



REPORT ON THE CROWDSOURCING CAMPAIGNS

Deliverable 3 (WP4)



Table of Contents

1 Introduction	1
2 Overview of the Campaigns	1
3 Campaign in Kenya	1
4 Campaign in Gabon	4
5 Campaign in Ivory Coast.....	6
6 Data Collection in South Africa	8
7 Summary of Lessons Learned	12
References	12

1 Introduction

This is the third deliverable of the CrowdVal project, which summarizes the crowdsourcing campaigns that took place during the CrowdVal project. Based on the experiences with these campaigns, we provide some generic lessons learned.

2 Overview of the Campaigns

Table 1 summarizes the campaigns that took place during the CrowdVal project. Three of the campaigns (Kenya, Gabon and Ivory Coast) had both an online LACO-Wiki data collection component (i.e., by visual interpretation) and a field-based component. In the case of South Africa, the data were only collected online using visual interpretation.

Table 1: Overview of the campaigns and data collected

Campaigns	Date of workshop	Date over which data were collected	Number of points collected in situ	Number of points collected online
Kenya	27-28 Sep 2018	Until Mar 2019	766 ¹	10636 ²
Gabon	12-13 Dec 2019	Until Apr 2019	293	1949
Ivory Coast	5-7 Feb 2019	Until Apr 2019	686	6128 ³
South Africa	N/A	June 2019	N/A	23000 ⁴

¹ Note that 485 points were collected previously and are displayed in the CrowdVal branch of Geo-Wiki

² There were 4369 points validated twice and then an additional 1898 disagreeing areas validated by an expert

³ There were 2428 points validated twice and then an additional 1272 disagreeing areas validated by 2 experts

⁴ There were 920 clusters with 92000 10m x 10m pixels validated, which is 23000 20m x 20m pixels

Each of the campaigns is now described in more detail in the sections that follow.

3 Campaign in Kenya

3.1 Training workshop

The training workshop was organized with the Regional Centre for Mapping of Resources for Development (RCMRD) in Nairobi, Kenya. In total, there were 28 participants, and the workshop was run by Steffen Fritz and Myroslava Lesiv of IIASA. The first 1.5 days of the workshop were devoted to training the participants in visual interpretation using LACO-Wiki. Myroslava Lesiv is an expert in visual interpretation and showed the participants different types of land cover in Kenya and what features to look for during the visual interpretation. There was also an interactive session in which participants could discuss examples that they encountered as they performed the visual interpretation, so that the whole group could learn from a variety of examples.

Before the start of the workshop, a systematic sample was created across Kenya at a spacing of roughly 12 km apart, resulting in a total of 4369 samples units to be interpreted. In the first session, blind validation was used; this is a feature in LACO-Wiki where users are asked to look at the satellite imagery and to choose a land cover type from a list. During the second session, the same sample was validated using enhanced plausibility, where the land cover type of the pixel was indicated; users could then accept that value or correct the land cover if it was wrong based on the visual interpretation. During the afternoon of day 2, Steffen Fritz gave an overview of LACO-Wiki Mobile (still at an early stage of the project) and users downloaded the app on their phone and tried it in the field.

3.2 Completion of the data collection

In total there were 4369 locations of the systematic sample that were validated twice using LACO-Wiki. Myroslava Lesiv then compared the two samples, identifying the number of locations at which disagreement occurred, which was 1898. Myroslava then went through each of these disagreeing locations and visually interpreted the location in order to produce a high quality validation data set for Kenya. This validation data set is available on the CrowdVal branch of Geo-Wiki (see Figure 1) for visualization and downloading. This sample has been used to calculate the spatial accuracy of the ESA 20 m prototype land cover map for Kenya as outlined in Deliverable D4.

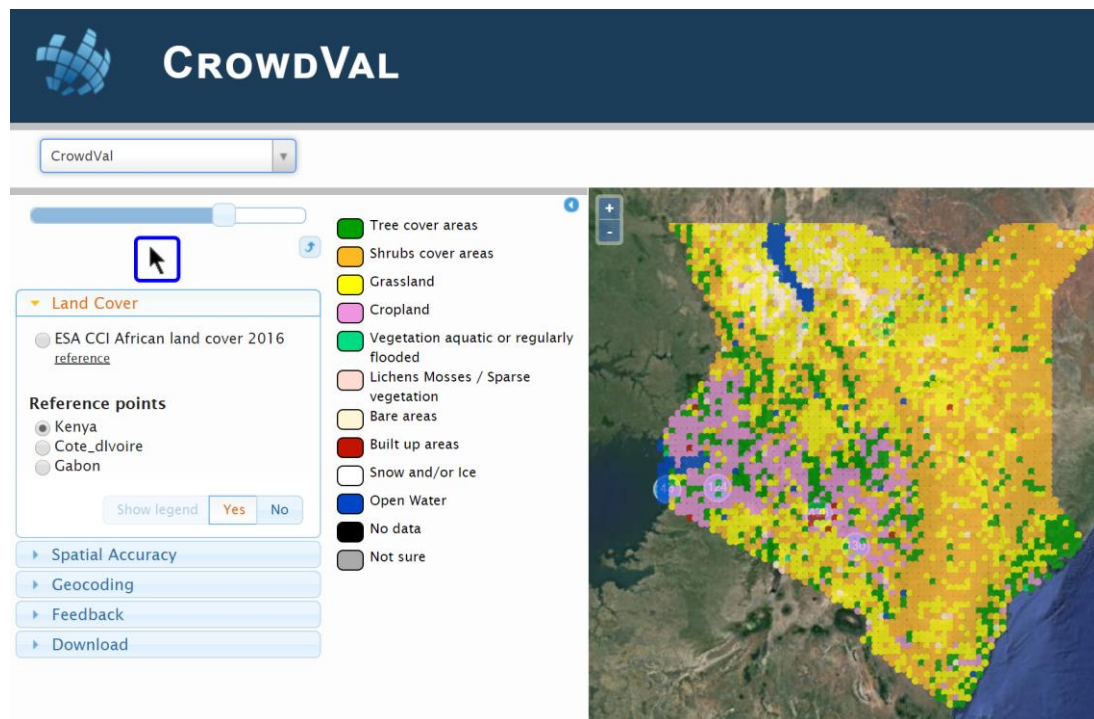


Figure 1: Systematic sample of locations visually interpreted for Kenya

During the workshop, LACO-Wiki Mobile was tested but there were various problems with the app. Moreover, the opportunistic functionality was not yet added. For this reason, some of the in situ data were collected by Myroslava Lesiv by taking geotagged photographs from her mobile phone. Figure 2 shows some examples of different land cover types taken while in Kenya.

A second trip to Kenya took place in March 2019. Additional data were collected with the mobile app during this trip and afterwards by local Kenyans, who were paid to collect the data. These in situ data have been uploaded to the CrowdVal branch of Geo-Wiki. The photographs can be viewed by clicking on individual points on the map.

In addition, 485 geo-tagged photographs with land cover types were added to the CrowdVal branch of Geo-Wiki. These were taken with the Geo-Wiki Pictures mobile app, which is the pre-cursor app to LACO-Wiki Mobile. This additional data set provides a further, rich source of data for training and verification of land cover.

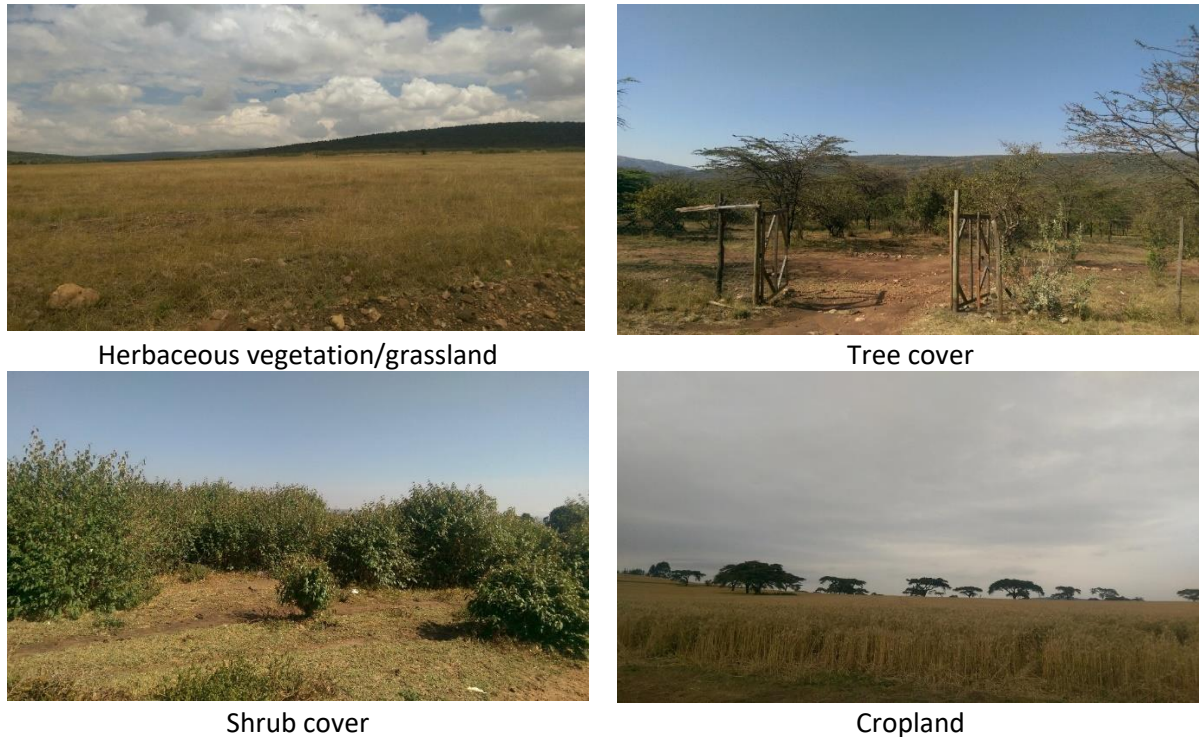


Figure 2: Examples of in situ land cover data and geo-tagged photographs collected by Myroslava Lesiv while in Kenya during the training workshop

3.3 Issues related to the campaign

There were some technical issues related to the slow internet connection, which meant that working with LACO-Wiki online was slow. However, there was enough time during two-day session to collect the validation data. As mentioned in section 3.1, the participants used LACO-Wiki first with blind validation (so users had to choose the land cover type), followed by enhanced plausibility validation (where the land cover at the location was revealed and users needed to accept or correct the land cover type). This experiment showed that enhanced plausibility validation resulted in higher overall accuracy. This is not unsurprising since being shown the land cover at a given location may bias the user towards choosing that land cover type more often. However, the advantage of enhanced plausibility validation is that for mixed pixels (e.g., 50% trees and 50% grass), the participants could select a land cover type based on the land cover map instead of choosing a class randomly from a few different options. This worked particularly well after participants were trained by Myroslava Lesiv in visual interpretation.

Myroslava observed that being a resident of Kenya did not mean that those taking part in the workshop had knowledge about the country's land cover. She noted some participants did not have a proper understanding of landscapes and hence visual interpretation was very difficult for them, e.g., they initially classified water as tree cover. However, she observed improvements in the performance of the participants over the course of the two-day workshop as they learned to recognize different landscapes. There were only a few participants who had travelled across the country for research purposes and shared their knowledge with others during the workshop. Hence there is an advantage in running such workshops with the participation of few good regional experts.

In terms of the mobile app, there were problems with some of the older mobile phones that could not run the Android version of the application during the workshop. At this point in the project (Sep 2018), an iOS version of the app was not yet available.

4 Campaign in Gabon

4.1 Training workshop

In total there were 20 participants at the workshop, which was organized by Stephan Pietsch (IIASA) in coordination with L'Université des Sciences et Techniques de Masuku (USTM) in Franceville, Gabon. The two-day workshop was split into two morning sessions, which were devoted to LACO-Wiki online and visual interpretation. The two afternoon sessions of the workshop were focused on in situ data collection.

Similar to the Kenyan workshop, Myroslava provided an introduction to visual interpretation. However, she was not present in Gabon but did this remotely via skype, followed by an interactive session in which participants could send the locations to Myroslava by email or skype chat. She then looked up these locations and fed back the land cover type to the group, along with the reasons why. Through this process, the group improved their visual interpretation skills.

For Gabon, a systematic sample was created using the same 12 km spacing as for Kenya, resulting in 1948 points to visually interpret. In the workshop, this sample was accessible from LACO-Wiki and was then classified only once by the participants using blind validation, mainly due to limited internet access.

This sample has been used to calculate the spatial accuracy of the ESA 20 m prototype land cover map for Gabon as outlined in Deliverable D4.

4.2 Completion of data collection

The participants were able to classify 1948 locations during the workshop. The remaining locations were classified after the workshop was completed and then the sample was quality checked by Myroslava Lesiv, which is shown in Figure 3. This sample can be viewed from the Geo-Wiki branch of CrowdVal and downloaded.

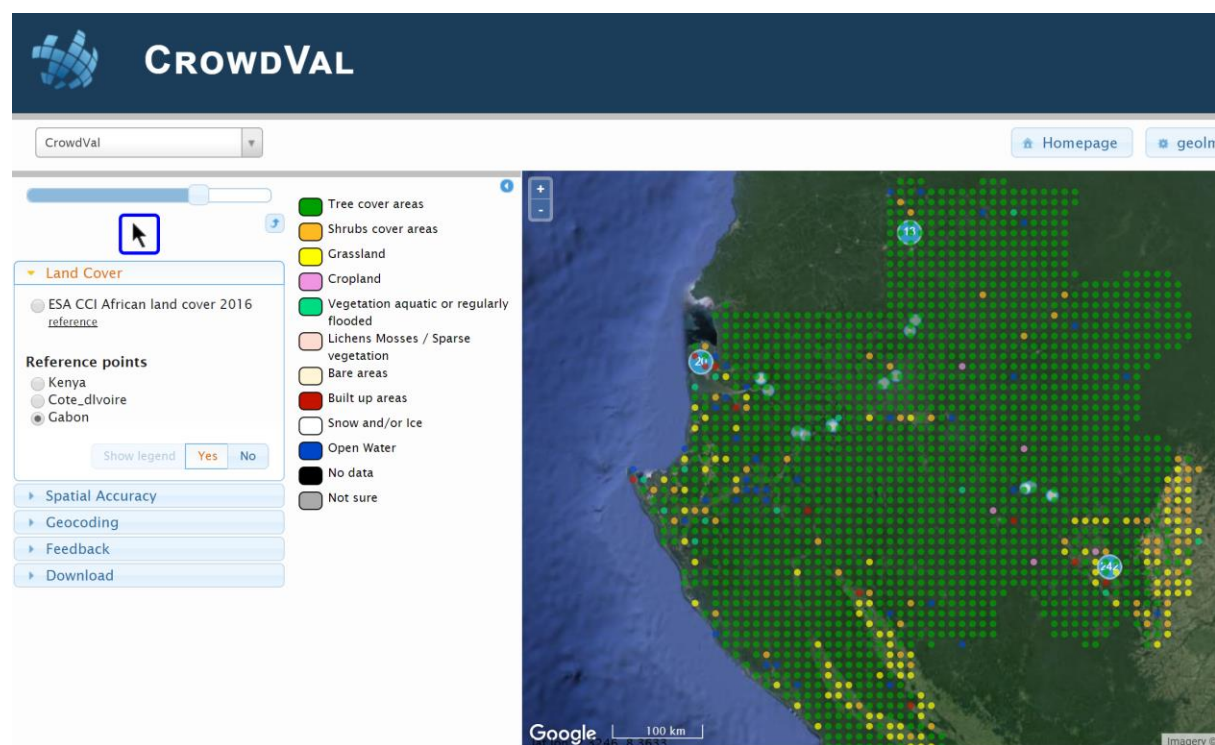


Figure 3: Systematic sample of locations visually interpreted for Gabon

Figure 4 shows the in situ data collected in Gabon during the workshop and afterwards until April 2019. Many of the locations are concentrated in the two urban centers or along roads. These data can be viewed from the CrowdVal branch of Geo-Wiki, clicking on individual locations to view the geo-tagged photographs (see the example shown in Figure 5). The in situ data can also be downloaded from Geo-Wiki.

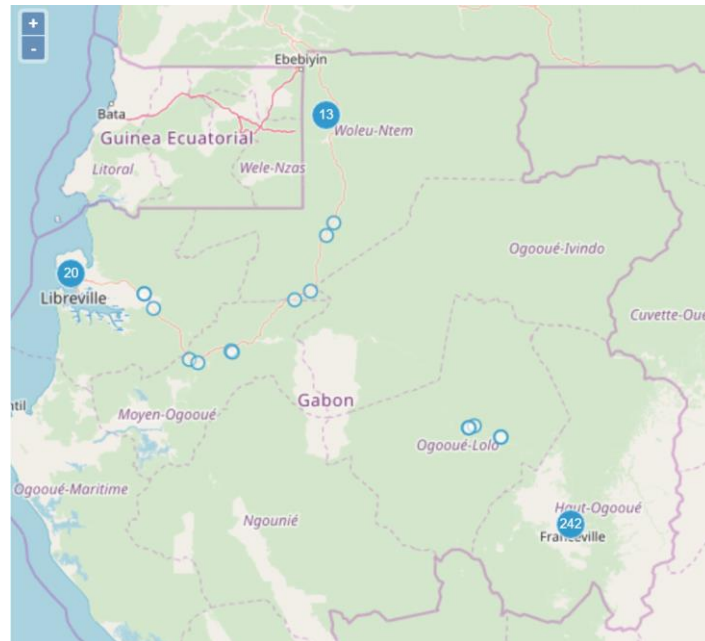


Figure 4: Locations of data collected in situ in Gabon shown in the CrowdVal branch of Geo-Wiki

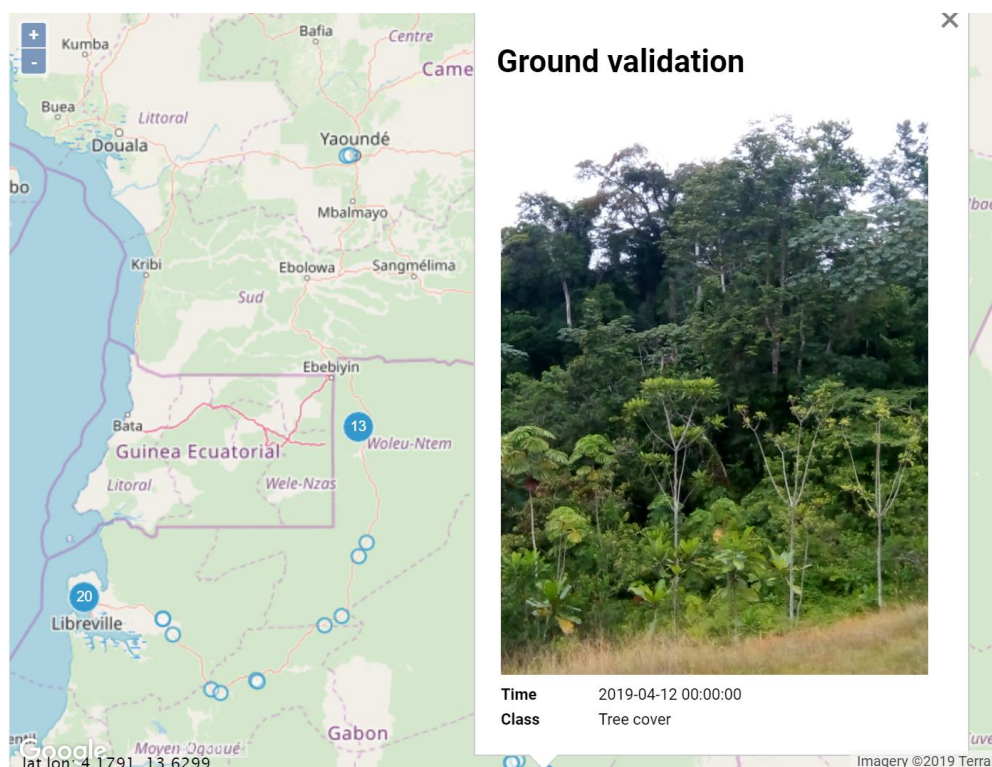


Figure 5: An example of a geo-tagged photograph recorded with the mobile app in Gabon, which is viewed from the CrowdVal branch of Geo-Wiki

4.3 Issues related to the campaign

The remote training of participants in visual interpretation and in the online version of LACO-Wiki by Myroslava Lesiv was successful. Although the internet was slow and the workshop was in French, translations were made by Stephan Pietsch to the participants, which worked well. Much of the in situ data were collected in close proximity to the major urban centers or along roads in an opportunistic fashion. The app did not work on all mobile phones, as many participants have older smart phones. Payments per point were an effective way to collect in situ data in a country like Gabon, where small, monetary incentives can be very attractive. Examination of some of the geo-tagged photographs indicated that the land cover was classified in the distance rather than at the location where the person was standing. Having some type of offset feature or the ability to move the point on the map would be one way to improve the accuracy of the land cover registration. More emphasis on how to do this in the training session may also have helped. Finally, some of this advice could be built into the app, to guide the user in classifying the land cover correctly and at the correct location.

5 Campaign in Ivory Coast

5.1 Training workshop

The final workshop that took place as part of the CrowdVal project was in Abidjan, Ivory Coast, during 5-7 Feb 2019. The workshop was organized in collaboration with the Université Félix-Houphouët-Boigny, which recruited the participants among the staff and the students. In total, forty-seven people from the university attended the workshop. The first and third day of the workshop involved an introduction to LACO-Wiki and data collection through visual interpretation.

As with the other workshops, a systematic sample was created across Ivory Coast with a 12 km spacing, resulting in a sample with 2,428 locations to be visually interpreted. Myroslava Lesiv provided training in visual interpretation remotely, in the same way as in the Gabon workshop, and provided interactive support to participants during the session. During the workshop, the sample was interpreted twice, once with blind validation and the second with enhanced plausibility, similar to the Kenya workshop.

Day two of the workshop was devoted to in situ data collection in the field. This involved both guided data collection of a dense sample of 500 points placed in a 10 km² area in the outskirts of Abidjan (Djorogobité) for the purpose of the workshop, as well as opportunistic data collection, i.e., land cover data collection at any location.

5.2 Completion of the data collection

The two visually interpreted samples from LACO-Wiki were compared for agreement; in total there were 1272 disagreeing samples. This is not surprising given the difficult landscapes to classify (see section 5.3 regarding issues related to the campaign). Brice Mora (of CS) and Myroslava Lesiv examined the disagreeing areas in LACO-Wiki to produce a final systematic sample for Ivory Coast, which can be viewed on the CrowdVal branch of Geo-Wiki (see Figure 6) as well as downloaded. This sample has been used to calculate the spatial accuracy of the ESA 20 m prototype land cover map for Ivory Coast as outlined in Deliverable D4.

Some of the in situ data were collected during the workshop but data collection continued until the end of April 2019. In total, 686 in situ data points were collected, which are roughly distributed across three locations in Ivory Coast (see Figure 7). These points can be viewed and downloaded from the CrowdVal branch of Geo-Wiki.

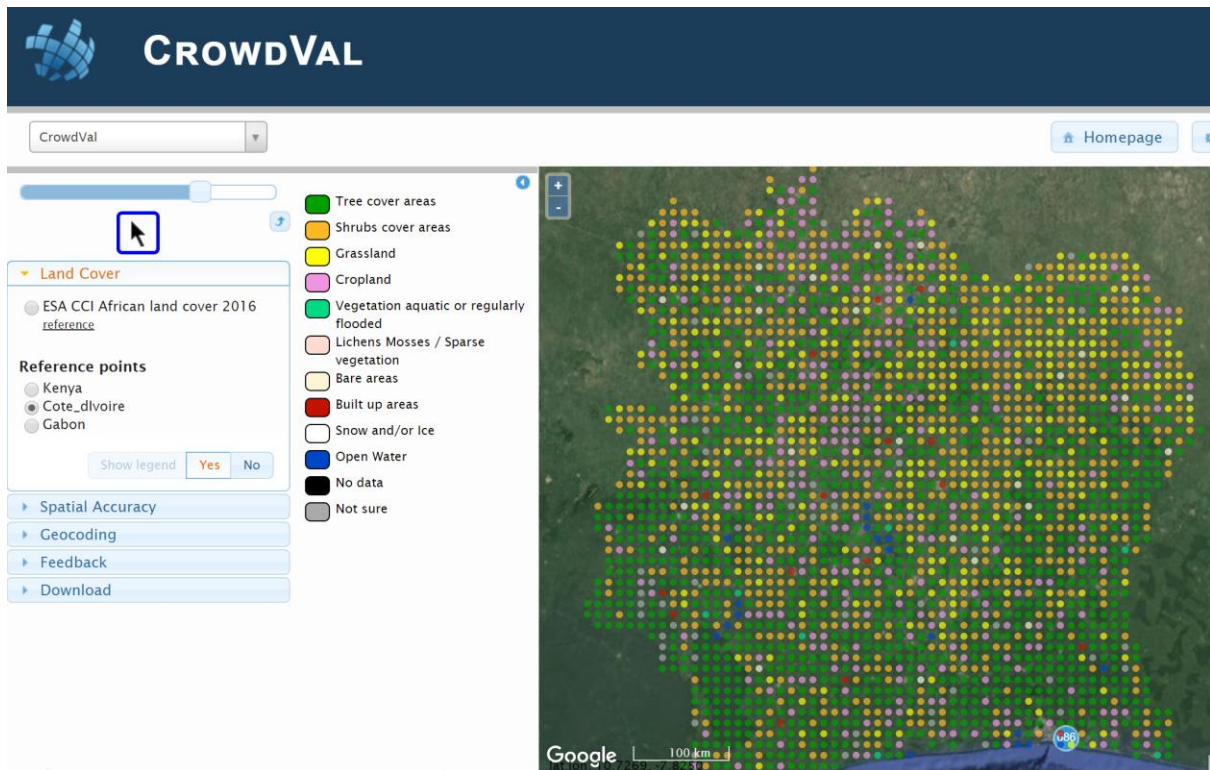


Figure 6: Systematic sample of locations visually interpreted for Ivory Coast

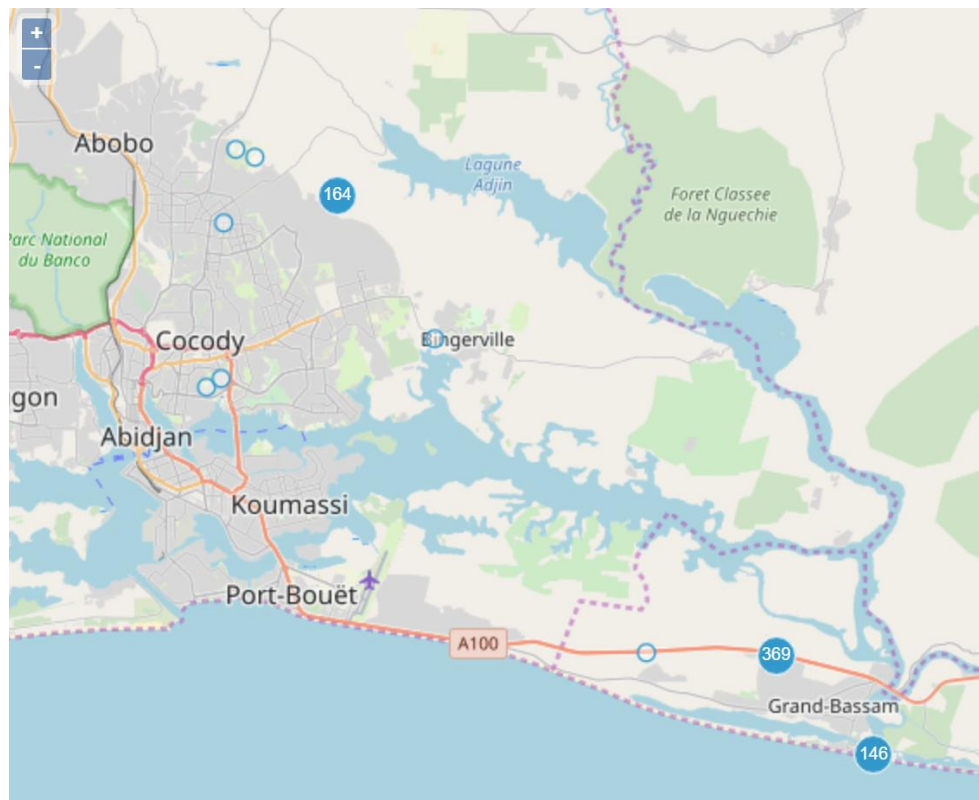


Figure 7: Locations of data collected in situ in Ivory Coast shown in the CrowdVal branch of Geo-Wiki

5.3 Issues related to the campaign

There were no major technical issues encountered with the online version of LACO-Wiki during the workshop. Minor issues were encountered, e.g., the zooming tool stopped working, the sample plot

location was lost when zooming out to large distances, and the pixel location sometimes moved on the map. However, these were generally solved by refreshing the page. Although the date of the satellite imagery from Bing maps is listed in the application, it would be very useful if the acquisition date of the Google Maps layer was also listed. NDVI time-series pixel-level charts would also be helpful in the visual interpretation. This is one of the features currently being added as part of the CrowdVal project. Unfortunately, it was not yet available during this workshop.

There were, however, major difficulties related to the photo-interpretation exercise as outlined below:

- Distinction between trees and shrub: contextual information could help when trees were present within the plot or the surroundings. However, discrimination was complicated when shrubs were photosynthetically active and no trees were present in or around the plot.
- Distinction between cultivated land and grassland: some cultivated areas look very much like grasslands with sparse tree cover, due to local agricultural practices. Only contextual information like geometric features delineating such areas aided with decent temporal spectral signatures can help to discriminate between these two classes. See Figures 8 and 9 for examples of this issue.
- Bare soil: some bare soil areas can be covered temporarily by sparse vegetation while others will remain permanently bare. Mixing of temporal spectral signatures between these two types of land cover may hamper correct bare soil classification, generating confusion with the grassland class notably. See Figure 10 for an example of this issue.
- Wetland: some wetlands are fully covered by tree species. Such wetlands were observed during the fieldwork activity. Such environments may be misinterpreted as tree cover (and misclassified if not appropriately represented in the calibration data set).

Classer l'objet:

■	Je ne sais pas (200)
■	Eau (10)
■	Zone urbaine (8)
■	Sol nu (7)
■	Zone inondée/humide (5)
■	Terre cultivée (4)
■	Zone enherbée (3)
■	Couverture d'arbustes (2)
■	Couverture arborée (1)


Commentaire

Soumettre

Information d'échantillon

Figure 8: Cultivated land that could be confused with grassland

<< Numéro de validation : 85 >>



Classification:

☒ Zone enherbée (3)

Correct Incorrect

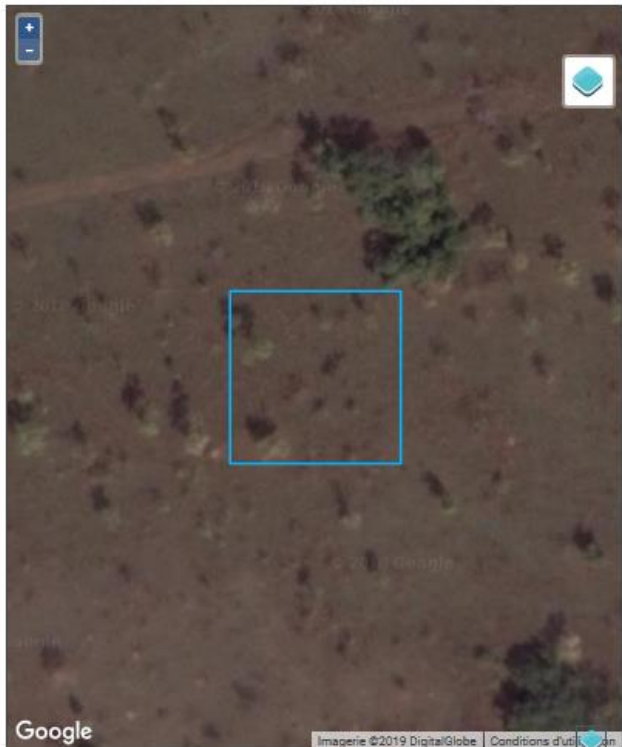
<input type="checkbox"/>	Je ne sais pas (200)
<input type="checkbox"/>	Eau (10)
<input type="checkbox"/>	Zone urbaine (8)
<input type="checkbox"/>	Sol nu (7)
<input type="checkbox"/>	Zone inondée/humide (5)
<input checked="" type="checkbox"/>	Terre cultivée (4)
<input type="checkbox"/>	Zone enherbée (3)
<input type="checkbox"/>	Couverture d'arbustes (2)
<input type="checkbox"/>	Couverture arborée (1)

Commentaire

Soumettre

Figure 9: Cultivated land that could be confused with grassland or bare soil

<< Numéro de validation : 37 >>



Classer l'objet:

<input type="checkbox"/>	Je ne sais pas (200)
<input type="checkbox"/>	Eau (10)
<input type="checkbox"/>	Zone urbaine (8)
<input type="checkbox"/>	Sol nu (7)
<input type="checkbox"/>	Zone inondée/humide (5)
<input type="checkbox"/>	Terre cultivée (4)
<input checked="" type="checkbox"/>	Zone enherbée (3)
<input type="checkbox"/>	Couverture d'arbustes (2)
<input type="checkbox"/>	Couverture arborée (1)

Commentaire

Soumettre

Information d'échantillon

ID de l'élément de l'échantillon 421557

Validé par [nom]

Figure 10: Grassland that could be confused with bare soil

During the photo-interpretation session, there was quite some discussion about the fact that some of the classes did not make sense, i.e., from a user perspective they were not relevant. Both the photo-interpretation and fieldwork in Ivory Coast demonstrated the need for a more detailed legend. Classes could subsequently be merged to obtain the desired 10-class map. The photo-interpretation itself was also tricky, even using these 10 high level classes. The input from the local experts was critical, and it took time to tackle the tricky examples. Ivory Coast is also a difficult situation, with a lack of time series data due to clouds. Finally, 20m is too detailed to produce a land cover map and larger pixels would be better.

In terms of LACO-Wiki Mobile, there were several issues encountered due to the use of old mobile phones and/or Android versions that are not supported by the app. The app worked fine on recent Android mobile phones but login issues occurred at times. There were also problems with the compass, which happened quite often.

Based on the experiences with the app, the following suggestions can be made for future improvements:

- Provide a default zoom level around the user location (e.g., a 300m radius circle display). At present, the zoom level is lost whenever the screen is turned from landscape to portrait screen display (and vice-versa), or when entering and exiting the class specification form. User experience on the ground demonstrated that users would greatly benefit from such a feature.
- Add a slider bar on the sample plot upload page.
- Clarify upload phases with icons and messages to facilitate the distinction between the following three phases: pending uploads, uploading, and upload complete.
- Keep providing distance to the plot location even within the data collection radius (100m in the case of the Abidjan fieldwork).
- Improve compatibility of the app with older phones and Android versions to take fully into account practical circumstances in developing countries.

Moreover, the area in Abidjan was constrained by topography, and large portions were not accessible. This demonstrates problems with the collection of validation data in the field, i.e., the difficulty in reaching locations. There were also a couple of points taken where the users looked into the distance and recorded the land cover there, i.e., 2 to 300 m away. This is similar to the experience in Gabon.

6 Data Collection in South Africa

During the mid-term report meeting held at ESA on 28 Feb 2019, it was agreed that we did not have to undertake an additional field-based campaign in Zambia since the resources were already spent in undertaking campaigns in Kenya, Gabon and Ivory Coast. However, since there was a large amount of spatial inaccuracy identified in the ESA 20 m prototype land cover map in South Africa, as documented by Lesiv et al. (2017), it was agreed that we would create a denser sample in South Africa, i.e., add an additional 920 locations (Figure 11). This would allow us to better understand where the problems are spatially and in which land cover classes. Note that the gaps in South Africa shown in Figure 11 are filled by validation data collected for another project (so will be used in the accuracy assessment) but they are not available for downloading from Geo-Wiki.

In addition, the data collection was different for South Africa compared to the other three countries. In particular, each location was sampled 100 times (a 100 m pixel was sampled every 10 x 10 m corresponding to Sentinel pixels) resulting in 92,000 validation points (see Figure 12 for an example of one location), which are then further reduced to 23,000 20 x 20 m pixels.

The process of data collection was completed by the end of June 2019. See deliverable D4 for the accuracy assessment for South Africa based on this additional sample. These sample points can be viewed and downloaded from the CrowdVal branch of Geo-Wiki (<https://www.geo-wiki.org>).

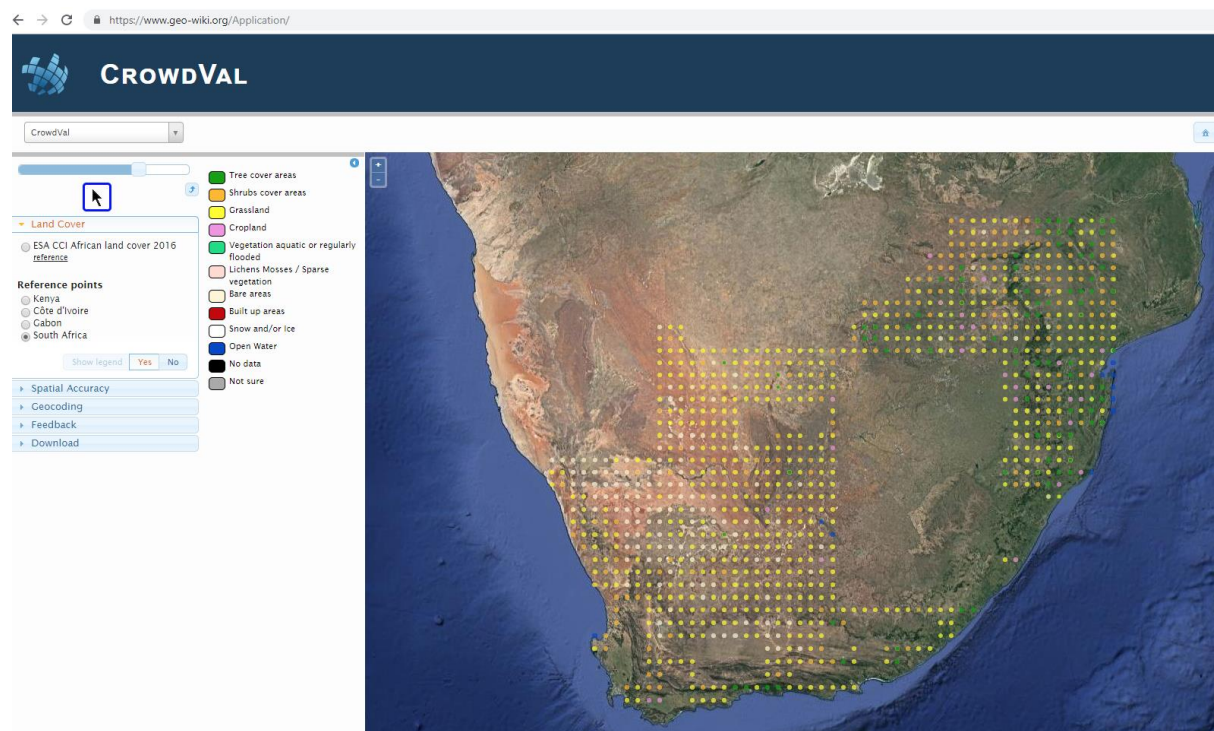


Figure 11: Systematic sample of locations visually interpreted for South Africa

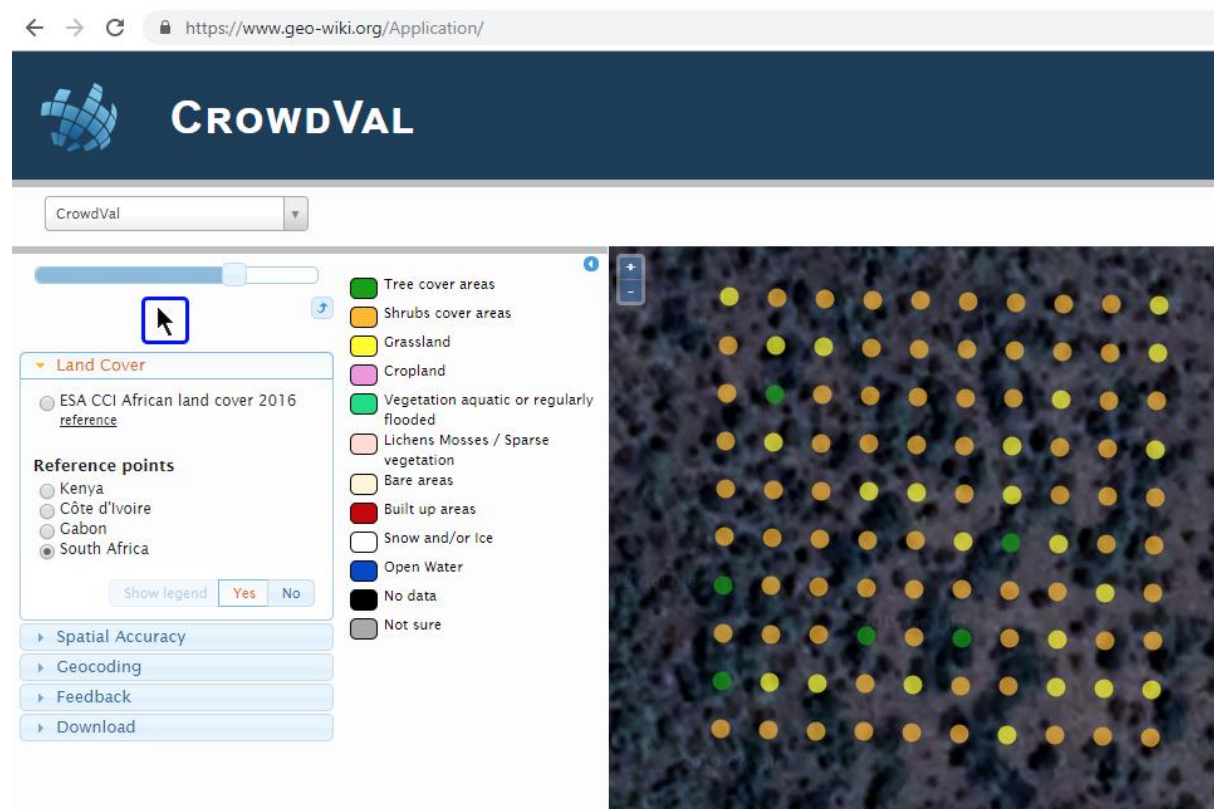


Figure 12: Example of one location in Figure 11 showing the 100 points sampled

7 Lessons Learned and Ideas for the Future

The three campaigns in Kenya, Gabon and Ivory Coast provided some interesting lessons regarding the visual interpretation and data collection in the field. In Kenya, there were very few participants who had local knowledge of land cover. Hence training in visual interpretation by an expert (Myroslava Lesiv) was critical in improving the ability of participants to recognize landscapes. In contrast, in Ivory Coast, local knowledge proved to be useful in classifying certain landscapes. Hence good training is essential but local knowledge can also be useful for some land cover types. This has implications for the classification algorithms applied to satellite imagery, as some features may not be picked up or would simply be misclassified if there is insufficient training or local knowledge.

The mobile app, although a useful tool for data collection, had some limitations. The first was technological as not all participants had smart phones capable of running the app. Difficult landscapes also made it hard to reach some of the points. Hence there is a tradeoff between in situ data collection based on a rigorous statistical sample (and the inability to reach some of these locations) and more opportunistic in situ data collection, where the latter could be used for training algorithms rather than validating a land cover map. Hence the mobile phone could be used more for collection of training data. The algorithm to place sample points along the road was not completed in time for these campaigns but this feature is now available and will facilitate in situ data collection in the future. More generally, the LACO-Wiki tools (both online and mobile app) could be recommended for use in building training data sets in addition to validation data sets.

A 10-class legend may be appropriate for large-scale mapping activities based on coarse spatial resolution images but is less appropriate for large-scale mapping activities based on high spatial resolution images. For example, the participants in Ivory Coast did not find some of the classes useful or detailed enough. Overall, even though random forests or classification tree algorithms¹ do not require calibration data sets with normal distributions for a given mono-temporal numerical independent variable, unimodal distributions should be sought.

The use but also the production of a validation data set for a 10-class high spatial resolution map (20 m) by means of photo-interpretation becomes tricky when such a map 1) results from an algorithm that used coarse spatial resolution samples for calibration, and 2) covers such a large and diverse area like the African continent.

In Ivory Coast, the local participants expressed an interest in repeating the exercise for the purpose of training students in visual interpretation and in situ data collection. This win-win approach could serve as a basis to maintain networks of local experts that could be activated in future land cover mapping and validation exercises.

References

Lesiv, Myroslava, Steffen Fritz, Ian McCallum, N.E. Tsendbazar, Martin Herold, Jean-François Pekel, M. Buchhorn, B. Smets, and et al. 2017. 'Evaluation of ESA CCI Prototype Land Cover Map at 20m. IIASA Working Paper. IIASA, Laxenburg, Austria: WP-17-021'. 2017.

¹ Assuming that one of these classification methods was used to develop the ESA 20 m prototype land cover map of Africa