dakota State University (SDSU) produced a global database of multi-temporal 20 km × 20 km sample tiles for the "epochs" 1990, 2000 and 2005 extracted from the USGS GLS archives, i.e. GLS-1990, GLS-2000 and GLS-2005 (available at http://globalmonitoring.sdstate.edu/projects/fao/index.html). GLS-2010 is under construction by USGS and related full Landsat Scenes will be available from USGS web site by end of 2011. This set of moderate spatial resolution imagery (GLS-2000 and GLS-2005 extracted from SDSU plus GLS-2010 to be downloaded from USGS) will be the main source of reference data for the validation exercise.

For the portion of the sample tiles that are not available from the GLS database or have persistent cloud contamination, other Landsat imagery or alternative moderate spatial resolution remote sensing data are planned to be used. For the 20 km × 20 km sample tiles located in tropical countries (i.e. c. 4,000 tiles of the systematic sample) and over Russia (c. 1,500 tiles of the systematic sample) the JRC has selected alternative Landsat imagery for "epochs" 2000 and 2005 (Beuchle et al., in press) and is presently in the phase of selection of GLS imagery for the epoch 2010 (from GLS-2010 dataset or complementary Landsat imagery with priority to Landsat-5 TM sensor). In the current text and for reading facility, this Landsat dataset over the tropics and Russia is called the "***TREES dataset***".

For the 2010 epoch, complementary imagery at finer resolution (20 m × 20 m or finer) will be obtained through ESA Category 1 proposals submitted in the framework of the CCI-LC and TropForest 2010 projects. This imagery will be used, to replace (when cloudy or missing) or to help in the interpretation of the main dataset of Landsat imagery to be used over the sample tiles,

1. In the framework of the ***ESA agreement with Third Parties Missions*** (TPM), the CCI-LC project intends to acquire a few hundreds of SPOT-4 images (at 20 m × 20 m resolution) at dates selected as close as possible to 2010 (both from the archive and programming requests for the year 2011) over sample plots which will pre-identified as lacking or low quality from the GLS-2010 dataset.

2. In addition, the ESA category-1 project entitled "***TropForest 2010***" allowed the acquisition of high spatial resolution satellite imagery for South America and South East Asia for the 2010 epoch. For the 1˚ × 1˚ confluence points, satellite imagery from the Advanced Land Observing Satellite (ALOS) Advanced Visible and Near-Infrared Radiometer (AVNIR) 2 sensor (at 10 m × 10 m resolution) or from DEIMOS-1 sensor (at 22 m × 22 m resolution) have been be acquired (circa 85% covered by AVNIR-2 and 15% by DEIMOS-2). For the 2˚ × 2˚ confluence points, satellite imagery from the Korea Multi-Purpose Satellite (KOMPSAT) 2 (at 4 m × 4 m resolution) have been acquired (presently 60% of 2˚ × 2˚ confluence points are covered). The acquisition of this dataset is being finalized by ESA and the resulting geometrically-corrected dataset will be analyzed by JRC in the TropForest project.

A complete description of these two Cat-1 proposals is provided in the annexes of the Data Access Requirement Document (DARD) [AP-6].

3.1.2.2 Auxiliary information

The moderate spatial resolution dataset will be used in combination with other information.

First, the experts will be able to consult very high spatial resolution data available from Google Earth. However, it has to be noted that their use could be limited due to the fact that the level of details can vary from site to site.

In addition, the possibility to consult existing reliable and detailed (spatially and thematically) land cover maps could also be offered to the experts (if they consider this could be a source of valuable information for their interpretation work).

Finally, multi-temporal annual profiles of spectral indices will also be made available to the experts in order to display seasonal variations of natural vegetation. Indeed, the high temporal resolution provided by the SPOT-VGT and MERIS time series gives us the opportunity to monitor the development of vegetation over time by examining the run of the NDVI profile. For each validation sample, NDVI values will be derived every 10 days from the MERIS and SPOT-VGT 2010 time series. The averaged 10-day NDVI and NDWI profiles over a multi-year period (9 years from 2003 to2011 for MERIS and 12 years from 2000 to 2011 for SPOT-VGT) will also be provided. Figure 8 illustrates this temporal information which will be provided to the experts based on the GlobCover experience (where the NDVI annual profiles from SPOT-VGT were related to the years 2005 and 2009 instead of 2010).

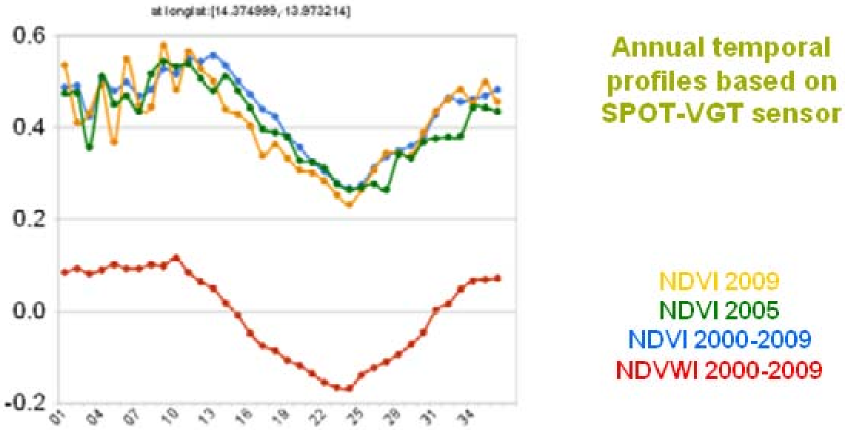

Figure 8. Examples of SPOT-VGT temporal profiles which will be provided to the experts for the validation

3.1.2.3 Summary

Table 5 lists all the dataset that will be used to build the reference database. Their technical characteristics and availability is exhaustively described in the DARD [AP-6].

Product	**Description**
Global Land Survey Database for 2000, mid-2000 and 2010	Nearly complete global coverage from the Landsat satellites available at no cost from USGS
TREES dataset	Sample units over South America and South East Asia for epoch 2010 (Landsat type imagery)
Images from the CCI-LC Cat-1 proposal	Sample units at the global scale (but mostly over Europe) for epoch 2010 (from SPOT 4 archive)
Images from the TropForest Cat-1 proposal	Sample units over South America and South East Asia for epoch 2010 (from ALOS AVNRI-2 sensor or from DEIMOS-1 sensor or from KOMPSAT-2 sensor)
Google Earth	High spatial resolution imagery freely available (for visualization) on Google Earth
NDVI profiles	10-day NDVI time profiles derived from the MERIS and SPOT-VGT dataset over 12 years (2000 to 2011)

Table 5. List of all datasets that will be used for validating the CCI-LC products

3.1.3. Existing reference datasets for cross-comparisons

Existing reference datasets at moderate resolution have been indentified from ongoing international initiatives in which UCL or JRC have been involved. These are the ***GLC2000 and GlobCover reference validation databases***. The datasets consist of two files of respectively 1,265 samples and 5,570 samples validated according to the Land Cover Classification System (LCCS) classifiers. These data will be used for the comparison with other global products in a confidence-building procedure.

3.2. Sampling

3.2.1. General requirements of the sampling frame

A sample is some part of a larger "population" (in this case, the whole CCI-LC product) which has been specially selected to represent the whole population. For a sample to be representative, it should reflect the similarities and differences found in the total population. The main objective of drawing a sample is to make inferences about the larger population from the smaller sample.

The sampling scheme will be designed with the following requirements:

- to be statistically valid for accuracy assessment of the CCI-LC products;
- to be reusable for future global products of similar type;
- to be designed before (i.e. independently) the CCI-LC product;
- to use the most recent picture of global land cover distribution (as best proxy of the actual land-cover distribution);
- to address the issue of rare classes with a strong impact on the climate system (urban areas, wetlands, etc).

Furthermore, it is expected to use such stratified sampling scheme to allow comparative assessment with other existing global land cover products.

Following these requirements, three questions will be addressed: the number of sample plots, their size and the way they are selected from the total population.

3.2.2. Number of sample plots

The sample size required for a given confidence level and a given acceptable error in the sample can be calculated from the binomial distribution (Stadelman et al., 1994):

$$n = \frac{p * q}{\left(\frac{E}{z_\alpha}\right)^2} \qquad (1)$$

where n is the number of sample plots, E is the allowable error in the sample, z_α is drawn from the normal distribution for the given level of confidence, p is the required accuracy and q is 1-p.

The allowable error stands for the error that is made when validating the product using a sampling strategy (instead of exhaustively assessing the product in any locations). For example, if the allowable error has been set at 5% and the validation process (based on a given sample) yields to a result of 77% successes, then it is safe to claim that an accuracy value between 72% and 82% would have been obtained if the whole accuracy had been validated.

The z-value (z_α) is directly related to the statistical normal distribution and a certain level of confidence. This level of confidence expresses the percentage that indicates how often a validation performed on the basis of a given sample dataset would yield a result that lies within the confidence interval. Considering the previous example (allowable error of 5% and accuracy of 77%) and adding the concept of this level of confidence (set at 95% for instance), we could say that we are 95% sure that the quality of the product falls between 72% and 82%.

Within the CCI-LC validation process, the different parameters of equation 1 will be set as follows:

p	**q**	**Confidence level (related to z_a)**
0.7	0.3	95%

Table 6. Values given to the equation 1 variable for this validation process

The validation of the GLC2000 product has yielded an overall area-weighted accuracy of 68.8% (Mayaux et al, 2006), based on the interpretation 1265 samples, and while the accuracy of GlobCover is 73% based on 3167 samples. It explains why 70% was considered as a realistic estimate of the accuracy of the CCI-LC products. It has to be mentioned that, paradoxically, the required sample size gets smaller with increasing *p* values. In other terms, if one puts the estimated accuracy of the map higher than it actually is, the required sample size gets smaller and it can offer the possibility to manipulate the results (Schouten et al., 2006). It would therefore seem wiser to use minimum target accuracy values (in this case 70%) to define the sample size.

Using such parameterization, this leaves the allowable error of the statistical estimate (statistical "precision") as unique variable. In the GlobCover Validation Plan (Schouten et al., 2006), a table demonstrated how the sample size is affected by various levels of this error (Table 7).

Precision	0.01	0.03	0.05	0.10
Sample size	8061	897	323	81

Table 7. Sample size with targeted accuracy of product 70% and a confidence interval of 95% for four different values of statistical precision (from Schouten et al., 2006)

The limited resources dedicated to the validation protocol also constrain the number of samples by a second parameter: the time for interpreting the data. The GlobCover validation exercise has shown that experts can interpret between 30 and 50 sample plots per day. If the validation workshop lasts one week, a single expert could generate between 150 to 250 validated plots. The number of experts that will be involved would be around 20 experts. Together they could generate around between 3000 and 5000 points.

To generate a precision of 0.03, around 900 sample plots would be needed for validating the global CCI-LC product. However, in order to be consistent with the previous exercises, we will select 2600 primary sample Units (PSU).

3.2.3. Size of sample plots

The elementary unit in the CCI-LC product is a pixel with a spatial resolution ranging from 300m × 300 m to 1 km × 1 km (depending on the sensor used to generate the product). However, the same unit placed over higher spatial resolution imagery (such as those used to build the reference database (see 3.1.2.1)) may represent something quite different. It is thus necessary to distinguish the minimum mapping unit (MMU) and the observational unit.

The MMU is actually a cartographic term. It defines what still can be depicted on a map, from a user's perspective. The introduction of a MMU is justified for the following reasons:

- Geo-location accuracy of the information. The absolute positional accuracy of the CCI-LC product is targeted to 1/3 pixel dimension;
- Image mosaics of MERIS scenes may result in radiometric information coming from a few adjacent pixels;

- Users are usually not interested in single pixels but in information at landscape level (corresponding to features that can be observed on the ground);
- Land cover of single pixels of the CCI-LC products will often be composed of mixed land cover classes.

Due to this last point (the fact that single pixels will often cover several land cover types), the observational unit (size of sample plots) has to be introduced next to the MMU. The observational unit is larger than the MMU and thus gives more weight to the neighbourhood of the pixel. The main reason for assigning more weight to the neighbourhood of the pixel is that it is not realistic for an expert to interpret land cover class of single MERIS-size pixels. The expert needs sufficient information (pixels in this case) to decide which land cover type is the dominant one.

For the validation of the GLC2000 product, blocks of 3 × 3 pixels at 1-km × 1-km resolution were analyzed (Mayaux et al., 2006), equalling to 900-ha surface. For GlobCover, blocks of 5 × 5 pixels at 300-m × 300-m resolution were analyzed (Bicheron et al., 2008; Bontemps et al., 2010), equalling to 225-ha surface. For the present exercise, it is envisaged to interpret plots of 3 × 3 pixels at 300-m × 300-m spatial resolution (corresponding to 3 × 3 MERIS full resolution pixels or to 81 ha).

3.2.4. Sampling design

3.2.4.1 General objectives

To satisfy requirements of design-based inference, the sampling design should be a probability sampling design, and the estimators should be constructed following the principle of a consistent estimation (Strahler et al., 2006).

Design-based inference is predicated on implementing a probability sampling design. The definition of probability sampling focuses on inclusion probabilities, where an inclusion probability is defined as the probability that a particular pixel will be chosen for the sample. Probability sampling requires these inclusion probabilities to be known for all pixels selected in the sample, and nonzero for all pixels in the population (the entire region mapped). Many probability sampling designs have been developed, including familiar designs such as simple random, systematic, stratified random and one- and two-stage cluster sampling. Adherence to probability sampling imposes some constraints on the sampling protocol to ensure that the inclusion probabilities can be determined.

As that data and resources limitations / availability must be optimised, the ***systematic sampling*** of the TREES dataset combined with a ***two-stage stratified clustered sampling*** will be considered because it is generally recognised as the most efficient sampling strategy (Strahler et al., 2006). In a first stage, the stratified random sampling will allow to select the ***Primary Sampling Units***, while the ***Secondary Sampling Units*** will be systematically distributed within the primary ones.

3.2.4.2 Selection of the Primary Sampling Units

Primary Sampling Units (PSU) correspond to a 20-km x 20-km boxes where moderate and/or high spatial resolution imagery (see 3.1.2.1) will be acquired. PSU will be selected using the TREES ***systematic*** sampling and a two-stage ***stratified*** clustered sampling.

In order to have an equal probability of selection for all potential PSU in the product, the projection of the map from which the PSU are selected should ideally be an equal-area projection. Because we take benefit of the ***systematic*** sample of the TREES dataset, we have progressively removed samples by latitude in order to keep this equal probability. The FAO FRA dataset indeed had reduced the sampling density by a factor 2 only at 60° of latitude (Figure 9).

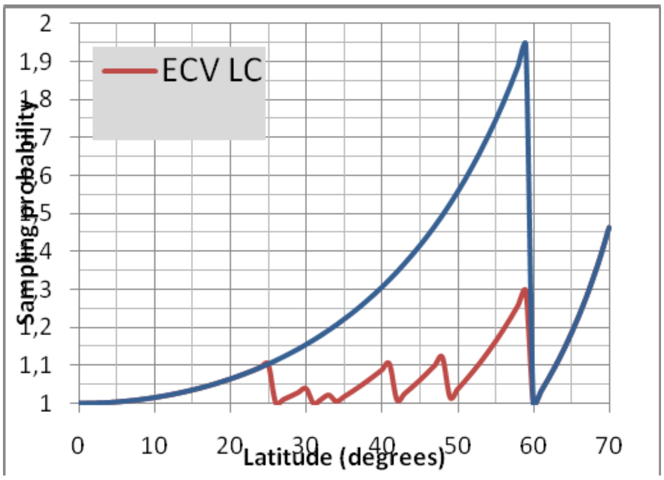

Figure 9. Sampling probability according the latitude of the samples in two cases: (i) sub-sampling at 60° N (FAO-FRA), (ii) progressive sub-sampling at 25°, 30°, 35°, 40°, 48° and 60°.

In a second step, the stratification of the entire population (i.e. of the whole product) is performed in order to reduce the number of samples necessary to get the required precision of the accuracy estimates, or alternatively to increase the precision for a given number of samples. Previous exercises have used the landscape fragmentation and composition for stratifying the population (Mayaux et al., 2006). The assumption is that homogeneous landscapes generate less classification errors, while some land cover classes are less subject to errors due to typical spectral response and good observation conditions. As a consequence, the expected accuracy of these classes is very high. For instance, in the GLC2000 validation, three land cover classes ("Bare areas", "Snow and Ice", "Water Bodies") had both a user's and a producer's accuracy higher than 95% (Mayaux et al. 2006).

In the CCI-LC validation protocol, two strata are defined (Figure 10).

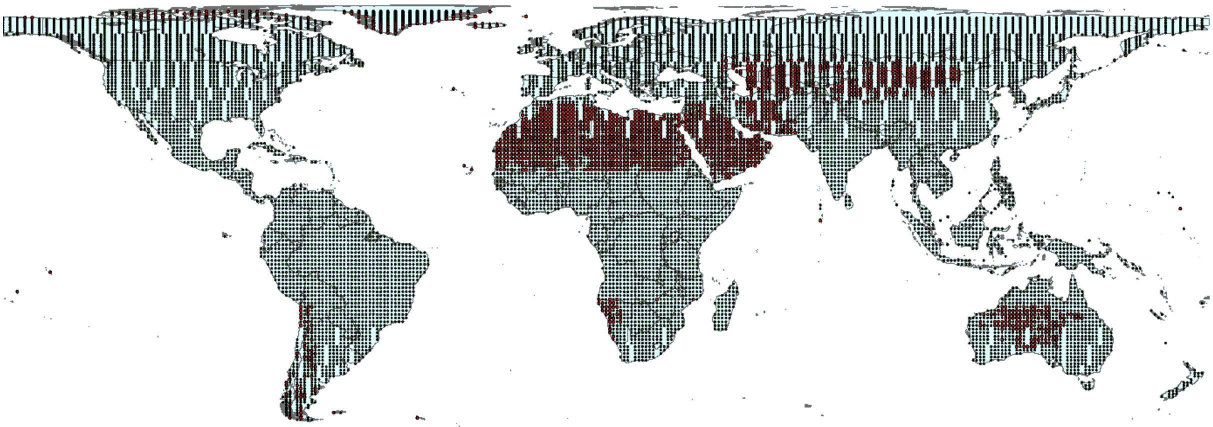

Figure 10. Total population considered in the systematic sampling procedure. Red dots correspond to stratum 1; black dots correspond to stratum 2.

The 2600 primary sample Units (PSU) are selected through a random stratified selection within the two strata: desert stratum with 100 PSUs and non desert stratum with 2500 PSUs. The final selection of the 2600 PSUs is illustrated in Figure 11.

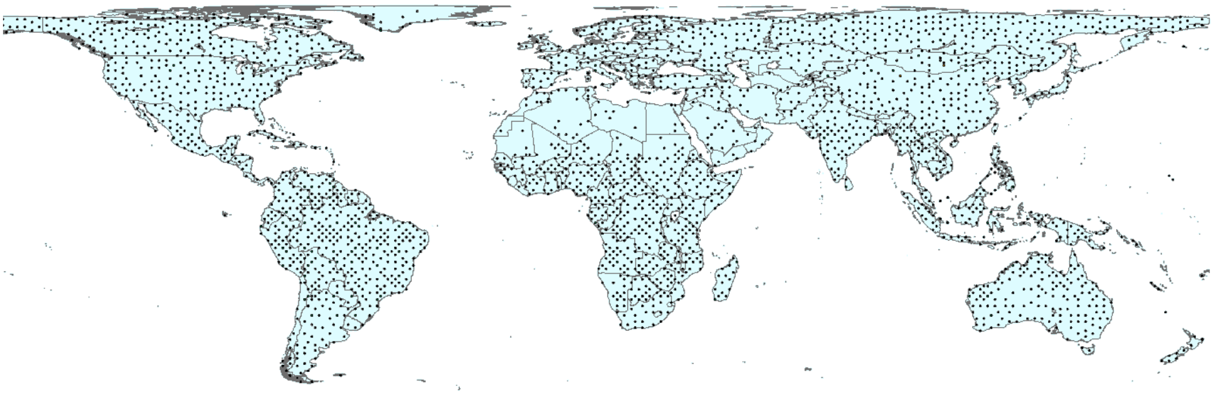

Figure 11. Selected sample frame displaying the 2600 PSUs

Each of these PSUs will be covered by a Landsat scene over the three periods (2010, 2005 and 2000), which are obtained from the GLS datasets (see section 3.1.2.1). For the 2010 epoch, additional sets of imagery at higher spatial resolution (20 m × 20 m or finer) will be obtained: (i) hundreds of SPOT-4 images will be acquired over Europe and Africa in the framework of the ESA agreement with Third Parties Missions and (ii) satellite imagery from ALOS-AVNIR 2, DEIMOS-1 or KOMPSAT 2 sensors will be acquired over South America and South East Asia thanks to an ESA category-1 project entitled “TropForest 2010”. Figure 12, Figure 13 and Figure 14 show the spatial distribution of these complementary datasets.

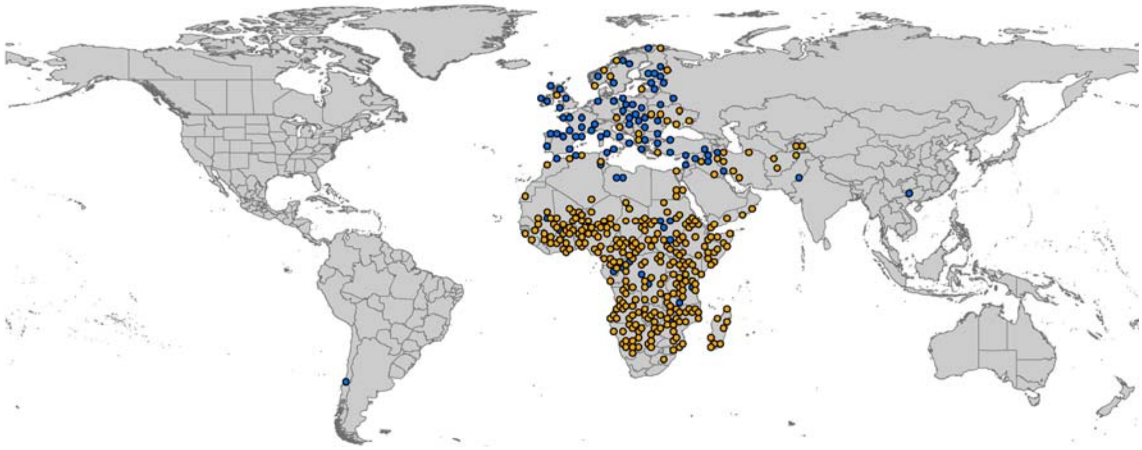

Figure 12. Spatial representation of SPOT-4 images collection (blue points referring to archive images while orange ones indicating new programming requests)

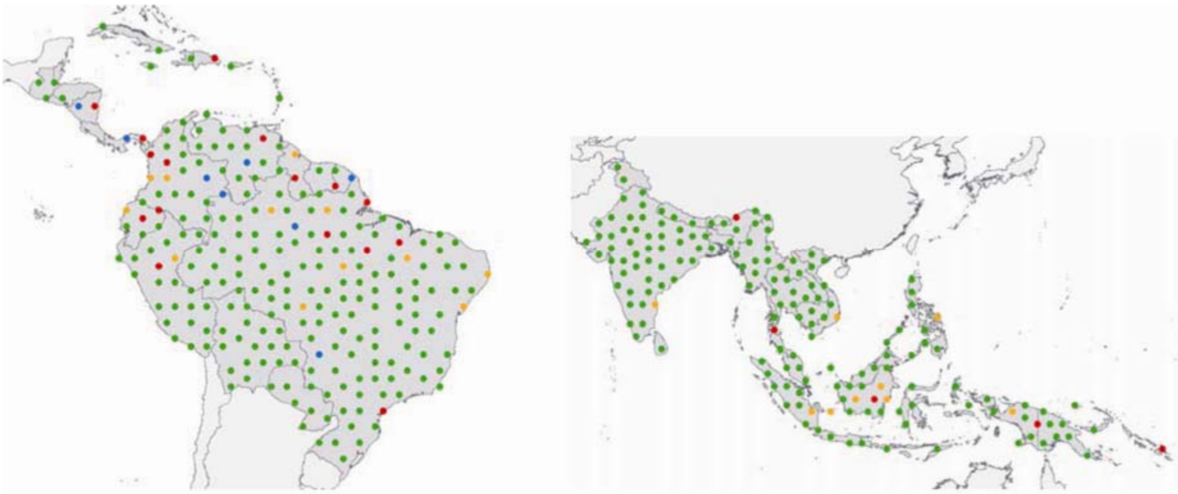

Figure 13. Availability of ALOS AVNIR-2 imagery (green points) complemented by DEIMOS-1 imagery (blue points) from TropForest project for CCI-LC PSUs: c. 95% of PSUs located at 1°×1° confluence points are covered

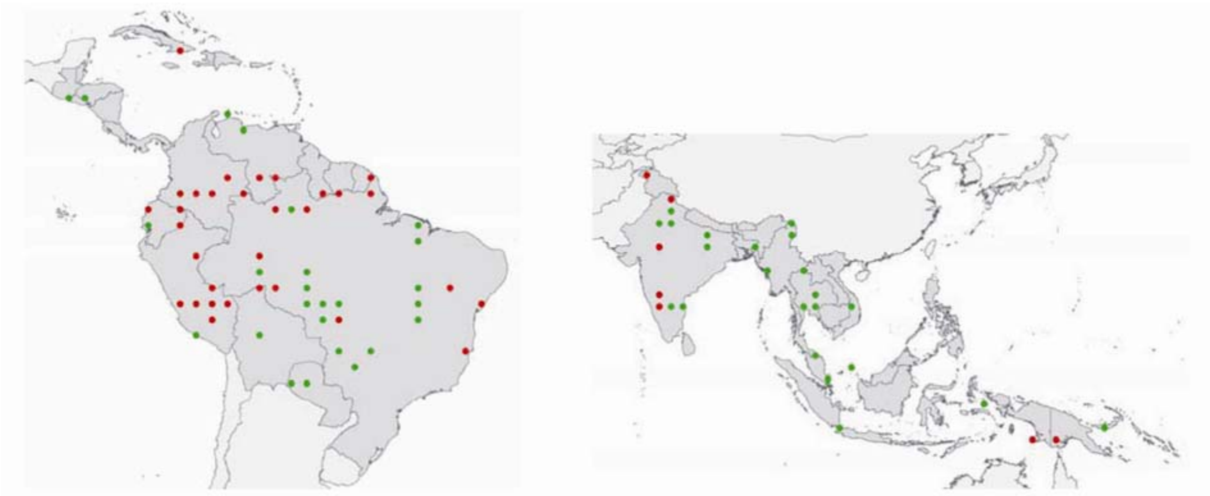

Figure 14. Availability of KOMPSAT-2 imagery (green points) from TropForest project for CCI-LC PSUs: 55.3% located at 2˚×2˚ confluence points in South America and Southeast Asia are covered (47 images over a total of 85 PSUs)

Table 8 summarizes, for the three epochs, the importance of the different validation datasets in terms of PSUs coverage (2600 PSUs in total).

Epoch	Validation dataset available	Area covered	% of PSUs covered
2010	GLS-2010 dataset	Global	~ 90 %
	SPOT 4 images	Mainly Europe and Africa	~ 20 %
	TropForest dataset (AVNIR-2 & DEIMOS)	Southeast Asia and Latin America	~ 15 %
	TropForest dataset (KOMPSAT-2)	Southeast Asia and Latin America	~ 2 % (c. 50 images)

2005	GLS-2005 dataset	Global	~ 90 %
	SPOT 4 images	Mainly Europe	~ 5-10 %
2000	GLS-2000 dataset	Global	~ 90%
	SPOT 4 images	Mainly Africa-Europe	~ 5-10 %

Table 8. Estimated percentages of PSU's covered by the different validation dataset for the three epoch

3.2.4.3 Selection of Secondary Sampling Units

A medium spatial resolution image will thus be acquired for each PSU. However, only a small part of each image will be analyzed. It is planned to interpret 5 "sample plots" within each image, i.e. to define 5 Secondary Sampling Unit (SSU) for each PSU. For the CCI-LC validation protocol, the 5 "sample plots" or SSU will be located at (i) the centre of each 20-km × 20-km box and (ii) at a distance of 4-km × 4-km from the centre of each box (Figure 15).

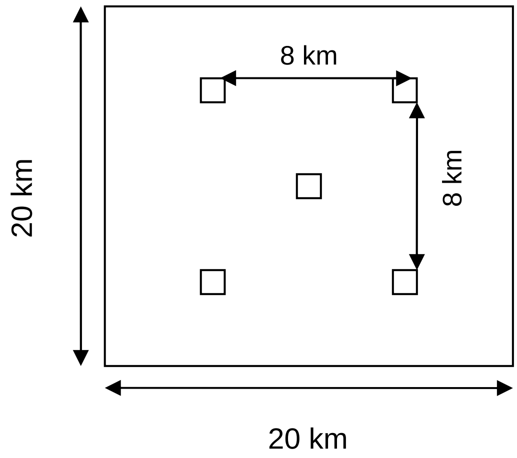

Figure 15. Selection of SSUs (900 m × 900m size) within a PSU

This nested sampling approach (i.e. the multiplication of the number of sample sites compared with the option of using only one single sample plot per box) provides a larger number of sample sites and therefore leads to lower standard error of accuracy estimates.

3.3. Image interpretation by local experts

3.3.1. General purpose

Previous validation exercises have massively used remote sensing specialists with local expertise to provide the land cover interpretation of the medium resolution imagery. This approach has proven to be the most efficient, but we recognize that this process of interpretation by experts has its own limitations leading to some uncertainties. Uncertainties are particularly caused by differences in interpretation skills and field knowledge, by limitations of the information contained in the satellite imagery, by difficulties in interpreting complex landscapes and by required familiarity and comprehension of the necessary land cover legend developed for a global map.

The interpretation of satellite imagery will be helped through the application of image pre-processing steps (e.g. segmentation process) which are intended to harmonize spectrally and correct the imagery from haze effects, and then allow for a more consistent identification and interpretation of land cover objects. Such strategy which combines advanced image processing techniques with a simple validation user interface is used by JRC and FAO in their respective TREES project and FRA-2010 Remote Sensing Survey.

The legend and the interface to be developed and used during the validation phase will take benefit from the developments made for the validation phases of the GLC2000 and GlobCover projects. In particular the interpretation legend will be developed according to the LCCS classifiers; the interface will combine vegetation profiles, ancillary information and Google Earth imagery (or similar quality imagery if existing and available).

3.3.2. Preparation of moderate and/or high spatial resolution imagery

In order to facilitate the validation, it is important to prepare all the data that the experts will access to perform their interpretation. CCI-LC products will be produced as raster products with a cell size of 300m ×300m or 1 km × 1km.

Over each PSU, the moderate and/or high spatial resolution satellite image which is part of the reference validation database (see 3.1.2.1) needs to be first radiometrically corrected and harmonized. The imagery will be pre-processed with a few automatic steps: orthorectification based on orbital parameters, extraction of 20 km × 20 km boxes and radiometric calibration.

The selection of a suitable approach for image interpretation depends on operational requirements. For this exercise, the possibility for rapid image interpretations is the main criterion. Indeed, on the one hand, the number of PSU is large (2600) and on the other hand, the number of experts and the allocated time are limited. For this reason, an object-oriented approach coupled with a pre-labeling procedure is envisaged with the following steps:

- Selection of an appropriate MMU for the automatic segmentation;
- Image segmentation;
- Collection of representative spectral signatures for land cover labeling;
- Automatic classification of segments with pre-assignment of land cover labels;
- Image interpretation by the expert consisting in a visual verification and final assignment of land cover labels.

A MMU of 5 hectares (equivalent to circa 250-m × 200-m size polygons) is considered for the automatic segmentation and interpretation of the satellite imagery. Within the TREES project processing chain, a finer “detection unit level” at about 1 hectare is used in a first automated segmentation and labeling step before aggregation to 5-ha objects for the interpretation phase. These criteria are based on initial tests performed by JRC on datasets for Papua New Guinea. The automatic operational land cover legend of the JRC automated chain includes a few land-cover classes: Tree cover, mosaics of trees and other land cover, other wooded cover, other land cover, water, clouds & no data. This can be used as initial land cover information which has to be refined for the “other land Cover” class. The experts will validate all the automatically labeled polygons of MMU 5 hectares, included in each SSU, i.e. in average 10 to 15 per SSU.

3.3.3. Image interpretation protocol

The sampled sites (SSU) need to be interpreted – i.e. labeled – by the expert. This section deals with the protocol that will be enforced for the labeling by the experts during validation.

Any object overlapping even partly the SSU box will have to be labelled. It may vary from 1 to 10 objects to label but in most cases, the sited are expected to be homogeneous. In addition, the labeling procedure should allow selecting several objects or all objects together to label them at once.

The process of samples interpretation can be ambiguous for three reasons: (i) inadequate quality of the reference imagery; (ii) heterogeneity of the landscape and (iii) limited knowledge by the expert.

- If the expert cannot derive the land cover because of poor quality of reference imagery, the sample has to be skipped. The expert must specify that no land cover class have been assigned to the sample because of insufficient quality of the data.
- If the landscape is heterogeneous, the expert has to explicitly specify that the landscape is complex. The segmentation procedure tackles this heterogeneity issue and will generate many small polygons in heterogeneous landscapes. But contrary to the previous exercises, the interpretation of the polygons can increase the interpretation quality in such environments. The reporting of the heterogeneous landscapes is still problematic since it will be difficult to build a 1 to 1 relationship between the map and the reference.
- If the expert is not sure how to interpret the sample (SSU), he / she can indicate a lower level of certainty. When there is serious doubt about the exact land cover class, the expert needs to indicate the classes from which the expert cannot choose with certainty. It is clear that more attributes than the dominant land cover classes are relevant, especially for the analysis of observed discrepancies between classification and expert's labeling.

An optimum attribute table must be designed for the validation. The following fields will be included:

- Name expert
- Date and time
- Dominant Land cover class (with classifiers)
- Level of certainty (default is an empty field)
- Comment (e.g. for indicating why the labelling was not successful, or to give the local name used for the concerned land cover type)

3.3.4. Graphical interface for image interpretation

For the interpretation of land cover state classes over the SSU, the experts will use a graphical interface to interpret Landsat-type imagery (Figure 16).

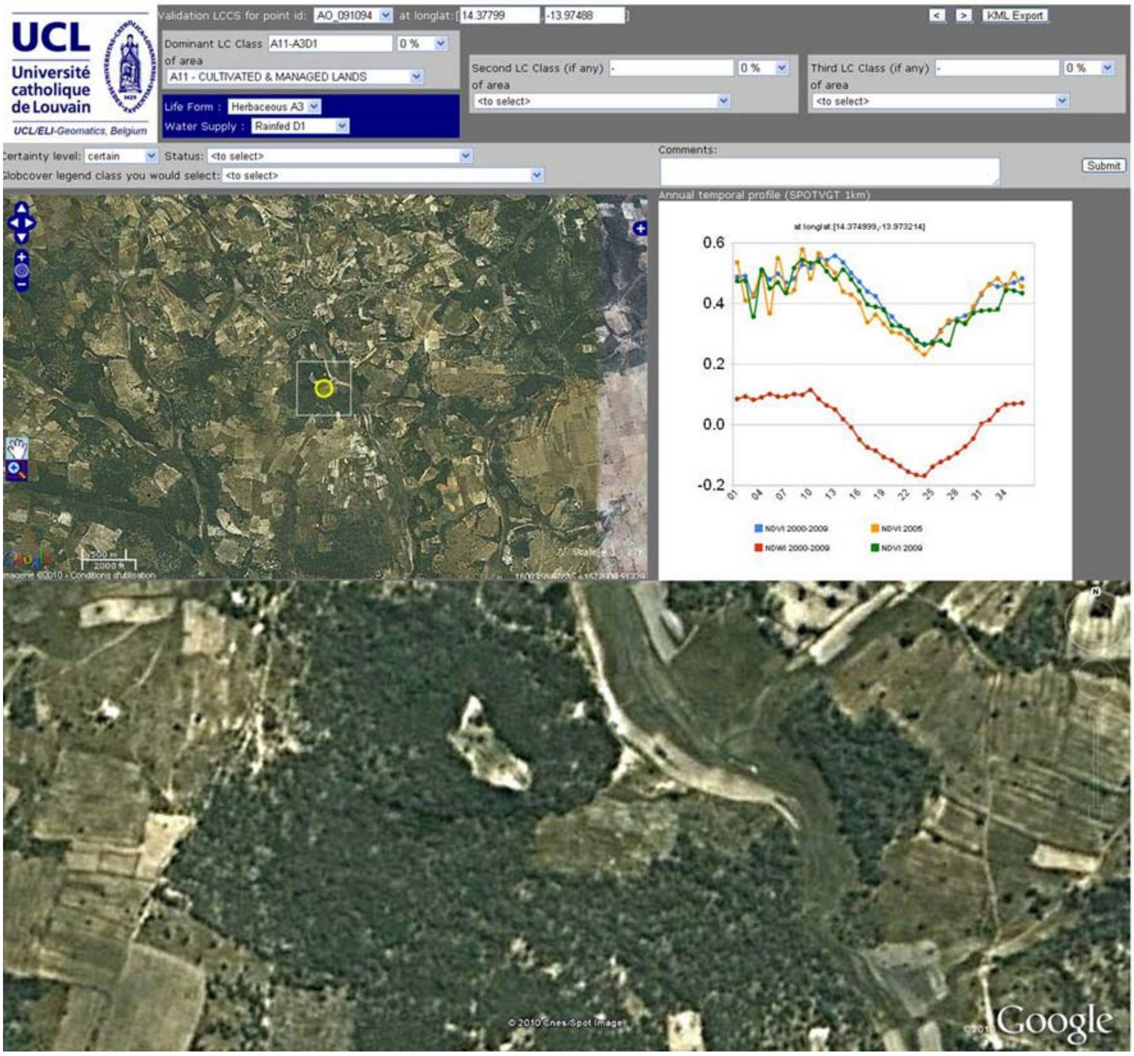

Figure 16. User interface for interpretation of sample plots available at UCL

This user interface is described in more detail in a manual which will be made available to the experts. The tool aims to be the best compromise between flexibility for the experts and productivity. One element of flexibility we have preserved is the use of LCCS classifiers and attributes rather than the use of legend entries. This does require a good knowledge of the LCCS system, and a substantial part of the manual is therefore dedicated to this system. The graphical tool will be adapted to the new protocol that includes the polygons. On the other hand, it must reduce the interpretation time by allowing grouping of polygons under the same label.

3.3.5. Quality Control Process

During the phase of SSU interpretation by experts, quality control procedures will be introduced including repetition of SSU interpretations by the same expert, comparison between interpretations from different experts (in particular comparing regional versus global experts), and analysis of very fine spatial resolution for a sub-sample set of sites.

- The repetition of SSU interpretations by the same expert and a subsequent comparison between interpretations from different experts will lead to a consistency assessment. This quantitative assessment will provide the level of relative uncertainty related to the human interpretation expertise and quantify its shortcomings.

- The analysis of very fine spatial resolution (5-m × 5-m to 10-m × 10-m) images for a sub-sample set of sites should in theory allow quantifying the absolute level of accuracy of the final products. However, it will be limited by the availability of this kind of imagery. Historical data at very high spatial resolution usually exists only for very limited zones and, when existing, are generally not freely available. Satellite imagery at such very fine resolution will be available for the year 2010 over South America and Southeast Asia through the TropForest 2010 project (from KOMPSAT-2 satellite at 5 m × 5 m resolution or from AVNIR-2 sensor onboard ALOS satellite at 10 m × 10 m resolution). If such imagery is available at JRC at the time of the interpretation phase by experts, a test case will be produced for these two continents.

In addition, a consistency check of the interpretation will be performed on different levels (LCCS classifiers, different validation classes, comparison with regional products) to determine potentially suspicious interpretations. Particular attention will be given to the heterogeneity of the landscape which has been found to be an important driver of interpretation ambiguities (Mayaux et al., 2006). Multi-scale interpretation options have been proposed to reduce the impact of such issues.

3.3.6. Validation dataset public availability

The validation dataset built and used in the CCI-LC project are planned to be made available to public within the framework of the CEOS working group on validation and calibration (WGCV). Table 9 lists those datasets and provides their specifications.

Validation dataset	**Specification**
GlobCover 2005	A shapefile (geographic Lat/Long representation based on the WGS84 ellipsoid) providing the 5*5 MERIS pixels blocks used as observational units for the validation exercise (each block being associated with a unique identifier) An access database providing the LCCS description for each observational unit (based on the unique identifier), each unit being described by 1 to 3 land cover type Only blocks interpreted as certain by the experts are provided
GlobCover 2009	A shapefile (geographic Lat/Long representation based on the WGS84 ellipsoid) providing the 5*5 MERIS pixels blocks used as observational units for the validation exercise (each block being associated with a unique identifier) An access database providing the LCCS description for each observational unit (based on the unique identifier), each unit being described by 1 to 3 land cover type Only blocks interpreted as certain by the experts are provided
CCI-LC	A shapefile (geographic Lat/Long representation based on the WGS84 ellipsoid) providing the PSUs used in the validation exercise (each block being associated with a unique identifier) An access database providing the proportions of the different land cover types present in each PSU/ (based on the unique identifier) Only PSUs interpreted as certain by the experts are provided

Table 9. List and specification of the validation datasets that will be available to the public within CEOS-WGCV

It should be noted that a phase of validation databases consolidation and verification will be necessary between the databases production and their release.

3.4. Reporting

3.4.1. Production of a validation report

A report will describe in detail the validation procedures, how they were implemented and the results. This validation report will analyze in detail the various parameters describing the accuracy of the map: contingency matrix, user's and producer's accuracy, Kappa statistics, and area statistics. The experience of the GLC2000 project (Mayaux et al., 2006) and the GlobCover project (Bicheron et al., 2008; Bontemps et al., 2010) will be used to provide a detailed validation report compatible with the state-of-the-art methodologies in this field.

Accuracies will be derived by comparing the CCI-LC products with the results of the independent interpretations of the SSU by the experts. Counting the frequency of matching results gives the eventual figures. To demonstrate this counting process, compare the confusion matrix in next table (Figure 17). For calculating the overall accuracy of the product, each class is weighted by the area it represents in the map.

Actual Class (↓)

Predicted Class (↓)	A	B	C	D	Σ
A	n_{AA}	n_{AB}	n_{AC}	n_{AD}	n_{A+}
B	n_{BA}	n_{BB}	n_{BC}	n_{BD}	n_{B+}
C	n_{CA}	n_{CB}	n_{C}	n_{C}	n_{C+}
D	n_{DA}	n_{DB}	n_{D}	n_{D}	n_{D+}
Σ	n_{+A}	n_{+B}	n_{+C}	n_{+D}	n

$$\text{Overall accuracy} = \frac{\sum_{k=1}^{q} n_{kk}}{n} \times 100\% \qquad \text{User's accuracy} = \frac{n_{ii}}{n_{i+}} \times 100\%$$

q = Number of classes

$$\text{Producer's accuracy} = \frac{n_{ii}}{n_{+i}} \times 100\%$$

Figure 17. Layout of a typical confusion or error matrix, showing computation of user's and producer's accuracies (from Strahler et al. 2006). The fields in grey mark the correspondence between classification and labelling by the expert.

A shortcoming of the overall accuracy is that it does not account for chance agreement. A complete random classification would also generate a certain level of accuracy. The overall accuracy does not tell how much better the obtained result is compared to the random classification. The Cohen's Kappa is an index used to make this comparison. It expresses the proportionate reduction in error generated by a classification process, compared with the error of a completely random classification. The closer Kappa gets to 1.0, the higher the accuracy of the data. Or, with other words, a value of 1.0 would imply that the classification process was avoiding all of the errors that a completely random classification would generate.

This Kappa index is frequently written as follows:

$$C_{Kappa} = \frac{Acc_{Overall} - Prob_{chance}}{1 - Prob_{chance}}$$

where $Acc_{Overall}$ = Overall accuracy and $Prob_{chance}$ = Probability that the agreement is due to chance.

The overall accuracy and the Kappa index can be calculated at various geographic scales. In the current validation, we will compute these values at three levels: (i) at global level, (ii) by continent and (iii) by biome. Indeed, the three metrics characterise the quality of the product in a complementary manner. The global overall accuracy allows for a comparison of the product with the previous products validated at the same level (IGBP-DisCover, GLC2000, and GlobCover datasets). However, even if the production of the seasonal mosaics and the classification of the final product are conducted at the global level, observation conditions and heterogeneity of the landscape can lead to different local accuracies. Accuracy can largely differ from one continent to another (with, for example, the difference in the number of observations due to a lack of receiving stations), from one biome to the other (due to different cloud coverage conditions). Therefore, the accuracy matrices will also be produced by continent and by biome, although the statistical significance of the conclusions will be limited by the reduced population of validation samples. Six continents will be analyzed (Figure 18), while the Holdridge Life Zones will serve as basis for the stratification (Figure 19). The Holdridge classification system characterizes the climates depending on altitude, latitude and humidity. The large amount of samples in the Asian region can lead in a later stage to a distinction between the boreal part on the one hand (Siberia) and the temperate and tropical part on the other hand.

In summary, the following accuracy measurements will be provided at global level, by continent and by biome: overall accuracy (with confidence interval), user's and producer's accuracy, Kappa statistics.

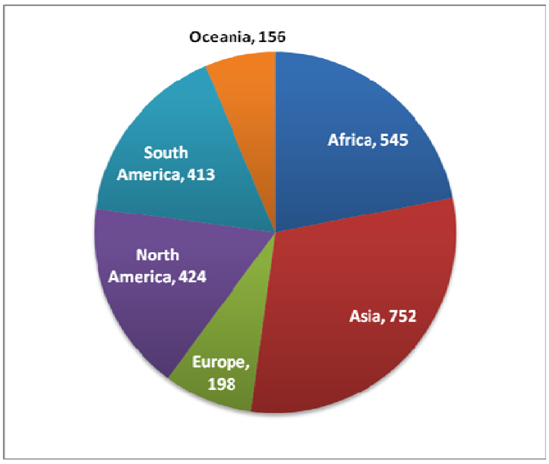

Figure 18. Distribution by continent of the 2600 samples used for the validation

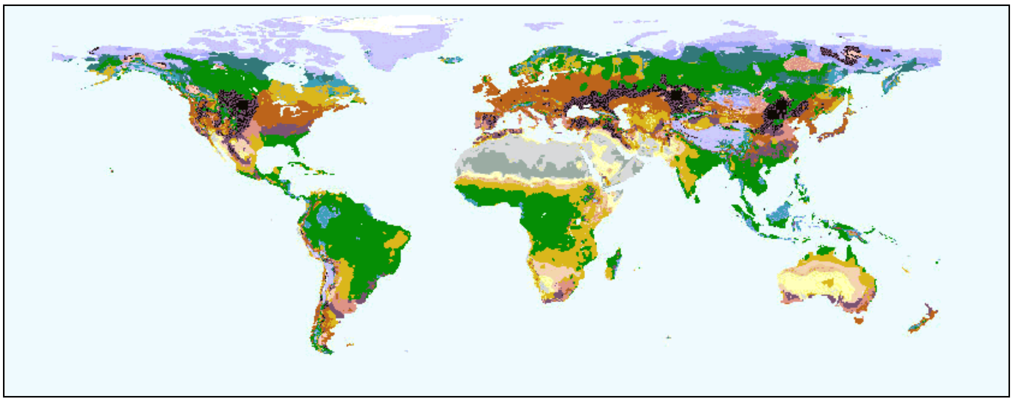

Figure 19. Map of the Holdridge Life Zones (generated by IASA, source FAO Geonetwork)

3.4.2. Accuracy from the user perspective

Apart from the overall accuracy, it is also possible to compute the accuracies of individual categories. This can either be the ratio of the correctly classified elements in each category and the total elements that were classified in that category (per row) or the ratio of the number of correctly classified elements and the total number of reference elements for that category (per column). These are called the user's and the producer's accuracy respectively.

The classical confusion matrix does not take into account the thematic distance between different classes. Indeed, in this matrix, a misclassification of a desert area as an evergreen forest has the same impact as classifying a semi-deciduous forest as and evergreen forest. Of course, these confusions have a totally different impact on the climate change applications and models. From both a producer and a user's point of view, we need to present a matrix where misclassifications between similar classes are weighted lower than misclassifications between dissimilar classes.

A number of different user perspectives can be envisaged for the CCI-LC: the carbon content, the net primary productivity, the methane emissions and specific requirements for the use of the CCI-LC product to feed into different climate models For each of these dimensions, a matrix of similarity between classes will be constructed, taking into account the physical distance between classes (e.g. the difference in carbon content) and applied to the classical confusion matrix as a weighting factor. This derived accuracy will provide the real accuracy for a specific use in the climate change models. For instance, the thematic distance will be quantified according the difference between parameters values associated to the respective Plant Functional Types (PFT) corresponding to the land cover state class.

The similarity will be described as the relative importance of different land cover classes for estimating model parameters. The importance of each differentiating two land cover classes is reflected in the relative similarity for each actual land surface parameter value. Thus, for each pair of classes (x, y) the similarity (Sim_{xy}) can be calculated by relating the specific parameter values for each class (Par_x, Par_y) to the overall range of parameter values across all classes (Par_{max} - Par_{min}):

$$Sim_{xy} = \frac{|Par_x - Par_y|}{Par_{\max} - Par_{\min}}$$

The similarity value is reported in percent with 100% representing the same parameter value for this pair of classes. As example (as derived from the URD [AP-5]), Figure 20 shows the average similarity from 9 parameter values for 12 land cover classes aggregated from 75 Olsson map classes (Hagemann, 2002). The areas with pink table cells have the highest average similarities among all land surface parameters. There is a tendency that they are located near the diagonal of the matrix reflecting somewhat the ranking of classes 1-12 from “forests” to “barren and water areas”. Most dissimilar are the non-vegetated and vegetated classes. A misclassification and confusion between two classes with large similarity will cause a much lower error in the quantitative parameter estimation than uncertainties among very dissimilar classes.

	Generalized Land Cover Legend	**1** Evergreen Needleleaf Trees	**2** Evergreen Broadleaf Trees	**3** Deciduous Needleleaf Trees	**4** Deciduous Broadleaf Trees	**5** Mixed / Other Trees	**6** Shrubs	**7** Herbaceous Vegetation	**8** Cultivated and Managed Veg.	**9** Urban / Built-up	**10** Snow and Ice	**11** Barren	**12** Open Water
1	***Evergreen Needleleaf Trees***												
2	***Evergreen Broadleaf Trees***	87.1											
3	***Deciduous Needleleaf Trees***	74.5	75.8										
4	***Deciduous Broadleaf Trees***	67.6	78.0	85.4									
5	***Mixed / Other Trees***	70.7	73.5	89.7	89.9								
6	***Shrubs***	54.2	59.1	78.1	80.0	78.6							
7	***Herbaceous Vegetation***	45.0	50.6	70.5	71.5	72.7	89.7						
8	***Cultivated and Managed Veg.***	52.5	61.5	75.3	82.9	80.0	92.3	87.4					
9	***Urban / Built-up***	34.2	33.2	53.1	48.3	55.3	58.9	65.9	55.4				
10	***Snow and Ice***	21.7	15.6	39.4	32.4	42.1	50.6	58.2	45.6	69.1			
11	***Barren***	36.6	30.5	54.3	47.3	57.0	65.5	73.1	60.5	78.5	85.1		
12	***Open Water***	30.6	24.0	48.2	40.6	50.1	57.7	65.1	53.1	75.6	88.9	89.7	

Figure 20. Matrix of the similarity between the 12 generalized land cover classes as average for 9 land surface model parameters.

The accuracy assessment information will provide an important base for the user interactions and assessments and applications of the CCI-LC products. Users will be informed about accuracies and potential limitations of the CCI-LC products for applications and allow the users to provide dedicated feedback and tune the use in climate models.

4. Comparison with other products

The comparison with other land cover products (based on international requirements) has the objective of building confidence in the CCI-LC products, thus increasing their use (i.e. for non-climate model applications) and integrating the results into other land cover monitoring efforts (i.e. MODIS land cover).

The comparison is, to some extend, driven by the notion, proposed by GOFC-GOLD, of "best" available map. Therefore, there is a need to quantify advantages of different maps (spatial resolution, temporal update, thematic accuracy, etc.) and seek opportunities to combine the most detailed and accurate areas of each dataset to a new global dataset, which is supposed to be more useful for a specific application (Herold et al. 2009).

The comparison will be performed for the most recent global land cover products and the CCI-LC. Table 10 provides details about the products that will be used for this comparison.

Product	Sensor	Spatial resolution	Temporal coverage
GlobCover 2005 (V2.2) LC map	Envisat MERIS	300 m	December 2004 – June 2006
GlobCover 2009 (V1.2) LC map	Envisat MERIS	300 m	2009
MODIS 500m LC map	Terra and Aqua MODIS	500 m	2001-2007
MODIS Vegetation Continuous Fields products	Terra and Aqua MODIS	500 m	November 2000 – December 2001
GLC2000	SPOT-VGT	1 km	2000
SYNMAP product (Jung et al. 2006)	SPOT-VGT	1 km	2000

Table 10. List of all global LC products to which the CCI-LC products will be compared

The comparison approach used will consider on the principles for harmonizing land cover information based on LCCS (Herold et al. 2006) and will be conducted on a per-class basis. Where possible, the effort will use probabilities considering the result from a comparative accuracy assessment of thematic classes (comparison by LCCS classifier) combined with spatial homogeneity of each product used (Goehmann et al. 2009, Herold et al. 2008). The methodology will provide the information to:

- Assess strengths and weaknesses of different products with particular focus on the land cover data products,
- Provide options and approaches of different products for different types of uses beyond climate models,
- Allow for derivation of information towards "best" available land cover information for a specific application,
- Support the assessment of land cover trends.

Further inter-comparisons will include the link between the land cover state product and its dynamic component, i.e. conditions, which are mainly derived from other projects and datasets (i.e. fire, phenology, snow cover etc). The inter-comparison between this integrated land cover product and the best available information globally available will use consistency checks and comparisons so discrepancies and potential divergence are understood and quantified per class and stratum as much as possible.

5. Temporal consistency assessment

This last assessment has the objective of providing users with information about the temporal consistency between the three CCI-LC products (derived from the three different "epochs" of 2000, 2005 and 2010).

The temporal assessment was asked by the users themselves. Their quantitative requirements are documented in details in the URD [AP-5] and are summarized in Table 11.

	Stability		
	GCOS	*CMUG*	*CCI-LC users*
Land cover products	> 85%	90-95%	> 85%

Table 11. Temporal stability requirements of the CCI users

The assessment will be performed between the three CCI-LC products, by comparing them one to one, on a per-class basis. The comparison will be preceded by an a priori land cover classes' erosion to avoid any border-effects. Indeed, despite a geometric accuracy much below the pixel resolution, the surface actually measured for a given pixel can significantly vary from one day to another. This is especially the case for the MODIS time series (which could be used in the classification chain).

According to such approach, any discrepancy between 2 products could be considered either as a land cover change or as a temporal instability. The results will be interpreted in the light of the outcomes obtained through the three previous validation steps: the confidence-building procedure, the independent statistical validation and the cross-comparison with existing products. These outcomes will provide valuable information about

- the per-class accuracy;
- the location of poorer classification performance;
- the influences about the landscape diversity and fragmentation;
- the types of errors which affect the land cover classes.

This kind of information is expected to be useful for discriminating the land cover changes from the temporal instabilities.