THE GOFC-GOLD/CEOS LAND COVER HARMONIZATION AND VALIDATION INITIATIVE: TECHNICAL DESIGN AND IMPLEMENTATION

M. Herold[1], C.Woodcock[2], S. Stehman[3], J. Nightingale[4], M. Friedl [2] and C. Schmullius[5]

[1] Wageningen University and Research Centre, Laboratory of Geo-Information Science and Remote Sensing, Droevendaalsesteeg 3, 6708 PB Wageningen, Netherlands, Email: Martin.Herold@wur.nl
[2] Department of Geography, Boston University, USA
[3]State University of New York, USA
[4] Sigma Space Corportation, NASA Goddard Space Flight Center / CEOS WGCV Land Product validation subgroup, USA
[5]Department for Earth Observation, University of Jena, Germany

ABSTRACT

A global effort to assess the accuracy of existing and future land cover products derived from a variety of satellite sensors over a range of spatial resolutions is being led by the Land Cover Implementation Team (LC-IT) of GOFC/GOLD (Global Observation of Land Cover Dynamics) in conjunction with the CEOS (Committee on Earth Observation Satellites) WGCV (Working Group on Calibration and Validation) LPV (Land Product Validation) subgroup. The first phase of this effort is complete and culminated in a publication of community consensus "best practices" for validation of global land cover datasets (2). The next phase is to implement the recommendations outlined in the "best practices" document. A "living database" of global randomized sample sites will form the basis of accuracy assessment for a host of global land cover products (GLC2000, MODIS land cover, GLOBCOVER, United Nation's Forest Resource Assessment (FRA2010), and the Mid-Decadal Global Land Survey. This "living dataset" will also be a community resource available for use in validation of regional or national mapping efforts using LCCS (UN FAO's Land Cover Classification System). Based on the known accuracy of existing land cover products, GOFC/GOLD will to develop and update a "best currently available" global land cover map. Individual geographic regions may be selected from different land cover products (global, national or regional), or they may be combined in various ways

1. INTRODUCTION

Several global and regional datasets have been derived in response to the need for worldwide and up-to-date information regarding land cover and related temporal dynamics. Product development was driven by different national or international initiatives; as a result, the subsequent mapping standards adopted reflect the varied interests, and the requirements and methodologies of the originating programs. Although efforts in harmonization and validation are mentioned in nearly all related mapping projects as well as in many other circumstances, there is only limited compatibility and comparability between these different maps. The GOFC-GOLD land cover implementation team and the CEOS WGCV LPV subgroup are developing a land cover harmonization and validation framework. In this communication we present our rationale, plans, and current status of this initiative. With the evolving monitoring programs for Essential Climate Variables (ECV), the issue of harmonization and validation of global land cover products requires further attention

2. INTERNATIONAL DRIVERS

There is a significant effort on the political and strategic level to move global and regional land cover observations into a more operational mode. Key initiatives foster an integrated and operational global land cover observation framework include:

- the Implementation Plan of the Global Carbon Observing System (GCOS) to the United Framework Convention of the Climate Change (UNFCCC) describes a number of specific tasks to provide operational and internationally agreed land cover information for reducing uncertainties in observing the global climate system,
- the UNFCCC fostered activities for observing land cover as an Essential Climate Variable (ECV) building upon the GCOS implementation plan,
- the Group of Earth Observation (GEO) fosters land cover observation as important for all areas of societal benefits. Part of a dedicated GEO work plan task urges activities for consistent and continuous land cover observations including a robust and sustained land cover product accuracy assessment (1)).
- the Integrated Global Observation Strategy (IGOS) has developed a dedicated land theme for Integrated Global Observations of the Land (IGOL, Townshend et al. 2007). The IGOS themes are currently in transition to GEO tasks and activities.

All these strategic and political initiatives are calling for efforts in harmonization and validation to reduce the current heterogeneity in land cover observations into a joint effort driven by user requirements, robust analysis of uncertainties, and synergy among mapping products

Proc. 'ESA Living Planet Symposium', Bergen, Norway
28 June – 2 July 2010 (ESA SP-686, December 2010)

based on international consensus. A thorough accuracy assessment is the foundation for any efforts to synthesize existing land cover characterizations, to service an informed user community, and to form the base for good science. International environmental protocols and agreements imply that all useful land cover products are being evaluated and possibly challenged.

3. INTERNATIONAL COORDINATION

The work presented here is embedded in the joint activities of the Committee on Earth Observing Satellites (CEOS) Working Group on Calibration and Validation's Land Product Validation sub-group (CEOS WGCV) and Global Observations of Forest Cover and Land Dynamics (GOFC-GOLD). There is particular need for coordinated international activities for a number of reasons:

- to establish a continuous process of engaging with political processes to understand their requirements and provide relevant technical contributions;
- to assess and translate existing (often broad) international user needs in operational land observation and monitoring strategies;
- to evolve and implement standardized approaches to land cover characterization and land cover validation;

The joint activities between the CEOS LPV subgroup and the GOFC-GOLD land cover office have received a further push during the CEOS WGCV-29 plenary meeting, September 30 – October 3, 2008 in Avignon, France (http://www.star.nesdis.noaa.gov/smcd/CEOS/WGCV29) and the 3rd GOFC-GOLD land cover symposium 13-17.October. 2008 in Jena, Germany (http://www.gofc-gold.uni-jena.de/sites/Jena08.php). Such activities are technical precursors to the implementation of an operational global land cover validation system is key component of observing land cover as ECV. Such technical efforts are part of a number of near term priorities:

- Continue cooperation between GOFC-GOLD and the CEOS land Cal/Val subgroup and support of international projects,
- Utilize and make available existing global reference databases (i.e. from GLC2000 and GLOBCOVER projects),
- Define standard accuracy assessment protocols for fine-scale land cover change and area estimates, in particular relevant for UNFCCC REDD (Reducing Emissions from Deforestation and Degradation) implementation.
- Planning and fostering implementation for an operational validation component of global land cover monitoring (i.e. for Essential Climate Variables)
- Use of "application specific weighting" for accuracy reporting to link land cover observations with areas of societal benefits,
- Continued communication and dedicated contributions with political and policy level including the UNFCCC and Group on Earth Observations (GEO work plan tasks) to ensure that efforts are salient (relevant and useful), legitimate (fair, transparent, involve international consensus and local experts), and credible (technically robust, consistent and continuous)

The CEOS WGCV LPV subgroup had initially defined a three tier validation hierarchy. However, in response to the evolving ECV monitoring activities, CEOS LPV incorporated a stage 4 validation component that defines an operational component to ensure that time-series land products are systematically validated:

Stage 1	Product accuracy is assessed from a small (typically < 30) set of locations and time periods by comparison with reference in situ and / or higher resolution airborne or satellite data. Spatial and temporal consistency of the product and consistency with similar products has been evaluated over selected locations and time periods.
Stage 2	Product accuracy is estimated over a significant set of locations and time periods by comparison with reference in situ and / or higher resolution airborne or satellite data. Spatial and temporal consistency of the product and consistency with similar products has been evaluated over globally representative locations and time periods. Results are published in the peer-reviewed literature.
Stage 3	Uncertainties in the product and its associated structure are well-quantified from comparison with reference in situ and higher resolution airborne and satellite 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 consistency with similar products has been evaluated over globally representative locations and periods. Results are published in the peer-reviewed literature.
Stage 4	**Validation results for stage 3 are systematically and operationally updated by independent actors for comparative assessment of existing products, when new products are released and as the time-series expands.**

While all previous global land cover validation efforts have focused on achieving Stage 3 validation, the aim for the future efforts should be to move towards a Stage 4, temporal validation.

4. OVERVIEW OF ACTIVITIES

Standard techniques and guidelines

The first step to ensure robust global land cover accuracy assessments is the development of standard methods. Current efforts are building upon a set of existing core accuracy assessment analysis methods that should be routinely adopted as a baseline for reporting map accuracy. These include employing probability sampling and consistent estimators within the design-based inference framework to generate estimates of the overall accuracy of the map as well as per-class accuracies and the variances of these estimates. Confusion matrices, user's and producer's accuracies should be published with the accuracy assessment, and the data used to derive these estimates should be archived and made accessible to the user and science community. The international effort to define consensus "best available" methods for global land cover accuracy assessment has been completed and published (2). In addition, international progress toward acceptance of the Land Cover Classification System (LCCS, 3) as the standard for defining legends for land cover maps points to a time when direct comparison of alternative land cover products will be possible. All global legends have been translated to LCCS (4).

In the next phase, the existing best practices document will focus on the area of standard accuracy assessment for fine-scale land cover change and area estimates, in particular relevant for REDD implementation (5).

Utilizing existing reference datasets

To date, there has been no systematic study that has used a true comparative accuracy assessment based on a suitable global reference database. This shortcoming is being addressed by the group activities through making best use of existing reference databases. The work currently involves the Global Land Cover 2000 (GLC2000) global database of reference sites (6) and the GLOBCOVER validation database (7). Both databases have been developed using a flexible land cover characterization based on classifiers of the UN Land Cover Classification System (3). Utilizing this reference database and a number of existing global 1 km datasets, the objectives of the paper is to:

- Process and analyze the reference database to a format useable for a global accuracy assessment,
- Apply a generalized global land cover legend of eleven classes based on LCCS,
- Perform a global comparative validation of five 1 km global land cover maps using standard accuracy analysis methods,
- Utilize the results and experiences from the comparative validation to specify requirements for an operational global land cover accuracy assessment framework for future efforts.

Designing an operational validation

Based on standard methods developed by the international community for global land cover accuracy assessment, the scheme in Figure 1 emphasizes several dimensions of such an initiative that are crucial for its success (2):

- The integration of in situ/local, regional, and global land cover observations,
- The combination of harmonization and validation activities towards interoperability, product synergy, and improved usability of land cover products,
- The importance of design-based sampling for selecting a statistically rigorous set of reference sites, their LCCS-based interpretation through regional networks, and the maintenance of this database as "living database" through regular updates and re-interpretations,
- The temporal dimension including an establishment phase and an operational phase,
- The different levels of land cover validation including primary validation, comparative validation (for both global and regional products), and updated validations including the assessment of change and incorporation of any new mapping products.

The accuracy analysis activities will focus on several validation levels (Figure 1):

- **Primary validation:** A statistically robust validation has already been completed for individual global land cover datasets (IGBP DISCover, GLC2000 etc.). This initiative will provide a consistent primary validation for all existing global land cover products.
- **Comparative validation:** The second step is a comparative validation of existing datasets through the comparison of appropriate accuracy measurements. The comparative assessment is essential for contrasting and comparing different datasets and the development of an interoperability strategy. The basic goal is to identify strengths and weaknesses of individual datasets relative to other land cover products. A comparative validation might also include regional land cover datasets.
- **Updated validation:** Given a regular update of the reference database, an operational and continued assessment of the accuracy and validity of datasets

can be established even after many years of their production.

- **Validation of new datasets:** The reference dataset is designed to provide accuracy assessments for any new global land cover product independent of the spatial resolution. An essential requirement, however, is that the legend development be based on LCCS to allow rigorous comparison between the new land cover product, the ground reference data, and any previous land cover map.

Initial expectations are that the following existing and planned global land cover products will be validated: GLC2000, MODIS Land Cover, GLOBCOVER, and supporting the FAO FRA 2010 remote sensing survey. Additionally, the test site database, or "living database", described below will also be useful for evaluating land cover products derived from the MDGLS dataset. The primary objective of the accuracy assessments will be the estimation of the traditional per-pixel overall and per class accuracy and the associated stand errors. These results will be provided at global as well as continental or regional scales. The primary steps of map accuracy assessment are sample design, response design and analysis, and the discussion below is organized according. A secondary set of accuracy estimates will relate to the accuracy of area estimates at various block sizes, as well as accuracy of landscape pattern.

Sample Design

The sample design will be based on a probability sample to ensure rigorous statistical inference is supported (9). The plan is to use 5kmx5km blocks as sample sites. The use of blocks as sampling units enriches the information in the sample by yielding many (depending on the spatial resolution of the land cover map) observations (i.e. pixels or polygons) at each sample site as well as the block level aggregate data. It also provides flexibility for use with land cover maps at different spatial resolutions. Given that high resolution imagery will need to be acquired and interpreted for each sampling unit, the use of blocks helps minimize the number of images required.

An independent data source will serve as the basis of the stratification. Stratification provides many benefits, including the ability to increase sample sizes within individual strata in an attempt to ensure that map classes with small areas are adequately represented in the sample, and to distribute the sample regionally to enhance the geographic representation of the accuracy estimates. The stratification chosen must be flexible enough to allow for augmentation of the sample size in a particular region (or country, for example) in support of improved precision of accuracy estimates for that area.

Response design

The reference data for the sample sites will be, ideally, based on interpretation of very high resolution imagery by regional experts. We expect to use SPOT 5 imagery (pan sharpened to 2.5m). The interpretation of the imagery will be done manually on standardized image products. The precise nature of the image products remains to be determined, and may include some degree of processing of the data that will reduce the effort level required from the local experts. For example, we may use automated image segmentation methods to define polygons, and the interpreter's task will be limited to labeling each polygon (following approaches developed at JRC as part of the TREES project). The intent is to use a minimum mapping unit of 1 hectare. At this scale, comparison with pixel sizes on the order of 300-500m (GLOBCOVER and MODIS) will be possible following aggregation of the ground truth data. It will also support assessment of Landsat (or similar)-based analyses with similar or larger minimum mapping units. The labeling of the polygons in the very high resolution imagery will follow the hierarchical conventions of LCCS, with the expectation that the local interpreters will be able to provide higher levels of thematic detail than is typically included in the legends of global land cover maps.

Analysis

Any map of interest would be subjected to a common analysis protocol that includes at least the methods identified in the joint GOFC/GOLD and CEOS Cal/Val document on "best practices" for accuracy assessment. The analysis will include an assessment of per pixel accuracy following the common practice of producing an error matrix and accompanying measures such as overall accuracy, and class-specific measures of user's and producer's accuracies. Accuracy estimates will be based on stratified random sampling formulas (8) to account for the sampling design implemented, and standard error formulas based on cluster sampling will account for the within-block correlation of the observations. Accuracy estimates can be provided for any subregion of interest, for example, by continent, country, or biome.

5. IMPLEMENTATION ISSUES

The implementation of validation activities for ECV monitoring requires the consideration of key issues:

- Users need to understand what datasets they should use for specific purposes and why, thus the validation has to have a user-driven component,
- Land cover validation is a continuous process and needs to be consistent for historical (i.e. AVHRR derived datasets) and future periods,

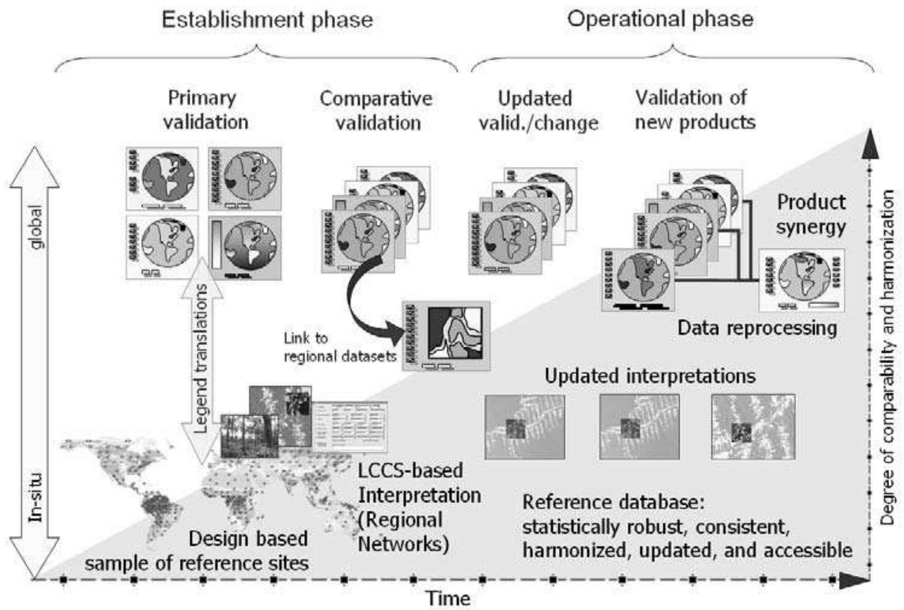

Figure 1: Schematic implementation framework for an operational global land cover harmonization and validation framework.

- Such efforts require additional resources for the interpretation of multi-temporal high-resolution datasets for reference (i.e., SPOT, KOMPSAT, RAPIDEYE) including costs for data, processing, and interpretation, sampling design, and comparative accuracy reporting for a quantitative assessment of multiple datasets at different times,
- Developing and maintaining a flexible and transparent land cover reference database of core sites provides the foundation of a long-term validation system.

The thematic validation of the global land cover products and particularly for evolving ECV monitoring products is required to be of an independent nature, follow international standards, and use existing reference datasets and related experiences as much as practical. However, the independent accuracy assessment for monitoring land cover as ECV poses additional requirements and challenges. The validation exercise needs to explore the options to perform validations for multiple dates in time. For historical periods the availability of reference data sets (i.e. for the 1990's) is rather limited. For the future there is need to develop the foundations and start implementing an operational monitoring system that ensures that progressing annual global land cover datasets can assessed consecutively. For the validation of land cover change, different protocols and approaches will need to be used. In addition, there are a number of technical challanges that have been discovered through the experiences of previous validation exercises from GLC2000, GLOBCOVER and MODIS and regional and national studies:

- Integration of different reference datasets for global land cover validation will be an issue. Many datasets exist for different times and utilize diverse standards. A critical assessment will be needed to ensure the validation results are comparable given the requirements for any product.
- High-resolution (spatial and temporal) satellite observations provide a suitable reference source. A compromise has to be made when balancing the cost for reference data and interpretations and the need to develop and implement an operational and efficient validation framework.
- Reference data sets are not free of errors and such quality is to be considered. In particular the issue of inconsistency between expert interpreters of the reference data has to be addressed.
- For the case of the operational global land cover validation, the issue of area frame sampling versus point sampling is important to consider and further community input on decision for implementation is required.
- Error traceability and error propagation along the processing chain for the different thematic products for various users will be more important than for previous exercises.

Although these critical validation issues are known they are most efficiently addressed in a step-wise approach in a process to build a consistent, long-term validation system for observing land cover as ECV.

6. TOWARDS A "BEST" GLOBAL MAP

The 'intercomparison' between products to evaluate their relative consistency and accuracy is very important for users when considering which of several products to use. It is also mandatory when combining several products into the 'best available product'. However, if two products are in good agreement, they could both be wrong: accuracy assessment is thus mandatory through comparison to independently acquired reference data.

The work aims to implement the concept to combine the strengths of the various land cover products to produce a "currently best available" land cover map, and make it freely available. The goal of this effort is to provide a community consensus best product that will help users decide which land cover products to use. Various strategies are possible for developing such a "currently best available" map and they will be explored. The key point is that they will be driven by the results of the accuracy assessment as applied to the combination of multiple maps. One approach is to select the resulting class for any particular location on the basis of the overall accuracy of the observed combination of classes from multiple source maps. In the places where existing maps disagree, the accuracy assessment data will provide a mechanism for deciding the most likely class. The approach to be employed is to use GIS-style map overlay analyses and ask questions like "what is the most commonly correct class" for various combinations of the source map. Another possible approach is to include entire maps for selected areas. For example, if a land cover map is produced for an individual country from high resolution imagery (Landsat, for example) and has demonstrated accuracies higher than even the combination of existing global maps for the area, then the high resolution map for that country could be included in the "currently best available" map. Four guiding principles will govern the process:

- More accurate is better
- Higher resolution (more spatial detail) is better
- More thematic resolution (more detailed definition of classes) is better
- More recent is better

A prototype exercise for combining global land cover maps based on accuracy is described in Goehman et al., (10).

REFERENCES

(1) Herold, M., Woodcock, C. E., Loveland, T. R. Townshend, J., Brady, M., Steenmans C. and C. Schmullius (2008a). Land Cover Observations as part of a Global Earth Observation System of Systems (GEOSS): progress, activities, and prospects, IEEE Systems, 2, 3, 414-423.

(2) Strahler, A., Boschetti, L., Foody, G. M., Friedl, M. A., Hansen, M. C., Herold, M., Mayaux, P., Morisette, J. T., Stehman, S. V., & Woodcock, C. 2006. Global Land Cover Validation: Recommendations for Evaluation and Accuracy Assessment of Global Land Cover Maps, Report of Committee of Earth Observation Satellites (CEOS) – Working Group on Calibration and Validation (WGCV), JRC report series.

(3) Di Gregorio, A. 2005. UN Land Cover Classification System (LCCS) – classification concepts and user manual for Software version 2. Available online at: www.glcnlccs.org.

(4) Herold, M., Woodcock, C., Mayaux, P., Baccini, A. and C. Schmullius (2008b). 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.

(5) Herold, M. & J. Sambale (2009). Report of the GOFC-GOLD/CEOS Workshop on Land Cover Change Accuracy Assessment, as part of the Third GOFC-GOLD Land Cover Symposium, 13 – 17 October 2008, Jena, Germany, available at www.fao.org/gtos/gofc-gold/series.html.

(6) Mayaux, P., H. Eva, J. Gallego, A. H. Strahler, M. Herold, S. Agrawal, S. Naumov, E. E. De Miranda, C. M. Di Bella, C. Ordoyne, Y. Kopin, & P. S. Roy. 2006. Validation of the Global Land Cover 2000 map. IEEE Transactions on Geoscience and Remote Sensing 44: 1728-1739.

(7) 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. 2008. GLOBCOVER products Description and Validation Report, http://postel.mediasfrance.org/IMG/pdf/GLOBCOVER_Products_Description_Validation_Report_I2.1.pdf

(8) Card, D. H. 1982. Using known map category marginal frequencies to improve estimates of thematic map accuracy. Photogrammetric Engineering & Remote Sensing 48: 431-439.

(9) Stehman, S. V. 2000. Practical implications of design-based sampling inference for thematic map accuracy assessment. Remote Sensing of Environment 72: 35-45.

(10) Goehmann, H., Herold, M., Jung, M., Schultz, M. & C. Schmullius (2009). Prototyping a "best land cover map approach" using existing global land cover datasets and comparative accuracy assessment. Proceedings of ISRSE 33.