Digital Earth AFRICA

Digital Earth Africa

Product Validation Strategy

Contents

Document version history

Version	Description	Date of issue	Author
0.1	First draft presented to the DE Africa Technical Advisory Committee	20 February 2020	Dr Meghan Halabisky, DE Africa Project Lead of Validation
0.2	Second draft	15 March 2020	Dr Meghan Halabisky, DE Africa Project Lead

1 Introduction

1.1 What is DE Africa?

Digital Earth Africa (DE Africa) provides a routine, reliable and operational service that will enable African nations to track changes across their countries and the continent in unprecedented detail, and to provide insights on a wide range of issues, including flooding, droughts, soil and coastal erosion, agriculture, forest cover, land use and land cover change, water availability and quality, and changes to human settlements. DE Africa leverages technology and services developed in Australia (implemented as Digital Earth Australia) to deliver a continental-scale platform and programme that democratizes the capacity to process and analyse satellite data. Routine decision-ready products and services will be made available in a model analogous to the operation of a weather service.

1.2 Why do we need validation and what is the intent of this process?

The mission of DE Africa is to process openly accessible and freely available data to produce decision-ready products. Decision-ready products include analysis-ready satellite imagery as well as derived products that map landscape features for the entire continent, such as surface water and vegetation cover. Validation is a fundamental component of any earth observation (EO) product. Validation is the process of independently determining how well a dataset represents the reality on the ground.

The primary goal of validation within the context of DE Africa is to provide confidence in the data products while also highlighting the limitations – where it works and where it fails – in order to enable appropriate application by the users of the product. The validation process is also an opportunity to engage with end users to ensure the products meet their needs.

Validation in the context of DE Africa serves three objectives:

1. To assess the accuracy of the map product while providing estimates of error and bias.
2. To work with DE Africa partner institutions to implement validation efforts.
3. To develop user engagement and uptake of data products throughout the validation process.

1.3 User-centered approach to validation

The DE Africa Technical Roadmap describes the data product development process as:

"an agile, user-centric approach to development that aims to ensure each product that is developed has a practical, real-world application that will enable positive business change for its users. All DE Africa services are designed with a user in mind, guided by the framework of enabling access and use of free and open data. This includes developing different services and platforms that cater to different user capability and user demand, along with diversity in technical capability to interact and use DE Africa's outputs."

A key component of the DE Africa mission is to work closely with the AfriGEO community, and to be responsive to the information needs, challenges and priorities of the African continent. As such, all levels of engagement with users must incorporate feedback before and throughout the validation process. This will allow for the integration of a variety of user perspectives in both the validation process and product development, while serving to create useful data products. The end goal is to increase the number of people who will benefit from DE Africa products and services through a process of co-production.

2 Overview of DE Africa datasets

A primary goal of DE Africa is to allow ingestion, storage, provision and analysis of structured geospatial data, particularly time-series of satellite images processed according to **CEOS**[1] **Analysis Ready Data (CEOS ARD)** standards. CEOS ARD have been processed to a minimum set of requirements and organized into a form that allows immediate analysis with a minimum of additional user effort and interoperability both through time and with other datasets. Analysis ready data from Earth observation satellites provides the foundation for Earth observation analysis and is the basis for the majority of the information products that DE Africa might be expected to produce. DE Africa will steadily increase the breadth and depth of its CEOS ARD collections with the addition of data from a range of satellite sensors and will provide free and open access to these standardised, analysis ready datasets.

A key objective of DE Africa is to use the CEOS ARD collection to produce **first order products** that cover the African continent (at full resolution). First order products map land cover features on the earth. Land cover is the observed biophysical cover on the Earth's surface including trees, shrubs, grasses, soils, exposed rocks and water bodies, as well as human elements such as plantations, crops and built environments. Earth observation data recorded over a period of time allows the observation of land cover dynamics. Classifying these responses provides a robust and repeatable way of characterising land cover types.

Development of data products is guided by identified needs from partner institutions and end users through the Technical Advisory Committee (TAC). The Technical Roadmap highlights and integrates the user-centric product design required to ensure that technical products remain relevant to the needs of DE Africa’s diverse stakeholders. Users can take first order products, such as land use and land cover maps and develop additional **second order products**, such as more detailed classifications, summaries of temporal trends, and estimates of land cover change (e.g. forest loss).

2.1 Product Types

DE Africa first-order products represent either categorical or continuous data and represent static or dynamic time periods. These two product characteristics require different approaches to validation, which needs to be considered in the beginning stages of validation.

[1] Committee on Earth Observation Satellites www.ceos.org/ard

2.1.1 Categorical v. Continuous

Categorical products provide a hard classification of raster pixels into a single class. An example of a categorical dataset is the Water Observations from Space (WOfS), which classifies satellite image pixels into water or non-water pixels. A validation approach of categorical data would compare the classification to the features actually on the ground that represent that class. The results of the comparison are expressed using an error matrix, which compares the predicted values from the mapped product to the reference data.

Continuous datasets provide quantitative estimates of land cover features. For example, *vegetation fractional cover* is an estimate of how much area within a pixel is covered by green or brown vegetation. A validation approach to *vegetation fractional cover* would compare the dataset to measurements of vegetation cover taken from sample units using high resolution imagery or measurements on the ground. The results of the comparison are expressed using statistical analysis, such as linear regression.

2.1.2 Static v. Dynamic

Static datasets are classified images that are not frequently updated (e.g. most traditional land cover maps). They may be derived from a single time period or from an image composite (e.g. least cloudy pixel within a year).

Dynamic datasets aim to capture dynamic environmental processes (e.g. hydrology, phenology). These datasets must be updated with high temporal frequency. Dynamic datasets may require an extra validation step to understand temporal errors (e.g. does the accuracy of the WOfS dataset decline as waterbodies get smaller, due to more mixed pixels or increased bottom reflectance?).

Most of the products envisioned in DE Africa are dynamic datasets derived from a time-series of image scenes. However, traditional remote sensing products have largely been static representations of land features at one moment in time. Matching the temporal window of the reference dataset to the data product of interest is an important consideration. For some features the accuracy of a dataset may change over time and be related to the dynamic nature of the feature.

3 Overview of DE Africa validation approach

3.1 Quantitative validation and standards

Validation is the process of comparing predicted or mapped values to a set of reference data that are considered to (better or more directly) represent the ‘reality’ on the ground. The three main components of validation include **sampling design**, **assessment of reference data** (response design), and creation of an **accuracy assessment** (error matrix) (Stehman et al 2019). Sample design is used to create the placement of the reference sample units. Reference sample units are assessed through interpretation of imagery, measurements taken through in-situ instruments, or visitation on the ground. An accuracy assessment is an accounting of the agreement and disagreement between the mapped value and the reference data.

The process of validation highlights dataset limitations by identifying the apparent disparity between the predictions of the satellite product and the reference observations. This includes the **error** (i.e. errors of omission and commission) and **bias** (e.g. systematic omission or commission of a certain class relative to other classes) in the dataset. If error and bias are not detected in first order products, they can have consequences for downstream uses of these products. Here, the focus is on design-based inference for the estimation of accuracy, however, model-based and Bayesian inference are approaches that are becoming increasingly used and will be considered in the future.

Our goal is to adhere to the following validation best practices (Stehman et al 2019). Validation will be:

1. **Map relevant:** Covers the spatial extent and uses a sample design that meets validation objectives.
2. **Statistically rigorous:** Implements a probability-based sample design.
3. **Quality assured:** Uses protocols to control the validation process to ensure the consistency and accuracy of the reference data.
4. **Reliable:** Results have low uncertainty as determined by low standard errors and report confidence intervals of overall and class accuracy. Acceptable standard error will depend on the application of the data.
5. **Transparent:** All details, positive or negative, that impact results are documented. The sample design, assessment of reference data, and accuracy assessment are well documented.
6. **Reproducible:** The validation can be repeated by following the methods outlined in the documentation and will produce the same results.

Any deviations from the elements listed above, when they cannot be met because of cost or other constraints, will be clearly documented. Selecting specific validation standards, such as the acceptable standard error, sample size, and appropriate sampling method needs to be evaluated through the application development team and informed by the users of beta and provisional products.

3.2 Engaging with users, and fitness for use

Quantitative validation produces an objective measure of the error of a data product, allowing the producers to broadly understand the product. In addition, quantitative validation allows for comparison of products by providing a baseline for improvement. However, quantitative information on its own is not sufficient to understand the usefulness of a product in a particular situation. In practice, *an inaccurate product may be very suitable* for some purposes, and conversely *a highly accurate product may be unsuitable for some purposes*. The fitness for use can only be judged against a specific use case and necessitates that users understand the products. In addition, it requires that the application development team understand the potential uses of the products.

A key outcome of product validation is that users have the confidence and knowledge to adopt and apply the product in appropriate circumstances. As stated above, this necessitates learning by both the creators of the data products and the users of the data product. A process of co-production is one that provides opportunity for shared learning for both creators and users of the data and is an important element that should be at the forefront of all stages of validation. Therefore, DE Africa will engage users in the product validation process to build familiarity with the products and develop an understanding of their potential usefulness and limitations, allowing users to correctly assess the fitness for use in a range of situations Figure 1). At the same time, DE Africa will work to create opportunities for users of the data products to provide feedback to DE Africa that can inform validation standards, improve the quality of data products, and assist in the development of new products.

3.3 Progressive improvement of data products

The quality of an EO product is a function of the satellite imagery and the processing algorithms used to create the dataset. As science progresses, improvements in satellites and processing methods will certainly enhance the quality of EO products. Validation provides an independent check on the performance of satellite sensors and processing algorithms and can be used to identify which dataset best represents what is on the ground. The ongoing phase of the product development cycle involves continuing to work with users to understand how products could be improved and to ensure that the positive business changes enabled by the product are maximised, captured and promoted. Validation must be an almost continuous component of the product development cycle (Figure 1).

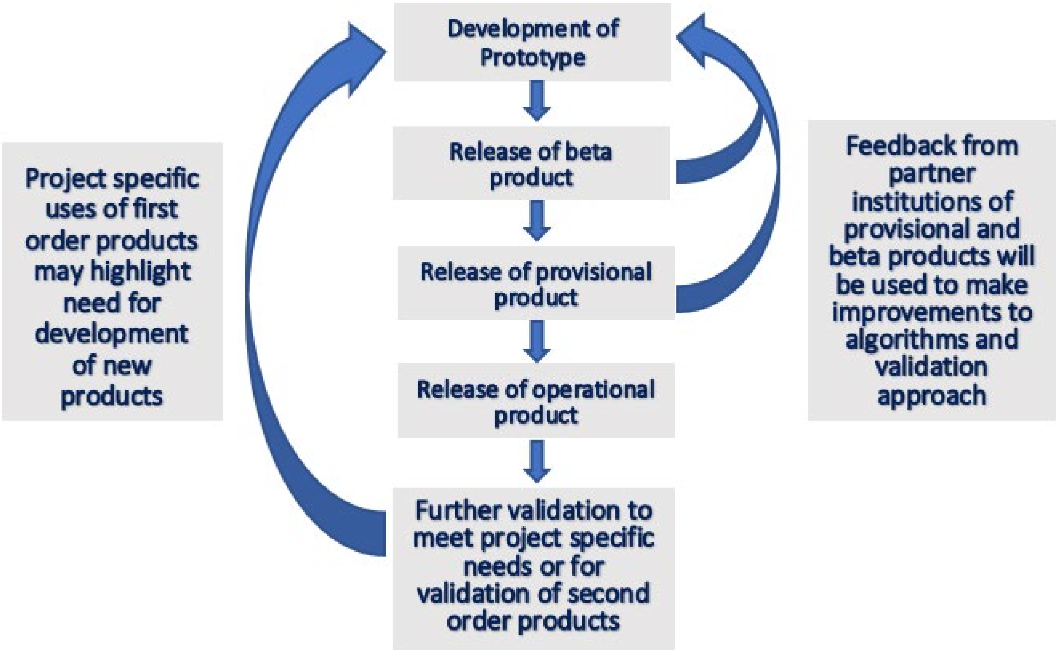

Figure 1. Schematic of integrating partner institution and end user feedback into the release of

While validation is a fundamental component of any EO product, DE Africa recognizes the need to balance the costs and time requirements of validation with the timely release of products. The progression in validation outlined here (Table 1) will allow for timely release of data products, while still adhering to validation best practices outlined above. Through efforts coordinated by DE Africa in the longer term, users will be able to work with the global Earth observation community to influence, improve and standardise processing techniques.

Stage 1 validation	A proven and published algorithm has been used (e.g., an algorithm developed in Australia or Switzerland) but has not yet been validated in Africa. A stage 1 validated product may meet the use requirements for some applications, but most will require further validation.
Stage 2 validation	Dataset is validated using an existing dataset that may not meet high standards for a variety of reasons (e.g. small sample size, contains sample bias)
Stage 3 validation	Dataset is validated using a reference dataset that meets DE Africa validation standards.
Stage 4 validation	Validation that goes beyond Stage 3 to meet specific project or user requirements.

Table 1: Validation stages for release of DE Africa products

3.3.1 Stage 1 validation

Typically, beta products will be released with minimal validation. A beta product represents the first, minimum viable product version of the newly developed concept and typically has not been validated using reference data collected in Africa. Beta products may rely on algorithms developed and validated in Australia or elsewhere, or over a limited area of Africa. Similarities between Africa and Australia in the climate classification are a guide to the transferability of certain Earth observation algorithms between Africa and Australia (Figure 2).

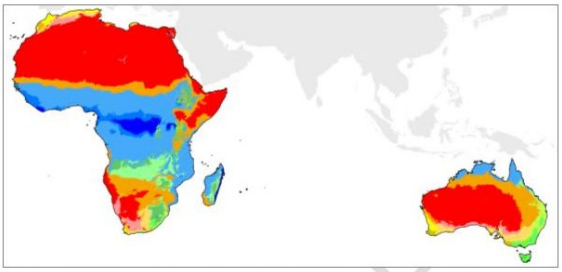

Figure 2. *Köppen-Geiger Climate Classifications for Africa and Australia. Each non-grey colour highlights a different climatic zone, and therefore a potentially different set of Earth observation products that can be applied to that area. Identical colours indicate similar climates, and a potentially similar set of Earth observation products that can be applied. Köppen-Geiger Climate Classification data taken from Beck* et al*. (2018).*[6]

3.3.2 Stage 2 validation

A product will be produced and iteratively improved in consultation with potential users to ensure any major issues are resolved before the product is officially published and routinely produced. Provisional products are data products that are validated using an existing reference dataset collected in Africa that may not meet acceptable validation standards for a variety of reasons (e.g. small sample size, sample bias). In many instances an existing reference dataset can be used for improving the training and validation of a provisional product and provide a first look at how a classification algorithm performs. In some instances, existing reference data will meet the standards of validation and can be used for release of a final product.

3.3.3 Stage 3 validation

Stage 3 validation will ensure that the product meets a level of accuracy that is acceptable for widespread application. Operational products from DE Africa are expected to meet a high standard of validation and follow validation best practices defined here.

3.3.4 Stage 4 validation - project-specific validation

In some cases, Stage 3 validation may not meet the objectives of a specific use and a secondary validation effort may be necessary. Validation at this level will be based on user requirements—likely at sub-regional, national or sub-national scales. Further validation will be tailored to meet specific project requirements (e.g. REDD+, SDG indicators or authoritative data requirements). In addition, second order products may be created that have different validation objectives. Project-specific uses may highlight the need for development of new first order products at the continental-scale or changes to the validation approach.

4 Validation methods

As a first step, the DE Africa application development team will determine the specific validation objectives for a first order product taking into consideration the data type, temporal frame, the spatial extent, and general use (Figure 3). Next, the DE Africa application development team will assess whether existing reference datasets can be used to validate first order products and if they meet DE Africa validation standards. If no existing datasets can be identified, then creation of a reference dataset using high-resolution imagery will be considered first. If reference data cannot be created through interpretation of high-resolution imagery a field campaign will be considered to collect reference data on the ground.

Where possible DE Africa will consider how to group validation efforts to allow for simultaneous validation of multiple datasets, reducing overall validation effort and cost. A framework that can validate multiple datasets must consider a sampling design that is appropriate for multiple classes, design of a hierarchical classification system (e.g. vegetation v. forest/shrub/grass), and a validation assessment tool that can allow for multi-purpose validation assessment (Tsendbazar et al. 2018).

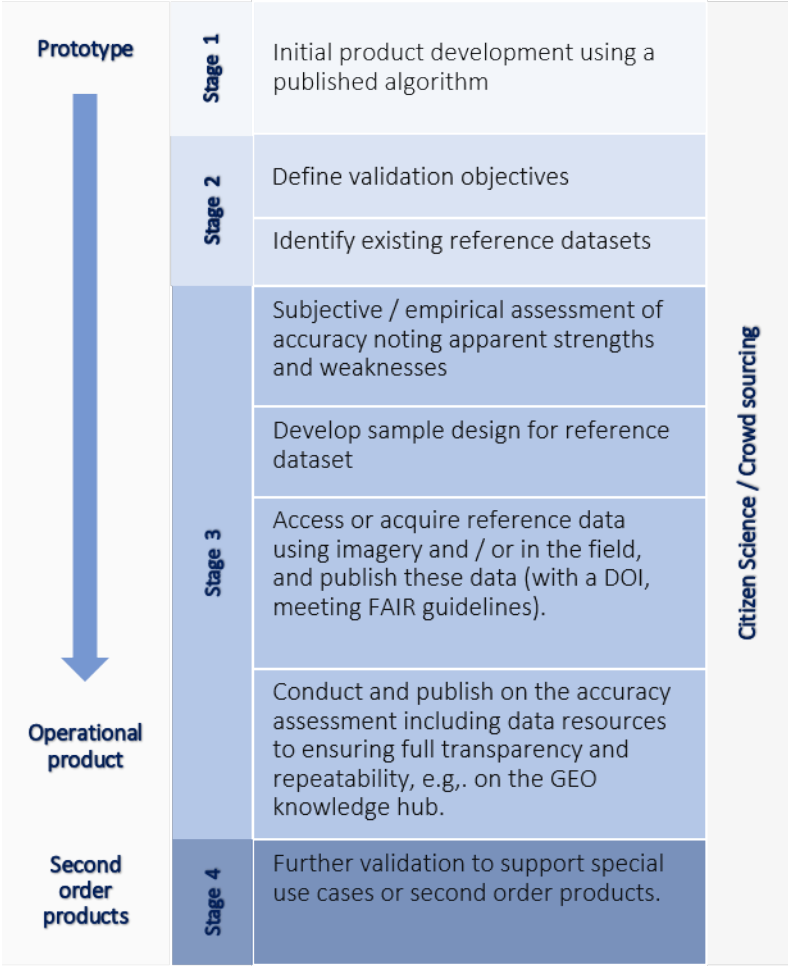

Figure 3. Workflow of validation process and how it aligns

4.1 Sample design

A well-designed sampling strategy provides statistical rigor and considers practical issues that impact cost effectiveness and feasibility. Sample designs can be generated using GIS tools in proprietary software such as ESRI ArcGIS or open access tools like Collect Earth Online.

Desired sampling criteria (Stehman et al 2019):

1. Uses probability sampling methods
2. Is easy to implement
3. Is cost-effective
4. Allows for increasing or decreasing sample size
5. Meets requirements of standard error for its intended use
6. Does not create bias or allows for correction of bias
7. Is spatially well-distributed across the continent

For guidance on developing a sample framework that meets the above sampling criteria see Appendix A.

4.2 Assessment of reference data (response design)

4.2.1 Validation assessment using existing reference datasets

The use of existing reference datasets can significantly cut costs of validation. Therefore, identification of existing reference datasets and evaluation if they fit the purpose of the dataset should be a first step in the validation process (e.g. does it match the spatial extent of the data product). If an existing dataset meets the validation best practices outlined here then it will be used in place of collecting new training and/or validation data. If an existing dataset marginally meets the standards of validation it will be used for development and validation of a provisional product release. In some cases, use of existing datasets could be supplemented to increase the sample size in general or for a particular class of interest.

Identified potential sources of existing reference datasets include Radiant Earth Foundation and NASA. Radiant Earth Foundation is “a non-profit organization working to empower organizations and individuals globally with open machine learning and Earth observation (EO) data, standards, and tools to address the world’s most critical international development challenges. [Radiant Earth] has created a common repository of geospatial training data and trained ML models to serve as a resource for a community of practice around benchmark training datasets and models to advance applications of ML techniques on Earth observations.” Reference datasets created by Radiant Earth are freely available for training and validation of any DE Africa product. DE Africa will work with Radiant Earth to align efforts where possible. Radiant Earth can serve as a secondary repository for any reference data collected through DE Africa.

4.2.2 Validation assessment using imagery

The use of imagery to assess sample units is an acceptable approach to create reference data that represent the “truth” on the ground. Ideally, the imagery used is a higher spatial resolution than the resolution of the dataset. For example, Sentinel-2 imagery with a pixel resolution of 10 meters could be used to assess the accuracy of a Landsat classification map with a pixel resolution of 30 meters. Assessing sample units using higher resolution imagery, as opposed to field data collection, can greatly reduce the costs of reference data collection while also putting fewer constraints on the placement of sample units (e.g. private property). This is acceptable as long as the features being assessed can be clearly identified in the referenced imagery.

In order to conduct validation using existing imagery an analyst will evaluate the sample unit using an image and select the feature that represents the sample unit. Sample units are most typically a pixel (or pixel unit), a polygon, or a point. In selecting a sample unit, it is important to understand the spatial alignment of the data product and the higher resolution imagery used to create a reference dataset. If datasets are not well aligned, then pixel units (e.g. 3 x 3 pixels) or polygons should be used over single pixel or point comparisons. For categorical datasets classes must be clearly defined so that photo-interpreters can translate what they see in the imagery to the correct class definition. For example, the definition of a forest cover class could be defined as any pixel where 60% of the pixel area or greater is covered in trees. A photo-interpreter could then decide if the sample unit (in this case a pixel) meets this definition.

4.2.2.1 Web-based validation assessment tools

Several web-based tools provide an easy to use platform for image interpretation of sample units. Two open access web-based tools are Collect Earth Online and Laco-wiki. Collect Earth Online and Laco-wiki are both flexible, customizable validation assessment tools that allows users to build a tool that fits their needs and can be accessed anywhere there is internet connection. Users can access a wide variety of satellite and aerial images through Google and Bing as well as OpenStreetMap. Additionally, users can upload their own imagery or connect to an image subscription service (e.g. Planet).

Collect Earth Online is “a joint NASA and USAID program in partnership with regional technical organizations around the world - and the FAO as a tool for use in projects that require land cover and/or land use reference data” (Figure 4). Laco-wiki is a project founded in 2014 and now operated by the International Institute for Applied Systems Analysis (IIASA) and GeoVille GmbH. The project was initially funded by the Austrian Research Promotion Agency (FFG) in the Austrian Space Applications Programme (ASAP). Each of these web-based tools allow users to establish permissions on who can assess the sample units and who can download the data. Assessment of sample units and downloading of data can occur anywhere there is access to the internet.

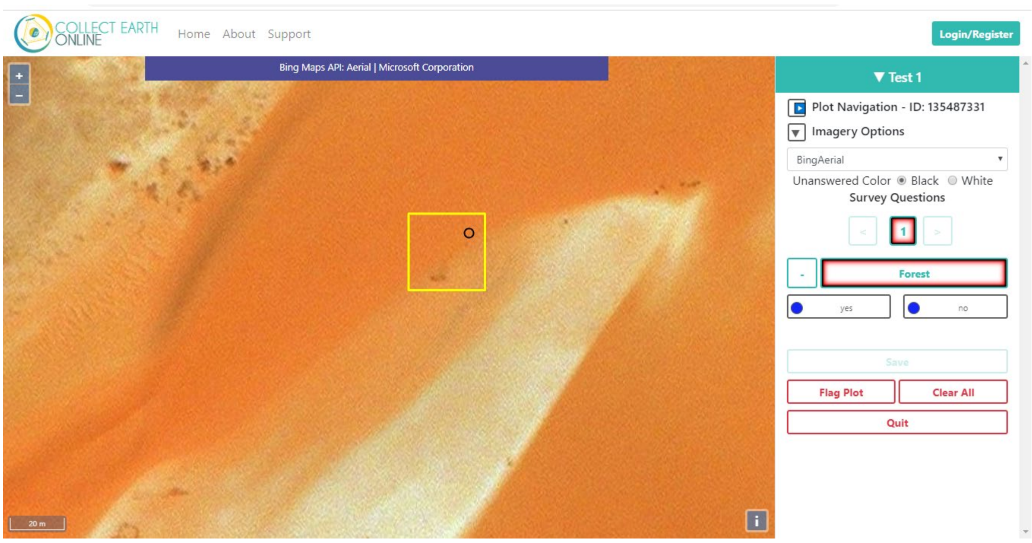

Figure 4. Example of assessing sample units using Collect Earth

4.2.3 Validation assessment using field data

Field data collection can be costly and requires an organized effort. Therefore, it should only be used if it is determined that image interpretation is not sufficient for development of reference data. Field validation methods may include identification of features that are not easily detectable in imagery (e.g. species of grass), collecting plot-level data, or in-situ measurements that can be used for modelling phenomena beyond land cover classification, such as a water quality index.

For many data products a field sample can provide clarity of what is on the ground. Some land cover classes may not be interpretable using high-resolution imagery (e.g. forested wetland). In addition, imagery may not be available for some areas or may not temporally match the time period of the data product. In some instances, it may be worthwhile to field validate a percentage of sample units. Field data collection can provide training and calibration of photo-interpreters as well as provide a better understanding of error attribution (where the error occurs).

In-situ data can be used for calibration and validation of modelled products. Development of a calibration-validation network that can be used to collect in-situ measurements and plot-level data may facilitate the extension of existing modelling methods into Africa. Specific protocols for field data collection should be clearly outlined before undertaking a field campaign.

4.2.4 Storage and accessibility of reference data

All reference data collected for DE Africa will be open and accessible via the DE Africa infrastructure and on open data platforms such as Radiant Earth. Reference data will include metadata that details the sampling design, references datasets used for stratification or masking, lists the data used for validation, and the date of validation.

4.3 Accuracy assessment

An accuracy assessment is a standard component of a validation process. An accuracy assessment does more than just describe the quality of a map - it provides a means to enhance its usefulness (Stehman et al 2019). An accuracy assessment compares the predicted values from the mapped product to the reference data using an error matrix (aka confusion matrix).

Any validation approach will need to report the results of the accuracy assessment which should include:

1.) producer's accuracy (i.e. omission) for each class

2.) user's accuracy (i.e. commission) for each class

3.) standard error or confidence intervals quantifying the uncertainty of accuracy and area estimates for each class

4.) overall accuracy of the dataset

5.) spatial accuracy over the continent. An accuracy assessment stratified by region or climate zone can help users understand areas where there are particular mapping challenges.

In addition, an accuracy assessment will include a description of error and bias written out in laymen terms. It should include a detailed assessment of where and why the error occurs for each class and include a description of limitations and considerations in the application of the data product.

5 5. Additional project-specific validation

5.1 Further validation of first order products

There may be some instances where additional validation effort at the country or regional level is required. For example, it may be important for a specific project to assess the area of a landscape feature, and the validation of the first order product does not contain enough sample units to provide certainty in the area estimates. DE Africa can assist users in further validation of first order products.

5.2 Validation of second order products

It is expected that first order products will be translated into second order products. Second order products may include but are not limited to; more detailed classifications, summaries of temporal trends, and estimates of land cover change (e.g. forest loss). These products may require a new validation strategy to assess the accuracy and uncertainty of the phenomena being evaluated. For example, validation of the first order WOfS product may not be sufficient if WOfS is being translated into a second order product, which tracks water quality for a particular region. DE Africa can assist users in validation of second order products. Second order products that have widespread applicability across the continent could be developed as first order products.

5.3 Validation support from DE Africa

For follow-up validation, GIS tools can be used (e.g. ESRI ArcGIS), or web-based platforms (Collect Earth Online and Laco-Wiki), which offer easy to use tools to assist with development of a sample design. Both Collect Earth Online and Laco-Wiki provide tools to create a probability-based sample design and a user guide to assist users through this process. DE Africa can provide general guidance on second order validation. DE Africa can provide assistance in the form of:

- Assistance with sampling design
- Support on validation assessment tools
- Training on validation methods including image interpretation and in-situ measurements.

6 6. Citizen Science

A main objective for DE Africa is to promote the use of DE Africa products and engage with the end user. Therefore, DE Africa will seek citizen science opportunities for every data product released. Citizen science serves several purposes, educating and engaging with citizens, reducing the cost of data collection, and providing contextual feedback for improvement on mapping algorithms. When well-designed, citizen science data can be used for training and validation of DE Africa data products (Leibovici et al 2017). The challenge with using citizen science data for validation is assessing the quality of data created by a diverse group with varied experience and training. However, there are quality assurance measures that can be taken to improve the quality of the reference data created from a citizen science campaign, such as calibrating and validating individual contributors (Leibovici et al 2017).

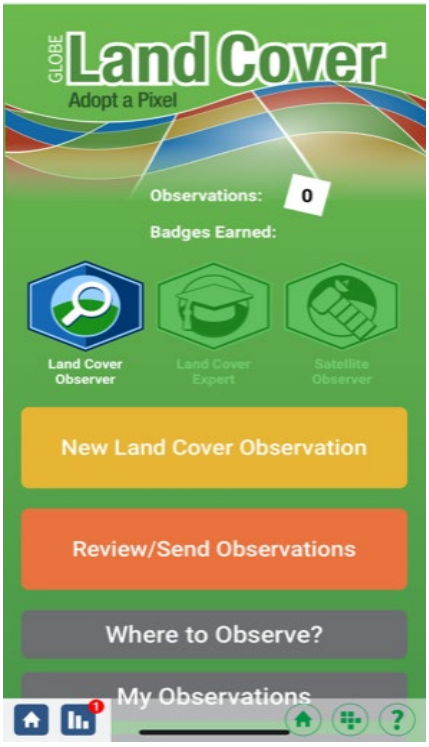

Figure 4: Example of NASA's citizen science program phone app, Globe Observer – Land Cover Adopt a Pixel.

The numbers of users and citizen science data collected in Africa is growing. Where possible DE Africa will leverage existing citizen science data collection tools such as the NASA Globe Observer (Figure 4), Collect Earth Online, or Laco-wiki. NASA Globe Observer has a well-designed mobile application for data collection and an established network within Africa. Three mobile applications developed by the NASA Globe Observer that have immediate relevance to DE Africa are the Globe LandCover (Figure 4), Globe Mosquito Habitat Mapper, and Globe Trees.

7 Putting validation into practice

A key objective for DE Africa is for continental data products to be taken up and used at the country level for decision-making, user engagement, development of additional use cases, innovation and policy. However, how a country will use these products can take several paths, and whether they will have different requirements and/or regulations related to certainty levels in alignment with authoritative data, for example, is unknown. The strategy needs to be put into practice by developing an implementation approach in consultation with Technical Advisory Committee (TAC) members, and one that will utilize skills and resources available through the DE Africa Program Office and partnerships with African-based institutions.

The implementation approach will align to the DE Africa partnership strategy and operational model such that those partner organizations supporting user engagement and applications will likely be utilized for implementing aspects of the validation strategy. Ultimately, the implementation approach needs to be co-designed with partner institutions, and therefore, it is likely that a small working group will be established inclusive of interested organizations from the TAC to develop the approach and user guide for how validation takes place.

8 References

Leibovici, Didier & Williams, Jamie & Rosser, Julian & Hodges, Crona & Chapman, Colin & Higgins, Chris & Jackson, Mike. (2017). Earth Observation for Citizen Science Validation, or Citizen Science for Earth Observation Validation? The Role of Quality Assurance of Volunteered Observations. Data. 2. 10.3390/data2040035.

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

Pickering, Jeffrey & Stehman, Stephen & Tyukavina, Alexandra & Potapov, Peter & Watt, Pete & Jantz, Samuel & Bholanath, Pradeepa & Hansen, Matthew. (2019). Quantifying the trade-off between cost and precision in estimating area of forest loss and degradation using probability sampling in Guyana. Remote Sensing of Environment. 221. 122-135. 10.1016/j.rse.2018.11.018.

Stehman, Stephen & Foody, Giles. (2019). Key issues in rigorous accuracy assessment of land cover products. Remote Sensing of Environment. 231. 111199. 10.1016/j.rse.2019.05.018.

Tsendbazar N-E, Herold M, de Bruin S, Lesiv M, Fritz S, Van De Kerchove R, Buchhorn M, Dürauer M, et al. (2018). *Developing and applying a multi-purpose land cover validation dataset for Africa.* Remote Sensing of Environment 219: 298-309. DOI:10.1016/j.rse.2018.10.025.

Appendix A: Guidance on developing a sample framework

A sample framework should consider the following questions (Stehman and Foody 2019):

1.) What is the appropriate sampling method?

The most common probability-based sample methods are simple random sampling, stratified random sampling, and systematic sampling.

2.) Will imagery or a field campaign be used or both? If imagery is used, what is the source?

Imagery used for creation of reference data should ideally be at a higher pixel resolution than the data product being validated. The pixel resolution should match the resolution of the dataset at a minimum. A field campaign is warranted if the target features cannot be assessed in imagery. A combined approach can be useful where sample units are first assessed using imagery and only those that are hard to assess in imagery are visited on the ground.

3.) Should clusters be used?

Cluster sampling is the practice of grouping sampling units together. Cluster sampling may be appropriate if there is a large field effort and travel to sites is costly or if individual image scenes must be purchased for high resolution image interpretation.

4.) Should strata be used?

Sampling stratification allows for distributing your sample units using strata (e.g. climate zone, features classes, region). Stratification is often used to ensure a minimum number of sample units within a rare class.

5.) What is the spatial scale that needs to be considered?

Is the dataset categorical or continuous? If it is categorical what are the classes that need to be defined. If it is continuous (e.g. fractional cover) how will the sample unit capture sub-pixel measurements (e.g. grid, transects).

6.) What is the temporal scale that needs to be considered?

Is the dataset static or dynamic? Does the temporal window of the reference data match the dataset? If the dataset is dynamic is it important that the sample design capture the temporal variation?

7.) What is the sample unit?

There are three common sample units; a point, a polygon, a pixel, or a group of pixels.

8.) What is the sample size?

How many samples are required to capture the variability within the dataset and reduce accuracy assessment uncertainty? As a general rule there should be a minimum of 50 samples per class.

9.) What is the sample frame?

Does the sampling cover the entire continent or only specific regions or areas (e.g. coastlines)? Do certain areas need to be masked out from the sample frame. For example, one may want to restrict field validation to lands with public access or areas near roadways.

10.) Are there any other constraints to consider?

It is important to think through any constraints that may impact the sample design. For example, are there enough cloud-free images available within the required temporal window for assessment of sample units?

Appendix B: Example – Developing a validation approach for WOfS in Africa

Water observations from space (WOfS) is a dynamic dataset that maps surface water across the continent at a 30m pixel scale using a time series of Landsat satellite imagery. The validation approach for the WOfS dataset is provided here as an example. The methods outlined here are not the validation methods that will be used for the WOfS product, but instead provide an example of what a validation approach as outlined in this document looks like in practice.

Step 1: Initial development of the dataset

WOfS uses a decision tree approach of multiple spectral indices calculated from Landsat satellite imagery and is based on training data collected in Australia.

Step 2: Define validation objectives and identify end user applications.

The objective of validation for the WOfS dataset is to determine the accuracy of mapping surface water throughout the continent, while understanding the associated error and bias. WOfS is a categorical (water v. non-water) dynamic dataset.

Step 3: Identify any existing reference datasets

We have identified one possible existing reference dataset. Radiant Earth is in the final stages of developing a global land cover reference dataset, which includes a water class. It will be made freely available for use in Spring 2020. The water class can be used for an initial validation of the WOfS product. When the reference data and metadata are released they will be assessed for compatibility with the above objectives. If the dataset is determined to be suitable for training and/or validation of the WOfS product it will be re-released as a provisional product.

Step 4: Develop sampling design to create reference dataset

1.) What is the appropriate sampling method?

We will use a stratified random sample.

3.) Will imagery or a field campaign be used or both? If imagery is used, what is the source?

We will use image interpretation for creation of a reference dataset. No field campaign is deemed necessary at this time. We will use Sentinel-2 imagery with a 10m pixel resolution. We will also use high-resolution satellite imagery, such as DigiGlobe as additional reference to determine the land features of our sample unit.

4.) Should clusters be used?

We will not use clusters as Sentinel-2 imagery is freely available throughout the continent. We will do an initial analysis to determine if there are any restrictions on availability due to cloud cover.

5.) Should strata be used?

Because water is a rare class we will consider stratifying our sample in two ways. The first is to use the Köppen-Geiger climate zones and the second is to stratify our sample by a simple water mask. We will evaluate existing water masks provided as part of Sentinel-2 and Landsat, but will likely create a water mask by thresholding of a simple water index such as the Normalized Difference Water Index (NDWI). We will stratify our sample points using the water mask. This is to balance our sample points into 1/3 water and 2/3 non-water points.

6.) What is the temporal scale that needs to be considered?

Surface water is dynamic and therefore the imagery used for reference data creation must match a certain WOfS observation give or take 10 days. The accuracy of WOfS may vary in wet and dry seasons. We will consider either a.) selecting only one time period in the wet season b.) selecting a time period in the wet season and one in the dry season or, C.) incorporating a range of Sentinel-2 image scenes across both wet and dry seasons. Selection of the temporal range will depend on how the data is applied.

7.) What is the sample unit?

We will consider a 3 x 3 or a 9 x 9 Sentinel-2 pixel unit for our sample unit. If we classify each pixel within our sample unit we will be able to use the reference dataset in the future if WOfS algorithms is applied to Sentinel-2.

8.) What is the sample size?

We used the following logic to determine sample size.

200 samples (class – 100 in the water class, 100 in non-water) *8 (climate zones) * 2 time periods (wet, dry) = 3,200 sample units.

9.) What is the sample frame?

The entire continent.

10.) What other constraints exist?

We will constrain sampling to clear Sentinel-2 images and check to determine there are enough cloud-free images available covering the continent.

Step 5: Assess reference dataset in field or using imagery

1.) What are the class definitions used for labelling of sample units?

We will develop a class definition for labelling of our sample units. Currently, we are using the following class labels; vegetation, bare earth, vegetation/water mix, clear water, & not clear water. We will develop a definition with image examples as a guide for each of these classes prior to any assessment effort.

2.) Who will do the assessment of reference data?

We will work with analysts at partner institutions to conduct the assessment of reference data using Sentinel-2 imagery. We will use Collect Earth Online for assessment of reference data. All reference data will be stored in Collect Earth Online and periodically downloaded and checked for quality assurance.

Step 6: Conduct an accuracy assessment

We will compare the reference data to the predicted values through an accuracy assessment, interpret the results of the accuracy assessment where we highlight the limitations of the dataset. We will publish the results.

Step 7: Identify citizen science opportunities

We have identified a scout jamboree in August, 2020 in Kenya. We will explore the potential of using the scout jamboree to initiate a citizen science campaign using the NASA Globe Observer app. This app can be used to collect land cover data, which can be used to provide context and potential training and validation for the WOfS product.

Step 8: Conduct any further validation

We are considering a second order product based off the WOfS dataset that models water quality. Development of a water quality product would require validation using in situ data. There is an opportunity to partner with GEO AquaWatch which is developing water quality demonstration projects in developing countries. We have been discussing with a local partner university in Ghana the possibility of installing an AERONET-OC station on Lake Volta.