Center for International Earth
Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY

MaHRV – Mapping Hard to Reach Villages

Portal Guide

Table of Content

1. Introduction

This portal aims to address gaps in spatial coverage of population distribution and to increase the visibility of off-grid, hard-to-reach populations, living in mangrove areas of coastal West Africa, to various development and adaptation actors. Such populations are often underrepresented and underserved in development projects, lack access to basic services including sanitation, health, and education and lag behind on achievement of SDGs. Providing information about the location, size and accessibility of very vulnerable, unaccounted populations to various agencies is a prerequisite to achieving sustainable development.
The portal was developed by The Center for International Earth Science Information Network (CIESIN) with the support from the World Bank Trust Fund for Statistical Capacity Building. The gaps in spatial coverage of population distribution were addressed by building a publicly accessible web application allowing to contribute data and to have access to population estimates in the mangrove areas of West Africa.

The approach adopted here innovatively combines analysis of high resolution satellite imagery, Voluntary Geographic Information (VGI) and statistical modeling. Village population estimates are obtained from in-situ reports via VGI or are estimated through statistical modeling based on in-situ data, village extents and other co-variates. An on-line mapping application allows to access population estimates. The estimates will be iteratively improved as more in-situ data becomes available through the VGI data collection interface.

The approach can be summarized as follows:

Other GIS layers

Data collected
in the field

Population modeling
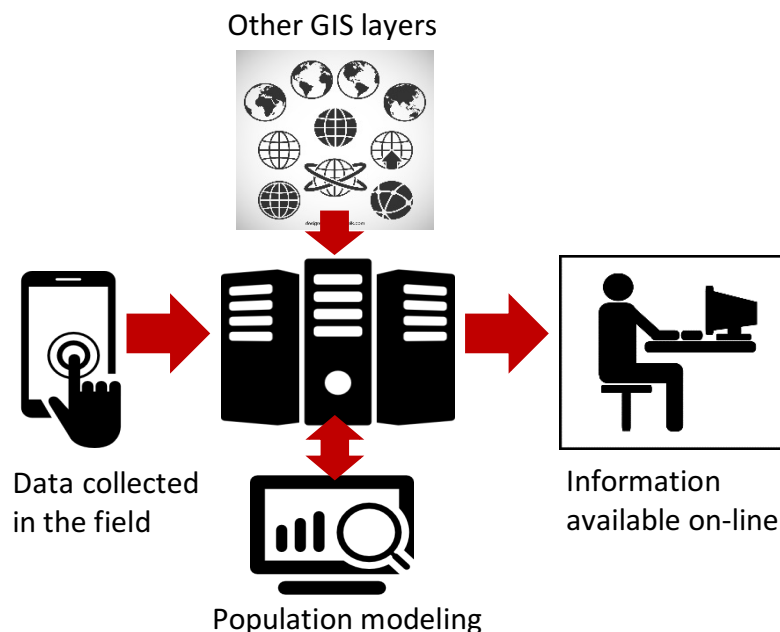
Information
available on-line

Figure 1: Schematic representation of the approach to generating on-line information about the populations living within the mangroves in West Africa.

More information on the approach and methodology can be found in the project report, available on the home page.

In addition to the home page, the portal has two main pages:

1. Information contribution page – for those who can collect and/or want to upload population data for villages in the mangroves of West Africa
2. Information display and download page – for those who want to obtain information/data on village populations living tin the mangroves of West Africa.

The next sections explain more in detail what is available on those pages.

2. Contribute Information

This page is designed to allow the upload of population data collected by various contributors who travel to mangrove villages. Whether you live in such a village or travel there for a project, or any other reason, your knowledge of the village – its population but also infrastructure, or other characteristics – is very valuable. Voluntary Geographical Information (VGI) process, also known as 'crowdsourcing' allows us to harvest and collate such knowledge and contribute to the general understanding of the vulnerabilities of the people living there and hurdles in achieving many of the SDGs.

To contribute information, a data collection device (cellphone or laptop) and a connection to internet (in the field or in the office) are needed. The process and technical characteristics of the data collection interface are described below.

2.1 Information contribution process

2.1.1 Registration

Information contribution starts with a registration – a one-time request to submit data - as a contributor with minimum information such as your name, email address and institution. After being approved contributors can contribute data with a user name that will be provided to them.

2.1.2 Data contribution

There are several options for data contribution, depending whether you are on or off-line.

The on-line Field Data Collection Guidebook, also available for download as PDF explains the process. In all cases, we expect that the information is collected and contributed to the portal with the knowledge and consent of the populations themselves and that the data collection interface (Survey 123) has been used in the field, or manually collected information is input into the online office based data collection interface. The interface is currently available in English and in French.

2.2   The VGI Application

2.2.1   Interface

The VGI interface is supported via the Arc GIS On Line (AGOL). It offers a number of possibilities for mobile and desktop data collection. In all cases the physical data resource is stored in a cloud-based Hosted Feature Layer supporting vector feature querying, visualization, and editing and having the capability to limit the possibility for data collection errors and maintain dataset integrity.

2.2.2   Survey software

The interface to collect data in handheld devices consists of a pre-populated Survey123 app because of its multi-language support and ability to support faceted responses where the response from one question impacts the possible responses of future questions. Figure 3 shows the interface.

Data collected in Survey123 forms can be uploaded directly from the field if internet is available (see 'Collect/Enter data from the field') or synced later from any other location where internet is available (see 'Enter field data from the office' and explanations below).

2.2.3   Off-line capabilities

Because we anticipate that the application will often be used in areas with no access to internet and maps, we offer the possibility to identify the location of the village on a map through the Baseline map in off-line mode. However, the volume of very high resolution data precludes storage and use of the full dataset for West Africa on a simple mobile device. Thus, quarter degree tiles were preprocessed from the ArcGIS Imagery Basemap into tile packages for the coastal study area that can be downloaded individually as needed. These TPKs (~100GB) are hosted in the CIESIN repository and are accessible via a webmap on AGOL. Figure 4 shows and example of the tiles developed for the application.
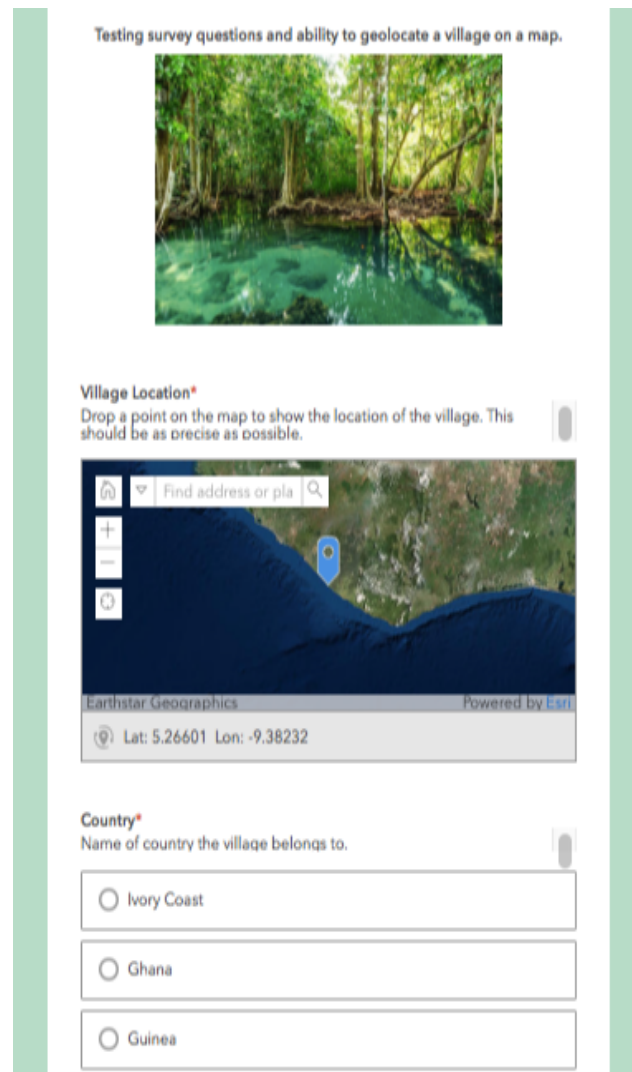

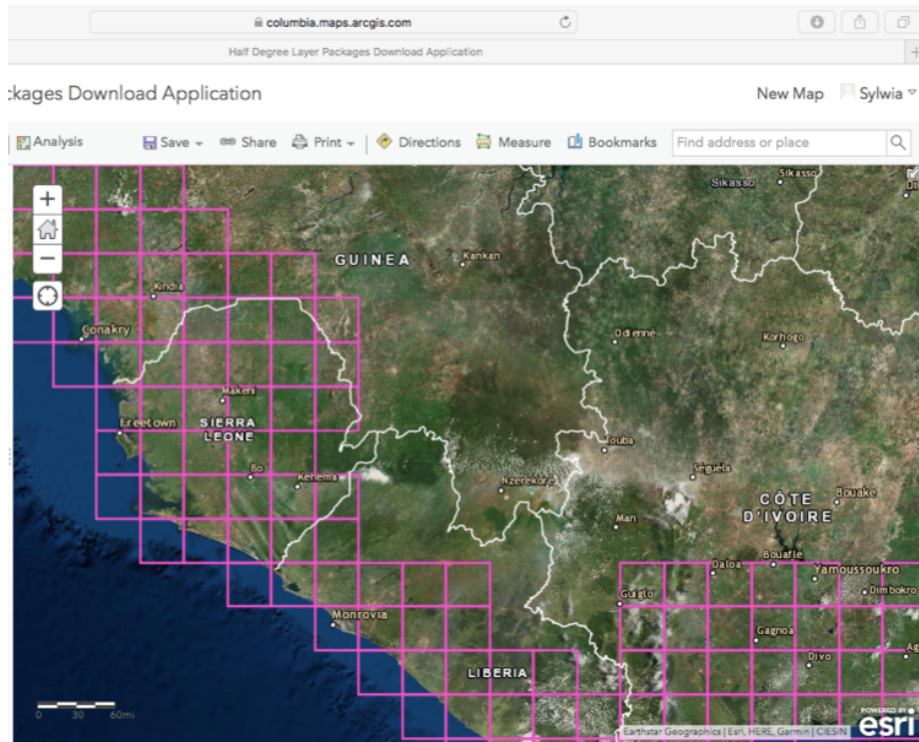*Figure 2: Screenshot of the VGI data collection interface.*

5

Figure 3. Map of the tiles developed for off-line use of the application

2.2.4   Enter field data from the office

Data collected with the Off-line capabilities of Survey123 can be synced at any time when an internet connection becomes available. Additionally, a browser-based Office VGI Data Entry application is available so that users might enter manually transcribed field data if needed.

3. Get Information

This page is designed to access information about villages located in the mangroves of West Africa. The information can be accessed in two formats:

- An interactive *map applicatio*n for viewing and querying data allowing to see all the villages within mangrove areas
- *Data download* as a comma-separated format file and Shapefile

3.1   The interactive mapping application

The mapping application allows to visualize the location and access the population estimates of the villages within the mangroves. It displays four layers: village location and extent, mangrove extents and district level boundaries. Only villages within or up to 100m from the mangroves are displayed. Village information can be accessed by clicking on the village. By clicking outside of the village information at the district level is displayed. The controls on the upper left side allow to zoom in and out and change the legend display.
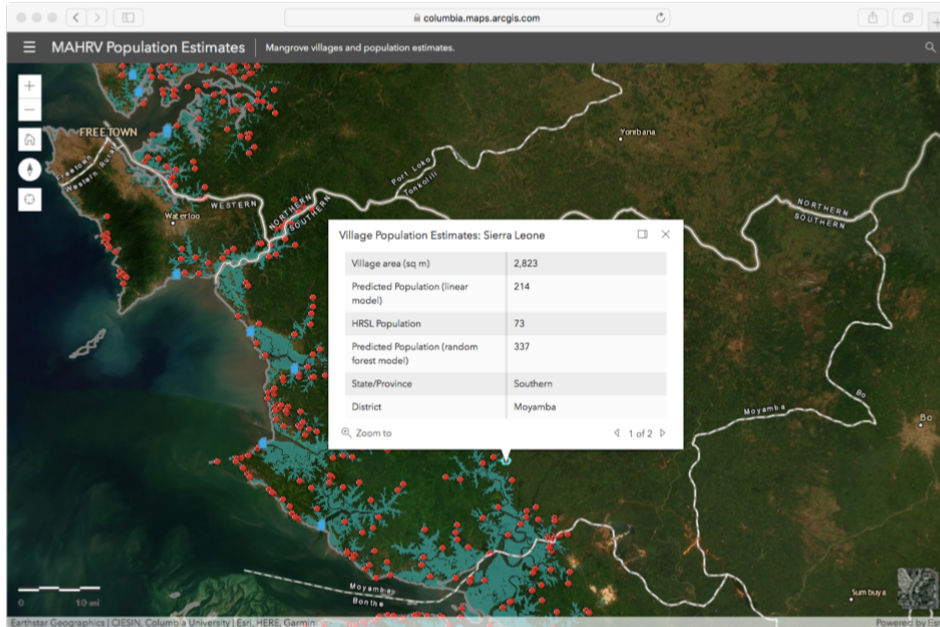
Figure 4: Screen capture of the Mapping interface displaying a pop-up information for a village.

3.2   Data download

The data can be downloaded as a zip archive which contains a file Geodatabase, shapefile, and Comma-Separated Values format for:

- Visited villages – data contributed through the VGI (point features)
- Village extents derived from the High Resolution Settlement Layer (polygon and point feature class), with population estimates from the HRSL, a linear population model, and a random forest population model.
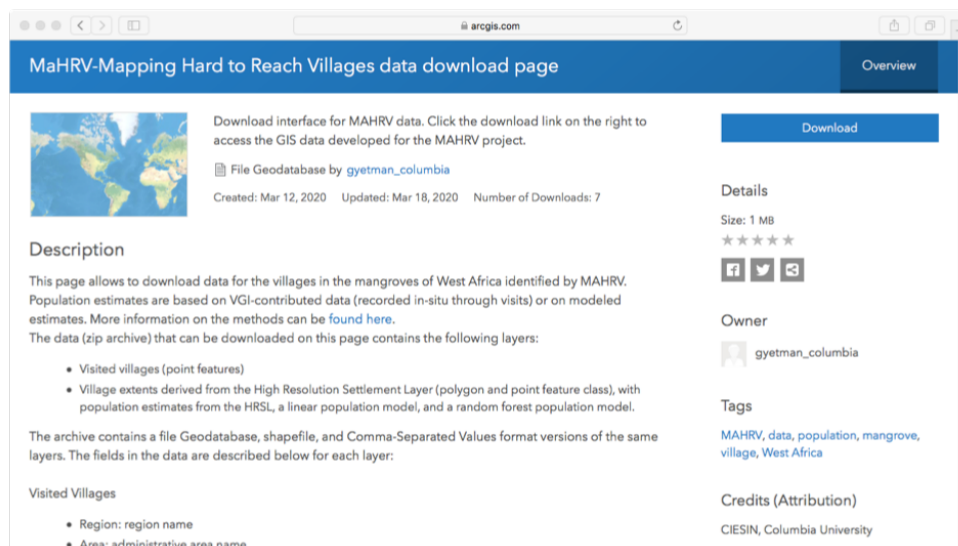

Figure 5: Screen capture of the download page

4. Technical considerations

4.1 Population models

The approach to population estimates capitalizes on advances in the application of machine learning to remote sensing problems by using the High Resolution Settlement Layer (HRSL; https://ciesin.columbia.edu/data/hrsl/) jointly developed by CIESIN and the Connectivity Lab at Facebook. Population estimates for the villages with no in-situ data were obtained through three modeling approaches:

- Linear model: based on a relationship between population recorded in villages visited and their extends from HRSL; this relationship is then used to estimate village population based on village extent.
- Direct HRSL estimate: average population density is derived from the total population and total villages extents within a given census unit; individual village population is obtained by using the average population density and the village extent. This model generally underestimates village population, as shown on figure 6[1].
- Machine learning (random forest) model: uses established relationships between population distribution and multiple biophysical characteristics; a random forest model was created using the available in-situ population data and standard biophysical covariates.
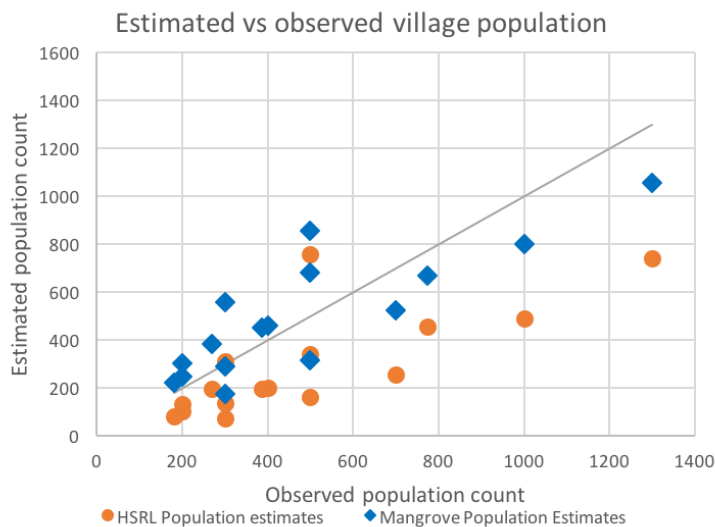

Figure 6: Population estimates using HRSL approach (based on settlement size and census data) and using village data and size to inform the model vs observed population size.

---

[1] Only limited number of data from visited villages was available for the very first models. As data become more available models and estimates will be updated

More about the models and their various characteristics can be found in the project report available from https://mahrv.ciesin.columbia.edu/home and from https://github.com/gyetman/mahrv

4.2   Village Clustering

The villages derived from HRSL sometimes had disparate polygons due to the 30m grid cell size and some missed buildings or disjoint settlement patterns. An example is shown in Figure 7. The villages were clustered based on spatial proximity using defined distance (DBSCAN), with a search radius of 250m. This reduced the count of HRSL-derived settlement polygons from 11,771 to 4,022 villages.

Figure 7. HRSL village polygons (shown in red) over satellite imagery. The disjointed nature of the village extents can be seen.

4.3   Technological Framework

Several technological solutions were considered, including open source and commercial solutions. Ultimately ArcGIS Online (AGOL) was selected as the primary technology utilized for the development of Online Services, Web applications, and Mobile Applications.

4.3.1.1   *Framework Selection Justification*

This selection is supportive of technology transfer. Because the technologies are implemented using cloud technologies, the feasibility of transferring to another organization (with unknown institutional knowledge) is increased as the ArcGIS client services can be customized by knowledgeable IT/GIS professionals, and hosted on local hardware if desired, or simply implemented in the cloud using well known templates. The data services can be downloaded from the cloud for local use, or managed entirely through an AGOL organization account. For

this project Web map templates, available from AGOL, were utilized to create an online user interface for data access and visualization, and for the input of VGI in the office.

4.3.1.2 Security of services

AGOL provides tiers of sharing options for all pieces of content. We have created a project-specific group that requires account login for access and only shared content to members of that group (presently, only members of the research team). For the initial phase, users will be vetted and added to the group. Subsequently we may decide to share the services more broadly so that any person with a (free) AGOL account might contribute to the data collection or view the data collection results.

Annex 1 – The verbal consent script and the questionnaire included in the VGI interface

A1.1 The verbal consent/Information sheet

[Interviewer, please read the following to the respondent.]

Good morning/ afternoon. My name is [interviewer's name] and I am here to ask you some questions about your village. As a knowledgeable person in the community, I am reaching out to see if you are willing to answer some general questions about this village.

Your participation is voluntary

Answering questions is voluntary, you can skip any question at any time.

Purpose of this research

We are interested in learning more about how many people live in this village, whether this human settlement is seasonal or permanent, and the various services available (for example, is there a health clinic in this village?, or is there a primary school?). With this research we will understand better the configuration and characteristics of human settlements within mangrove forests.

How long does it take?

The survey takes around 15 minutes to complete.

Risks and possible discomforts

If you choose to participate, we anticipate minimal risks and only the minor discomfort that might accompany answering a questionnaire.

Potential benefits

We cannot promise any benefits to you or others from taking part in this research. However, possible benefits include contributing to the general knowledge about communities living near or within mangrove forests.

What happens to the information collected?

I will store all your answers in this mobile device [show smartphone/ tablet]. After we finish, I will save all the information here, and once I get access to the internet, I will send the data to a place where answers from other respondents, from other villages, are grouped together. All the information that is collected is stored and used securely.

Privacy protection

I am not collecting any identifiable information from you. I will neither take your name or contact information. My interest of interviewing you is only to collect information about the village only.

Contact information

The research for this study is performed by the Center for International Earth Science Information Network (CIESIN) Research Team, at Columbia University. Their mailing address is: 61 Route 9W Palisades NY 10964, USA. Their telephone number is: +1-845-365-8988. Contact them if you think the research has harmed you; if you have questions, concerns, or complaints about this research; or if you wish to withdraw your answers from this study.

This research has been reviewed by the Columbia University IRB. Their mailing address is: Columbia University Medical Center, 154 Haven Avenue, 1st Floor, New York, NY 10032. Their telephone number is: +1-212-305-5883. Contact them if you are unable to contact the research team; if the research team does not respond to your contact; if you want to talk to someone besides the research team; or if you have questions about your rights as a research participant.

Do you agree to participate in this study?

Yes   → [continue with the questionnaire]

No   → Are there any questions I could answer?

      → [if still no] Thank you for your time!

A1.2 The questionnaire

1. Name of individual entering data (automated, based on controlled access).
2. Date -start (automated)
3. Date -submitted (automated)
4. IMEI (automated, if using mobile device)
5. UUID (automated, for each submission).
6. Location (latitude, longitude, precision if collected in the field using mobile device)
7. Administrative hierarchy (nested, pre-populated domains, as applicable)
    a.   Country
    b.   Admin level 1 (e.g. state)
    c.   Admin level 2 (e.g. county)
    d.   Admin level 3 (e.g. township)
    e.   Admin level 4 (e.g. village)
    f.   (varies per country)
8. Traditional hierarchy (nested, pre-populated domains, as applicable)
    g. Country
    h. Level 1 (e.g. zone)
    i. Level 2 (e.g. area)
    j.   (varies per country)
       9. Name of village or human settlement
       10. Alternative name of village or human settlement
       11. Estimated total population within the village
       12. Year of estimated total population within the village

13. Estimated number of houses (dwellings) within the village
14. Year of estimated number of houses (dwellings) within the village
15. Is this a seasonal or permanent settlement?
16. Is there a {…} in the village? Please select all that apply
a. Health facility
b. Health post
c. Nursery/ Pre-K school
d. Primary school
e. Secondary school
f. Vocational/ technical school
g. Market
h. Church
i. Mosque
j. Other place of worship

17. If not how far is the {…} using the best transportation (estimate in distance or time)
   a. Health facility
   b. Health post
   c. Nursery/ Pre-K school
   d. Primary school
   e. Secondary school
   f. Vocational/ technical school
   g. Market
   h. Church
   i. Mosque
   j. Other place of worship

18. If not what is the best transportation to get to
a. Health facility
b. Health post
c. Nursery/ Pre-K school
d. Primary school
e. Secondary school
f. Vocational/ technical school
g. Market
h. Church
i. Mosque
j. Other place of worship

19. Source of information, please select:
   a.   Personal knowledge
   b.   Village authorities

c. Village resident
   d. Local official
   e. Regional official
   f. Census worker
   g. NGO worker
   h. Other

20. Notes