

LACO-WIKI: AN OPEN ACCESS ONLINE PORTAL FOR LAND COVER VALIDATION

L. See^{a,*}, C. Perger^a, M. Hofer^b, J. Weichselbaum^b, C. Dresel^c, S. Fritz^a

^a Ecosystems Services and Management Program, International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, 2361 Laxenburg, Austria - (see, pergerch, fritz)@iiasa.ac.at

^b GeoVille Information Systems GmbH, Sparkassenplatz 2, 6020 Innsbruck, Austria – (hofer, weichselbaum)@geoville.com

^c University of Applied Sciences, 2700 Wiener Neustadt, Austria - dresel@fhwn.ac.at

Commission II, WG II/4

KEY WORDS: Land Cover, Validation, Open Access, Standards, Accuracy Assessment

ABSTRACT:

The LACO-Wiki tool represents an open access, online portal that offers standardized land cover validation at local to global scales. LACO-Wiki integrates the LACOVAL prototype for land cover validation and the Geo-Wiki system for visualization, validation and crowdsourcing of land cover. This paper presents a conceptual overview of the LACO-Wiki system and describes the main validation workflow, in which the user uploads the map for validation, creates a validation sample, carries out the sample interpretation and generates a report detailing the accuracy assessment. In addition to a land cover validation tool, LACO-Wiki is also intended to become an open access repository for calibration and validation data that can be used by the land monitoring community to improve future land cover products.

1. INTRODUCTION

Land cover validation is an important step in the process of generating a land cover map derived from remotely-sensed imagery. Validation is a quality assurance process and involves quantifying the accuracy and the associated uncertainties in these derived products (Justice et al., 2000). Users of these map products utilize accuracy measures as a form of confidence regarding how well the maps represent the land surface, which informs their fitness-for-use as inputs to models or for subsequent decision-making purposes. Map producers also need accuracy measures to understand how well their classification algorithms work and where they can make improvements. However, validation exercises are usually undertaken in-house with tools and procedures that are specific to a particular institute or organization. There is little standardization or a set of open tools available for land cover validation.

An initial prototype was developed, called LACOVAL (Land Cover Validation). Funded by the European Space Agency, the LACOVAL prototype attempted to provide some standardization to the validation process. The LACOVAL prototype was trialed with users from the European land monitoring community in the context of Copernicus land monitoring. However, areas for improvement, e.g. in the areas of data handling, access and sharing, ease-of-use as well as enhancements to the capabilities of this prototype were identified. The LACO-Wiki tool is the successor to LACOVAL and is currently in development. A more detailed investigation of additional requirements has since been conducted to ensure that LACO-Wiki will cover the functionalities necessary for an operational and standardised validation solution for the land monitoring community.

Therefore, LACO-Wiki will provide a framework for assembling land cover validation methods and workflows in one web interface including:

- Online storage and management of all necessary geoinformation data;
- Guidance for operators through the entire validation process in the form of ‘wizards’;
- Methodologically sound sampling designs and rapid interpretation of samples in a user-friendly environment; and
- The generation of state-of-the-art accuracy reports for communication to map users.

This paper presents an overview of the LACO-Wiki tool to date, which brings together standardized land cover validation methods and workflows into a single portal. This includes the storage and management of land cover maps and validation data; step-by-step instructions to guide users through the validation process; sound sampling designs; an easy-to-use environment for validation sample interpretation; and the generation of accuracy reports based on the validation process. LACO-Wiki merges the LACOVAL prototype with the architecture and database design of Geo-Wiki, a tool for improving global land cover (Fritz et al., 2012). LACO-Wiki represents the first online solution for land cover map validation, where one of the main design goals has been to simplify the workflows as much as possible so that the tool can be used in a wide variety of contexts. The tool has been developed for validating land cover data sets from government departments (e.g. in the context of the Austrian Land Information System LISA) and for international institutions such as the European Environment Agency, e.g. Copernicus layers. LACO-Wiki is also aimed at validation of land cover maps in an educational and research context and by map producers at all scales including global land cover.

A longer term goal of LACO-Wiki is to build a reference database for calibration and validation data globally as different users validate their maps and contribute the data to this repository. A wide range of potential stakeholders interested in

* Corresponding author

land cover validation can therefore benefit from the current and future developments in LACO-Wiki.

The remaining sections of this paper describe the conceptual overview of the LACO-Wiki system and the main validation workflow, followed by conclusions.

2. THE LACO-WIKI SYSTEM

2.1 Conceptual Overview and Architecture

A user requirements analysis was undertaken with stakeholders from the European land monitoring community, GOF-C-GOLD (Global Observation for Forest Cover and Land Dynamics) and industry partners to develop the main system requirements. These requirements have been organized into nine sets of features (Figure 1). The five central features represent the main steps in the validation workflow while the user management, data sharing, system performance, and security are cross-cutting.



Figure 1. LACO-Wiki feature overview

Users access the LACO-Wiki system via existing applications such as Geo-Wiki, Facebook and Google, which simplifies access and minimizes risks associated with security weaknesses.

To ensure good system performance, the system architecture (Figure 2) is scalable in terms of storage and processing capacity. The user accesses the LACO-Wiki portal, which distributes the data and the processing tasks across multiple DataStore servers (responsible for data management and publishing), and Worker servers (responsible for the data analysis and generation of the validation samples). RESTful Web services are used for communication using the service-oriented architecture (SOA) design pattern.

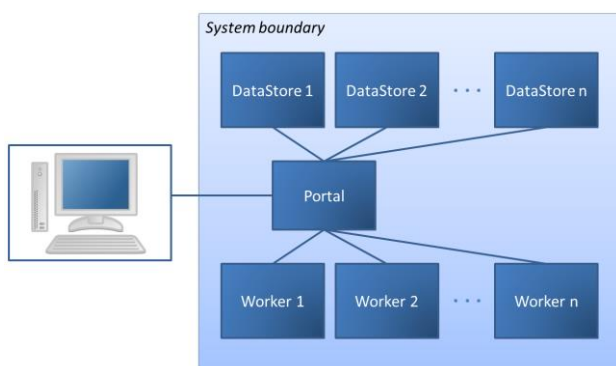


Figure 2. The LACO-Wiki system architecture

Data sharing functionality determines how the data are managed and shared as well as how results are processed between users. The user who uploads the map and creates the validation sample also defines the data sharing settings for published data sets, validation samples, validation sessions and reports. It is possible to create groups of users to share the workload of a single validation session (or campaign). Multiple users can be given samples at the same location in order to examine agreement between interpreters and judge the difficulty of the interpretation tasks. To ensure security, the data sets and maps cannot be accessed directly by the users but rather access is provided via the portal.

The LACO-Wiki system contains Bing Maps as base imagery and may include Google Maps imagery in the future. Users will be able to upload their own base imagery as well.

2.2 The Validation Workflow

This section describes the five features in Figure 1 that represent the steps in the validation process.

2.2.1 Data upload and legend definition

The Datacenter is the starting point for creating a new data set via uploading or managing a saved session in which data have been previously uploaded (Figure 3).



Figure 3. Starting the workflow from the Datacenter

The map that is uploaded can be a continuous or categorical raster or vector data file (i.e. shapefiles, geotiffs, etc.). The user enters a name, whether the data set is categorical or continuous, a description and then drags and drops a file to the Files box or browses for a file (Figure 4).

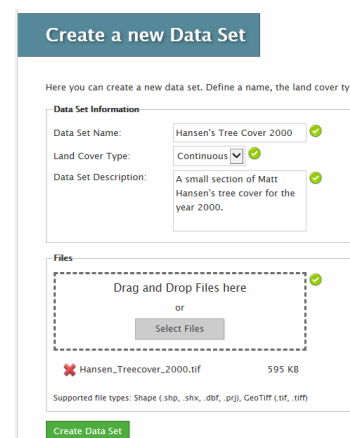


Figure 4. Uploading a data set to the system for validation

Once the data set is loaded, the user is automatically redirected to the Data Set Details page (Figure 5), which allows the user to preview the image and provides basic information about the data set.

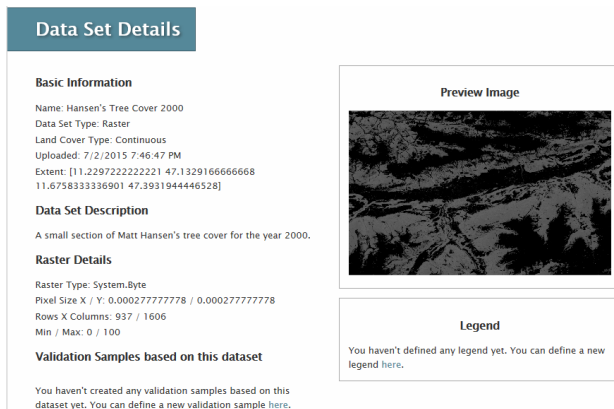


Figure 5. A page showing the details of the data set uploaded and a preview of the image

The message on the bottom right of Figure 5 indicates that the legend has not yet been defined. The Legend Designer allows the user to change the values, ranges and colours selected, which are read directly from the file. Once the Legend is saved, the new colours are applied to the previewed image (Figure 6).

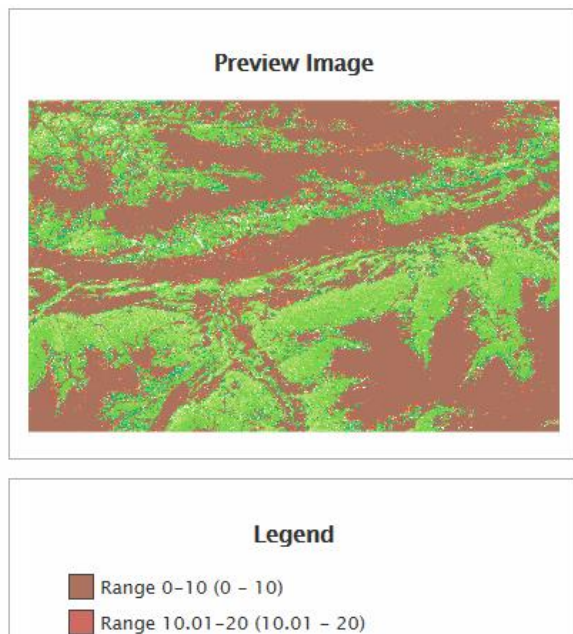


Figure 6. The legend applied to the image as created by the Legend Designer feature

In the future, users will be able to upload their own or pre-existing legends, e.g. the Corine Land Cover (CLC) legend. Users will publish their datasets via an OGC compliant Web Map Service (OGC WMS) and specify the terms and conditions of access and use in this step.

2.2.2 Validation sample definition and generation

Users define the area of interest by designating the bounding box, drawing a polygon or uploading a polygon feature.

Samples are then generated using different methods including random (Figure 7); systematic; and stratified. There will be an option added in the future to augment samples from a user-contributed dataset or a combination of these methods.

Create new Sample Collection

Sample Name: ✓

Sample Description: ✓

The given dataset contains a **raster file** and supports the following sample types:

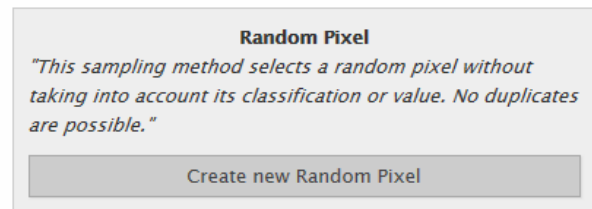


Figure 7. Setting up a randomly selected validation sample

With the stratified method, users can define class-specific sampling based on area statistics. Recommendations on minimum sample size per class (based on the required confidence interval and maximum acceptable uncertainty) are provided for stratified sampling while guidance is given on appropriate sample sizes for the random sample method.

For raster data sets, the size of the sample pixels is based on the resolution of the uploaded map for validation. For vector data sets, validation is object-based. Here there are two different ways of selecting objects for validation. The first is based on a random selection of objects regardless of the size of the objects. The second is a random selection based on the size, which has been shown to have some advantages over pixel-based validation (Radoux et al., 2008). Moreover, if the sampling is based on object area, then the confusion matrix is weighted by the areas of the objects.

2.2.3 Validation session definition

Once the validation sample is created, the user can begin their validation session (Figure 8). Users can also continue where they left off as sessions can be saved or they can review the validations from a session.

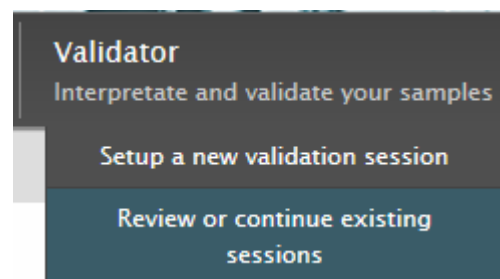


Figure 8. Setting up a randomly selected validation sample

Users can specify three different types of validation method:

- **Blind validation:** Users interpret the image and choose a value from a class legend or a percentage value, e.g. percentage forest cover. The map that is being validated is not provided to the user. This is the approach used to collect data via Geo-Wiki (See et al., 2015). The idea is that the user is not influenced or biased in their classification of an object or pixel.
- **Plausibility validation:** Users are provided with the actual value from the land cover map at that sample location and they must agree or disagree with this value. This is an approach that was used in the original version of Geo-Wiki (Fritz et al., 2009) whereby users examined an area and then determined whether the GLC-2000, MODIS and GlobCover land cover data sets captured what was visible from Google Earth in a good or bad way or whether the user was unsure.
- **Enhanced plausibility validation:** This method is the same as plausibility validation except that users can provide a corrected value when they disagree with the actual map value or select “interpretation is not possible with current reference data” (i.e. cloud cover in the reference data). This type of approach could be used to improve land cover products interactively and moves beyond a pure validation approach. Users would effectively be part of the mapping process, which would be similar to an OpenStreetMap or more of a Wiki type approach (Ramm et al., 2011).

Validation can also be enhanced via the use of secondary data, e.g. geotagged photographs at the location of the sample or profiles of vegetation health status, e.g. NDVI (Normalized Difference Vegetation Index) pulled in from another system via a web service. Both of these are features of the current Geo-Wiki tool (Fritz et al., 2012) and will be incorporated into LACO-Wiki in the future.

2.2.4 Sample interpretation

This feature allows users to specify which layers are used for the validation and it determines how the samples are presented, e.g. systematically, randomly or whether users can revisit and revalidate previous samples.

2.2.5 Reporting and download

The feature creates reports in various file formats and styles, where different accuracy measures are available and selectable by the user (Figure 9). There is a number of accuracy measures currently built into the system. Using a contingency matrix, a series of standard measures are calculated (Congalton and Green, 2009), including overall accuracy with confidence limits, which is the proportion of correct classifications in relation to the total as well as producer’s and consumer’s accuracy, which reflect different kinds of misclassification. Producer’s accuracy or commission errors reflect classes that were incorrectly identified in relation to the reference data class (e.g. forest classified as an urban reference class) while user’s accuracy or omission errors reflect the exclusion of classes (e.g. urban areas in the reference set that were misclassified as forest). Another common accuracy measure available for reporting is the kappa coefficient, despite the fact that there are known problems with this measure (Foody, 2004, Pontius and Millones, 2011).

Less commonly used accuracy measures are also reported. The average mutual information (AMI) (Finn, 1993) is derived from information theory. This can be thought of as the predictability of validated classes as a function of the mapped class, and the amount of information conveyed in this prediction. If a base 2 logarithm is used, then units are in ‘bits’. AMI can also be normalized to the theoretical maximum amount of information possible given the distribution of categories in a map and is expressed as the percentage AMI. This is the value reported by LACO-Wiki. Finally, the quantity and allocation disagreement (Pontius and Millones, 2011) are provided.

Other measures will be added in the future based on user demand. Which measures users choose, and how they interpret these measures, is entirely up to them. The goal here is to provide a range of choice for the user.

It is also envisaged to implement the possibility to define workflow templates to facilitate greater usability of the system for non-experts.

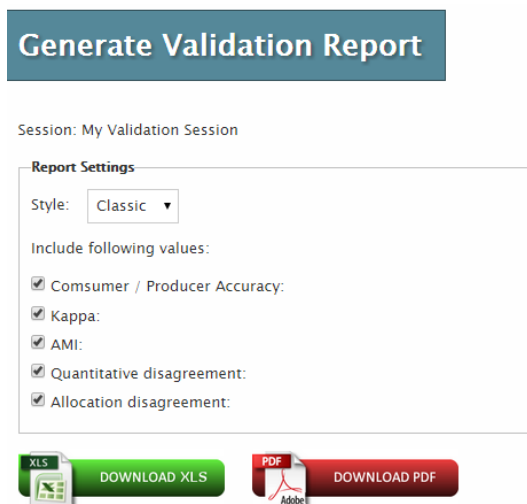


Figure 9. Generating reports in the LACO-Wiki system

The user can also download the raw sample data, the validation points and the contributed land cover maps, depending upon the access and data sharing rights.

The implementation, testing and validation phase will take place iteratively with feedback from stakeholders as the platform develops. Validation will specifically target the requirements of European land monitoring stakeholders, addressing the product portfolio from Copernicus land monitoring (e.g. CORINE Land Cover, High Resolution Land Cover Layers, Urban Atlas, and Riparian Zones – Figure 10). In addition, LACO-Wiki is intended to become a one-stop-shop for land cover validation for value-added monitoring applications worldwide.



Figure 10: LACO-Wiki will address the validation of European land monitoring products

3. CONCLUSIONS

This paper presented an open, online solution to map validation called LACO-Wiki, which is intended to fill a current gap in the area of land cover validation. Since the system will be open, LACO-Wiki will be available as a resource to students who need to create land cover maps as part of projects and postgraduate research. Researchers can use this tool to examine different methods of accuracy assessment while businesses can create value added land cover, land use and land change products using this solution, e.g. in the validation of EU Copernicus products.

LACO-Wiki also has the potential to become a substantial repository for calibration and validation data that can complement authoritative reference data portals such as that offered by GOF-C-GOLD (Global Observation for Forest Cover and Land Dynamics). The LACO-Wiki repository will also be extended to land cover maps, which is a significant step forward as both validation data and land cover maps are rarely shared among stakeholders, yet represent a valuable resource for map users and the land monitoring community. Finally, there are plans to integrate crowdsourced in-situ data into the platform, for example, geotagged photographs collected through the Geo-Wiki pictures smartphone app, geotagged photographs from repositories such as Flickr and Panoramio, and the well-documented landscape photographs from the Oklahoma Field Photo Library. Examining the value of photographs for land cover calibration and validation are at the cutting edge of current research, see e.g. Foody and Boyd (2012) and Leung and Newsam (2014).

LACO-Wiki is currently being tested but will become available at <http://www.laco-wiki.net>. In the meantime, the LACOVAL prototype, on which LACO-Wiki is built, can still be used at <http://lacoval.geo-wiki.org>.

ACKNOWLEDGEMENTS

This research was supported by the Austrian Agency for the Promotion of Science (FFG) under ASAP-10 in the LACO-Wiki project (No. 844385).

REFERENCES

- Congalton, R.G. and Green, K., 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton FL, USA: CRC Press/Taylor & Francis.
- Finn, J.T., 1993. Use of the average mutual information index in evaluating classification error and consistency. *International Journal of Geographical Information Systems*, 58, 349–366.
- Foody, G.M., 2004. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 70 (5), 627–633.
- Foody, G.M. and Boyd, D.S., 2012. Using volunteered data in land cover map validation: Mapping tropical forests across West Africa. In: *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*. Presented at the Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany: IEEE International, 6207–6208.
- Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F., and Obersteiner, M., 2009. Geo-

Wiki.Org: The use of crowdsourcing to improve global land cover. *Remote Sensing*, 1 (3), 345–354.

Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., van der Velde, M., Kraxner, F., and Obersteiner, M., 2012. Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling & Software*, 31, 110–123.

Justice, C., Belward, A., Morisette, J., Lewis, P., Privette, J., and Baret, F., 2000. Developments in the ‘validation’ of satellite sensor products for the study of the land surface. *International Journal of Remote Sensing*, 21 (17), 3383–3390.

Leung, D. and Newsam, S., 2014. Land cover classification using geo-referenced photos. *Multimedia Tools and Applications*, 1–21.

Pontius, R.G. and Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32 (15), 4407–4429.

Radoux, J., Defourny, P., and Bogaert, P., 2008. Comparison of pixel- and object-based sampling strategies for thematic accuracy assessment. In: *GEOgraphic Object Based Image Analysis for the 21st Century*. Presented at the GEOBIA 2008 - Pixels, Objects, Intelligence, Calgary, Alberta, Canada: ISPRS Archives.

Ramm, F., Topf, J., and Chilton, S., 2011. *OpenstreetMap: Using and Enhancing the Free Map of the World*. Cambridge, England: UIT Cambridge.

See, L., Fritz, S., Perger, C., Schill, C., McCallum, I., Schepaschenko, D., Duerauer, M., Sturm, T., Karner, M., Kraxner, F., and Obersteiner, M., 2015. Harnessing the power of volunteers, the internet and Google Earth to collect and validate global spatial information using Geo-Wiki. *Technological Forecasting and Social Change*.