Use of GIS and Remote Sensing to Build a Sampling Frame for Household Surveys in Somalia

Authors: Josyline Gikunda, Amina Omar, Umikaltuma Ibrahim, Mariam Alwi, Richard Ng'etich, Felix Mulama, Samwel Andati

1. Introduction

Somalis have endured a long spell of absence of comprehensive information on themselves: on population and important social and economic characteristics. Before the Population Estimation Survey (PESS) in 2014, the last available information was from a census conducted in 1975, which published limited results; and the findings from another population census conducted from 1985 to 1986 that were not published officially. To fill this crippling gap, and support Somali authorities and their partners to design policies and plans based on the realities of Somalis on the ground, and in consultation with the UN Country Team, the United Nations Population Fund took on the lead role in coordinating the Population Estimation Survey. PESS was the first extensive household sample survey to be held in decades that provides reliable population estimates and information by geographical areas, among other details. For Somali regions recovering from many years of war, it was extremely challenging to plan and carry out a survey of the nature and scale of PESS. This was mainly due to a capacity gap coupled with insecurity and lack of access to some areas. (UNFPA/PESS, 2013).

The design of sampling frames during the Population and Estimation Survey (PESS) done by UNFPA, was a major milestone for Somalia. Prior to the survey, the country was without a sampling frame and surveys were conducted on the basis of lists of settlements. However, there was need for a more up-to-date and extensive sampling frame that covers Urban, Rural, Internally Displaced Persons (IDPs) and Nomadic settlements in the country to be used for future household surveys. As such, an updated frame was developed for the 2018/2019 Somali Health and Demographic Survey (SHDS)

A sampling frame is a complete list of all sampling units that entirely covers the target population from which a sample is drawn. It provides a means for choosing the particular members of the target population that are to be interviewed in the survey; which an easy scientific method of investigating the entire population to assess an outcome of interest. A sampling frame must also provide a clear identification of every sampling unit and its elements should be ordered in a way that a random selection of a sample can be done in an efficient way(Iglesias, 2013).

A perfect sampling frame should be:

1. Complete, meaning that it covers all the elements of the population.
2. Accurate, meaning that every element of the population is included in the frame once and only once
3. Current/up-to date

Although these properties are ideal and are unattainable in household surveys, they should be strived for in the construction of a frame and when using a frame that already exists. One of the key benefits of a sampling frame is the possibility of making multiple selections of samples for a number of varied surveys. Researchers face difficulties in getting an up to date line list of sampling units, especially in an environment where rapid changes occur frequently. Particularly in household surveys, it is difficult as newly constructed dwelling structures are seen as changes occurring both in urban and rural areas; hence a need for an efficient and effective way to monitor the changes (Elangovan et al., 2016).

Geographic information plays an integral role in conducting household sample surveys. Sampling and observation units are connected in some way with geographic space, and often results are summarized by geographic regions. In the recent past, digital geospatial data has been introduced into several components of sample survey processes (Nusser, 2007). With recent advancements in Geographic Information Systems (GIS) and Remote Sensing (RS), these technologies can be exploited extensively for spatial sampling frame design for household surveys. They are particularly useful when an area is inaccessible due to security concerns which is the case in Somalia. This paper focuses on use of the growing availability of GIS tools and Remote Sensing including high resolution satellite imagery, Google Earth, handheld Global Positioning System (GPS) devices, and location-enabled applications on mobile phones such as GPS Essentials; to develop a sampling frame for Somalia household surveys.

2. Objective

The main objective of the study is to utilize Geographic Information Systems (GIS) and Remote Sensing (RS) technologies to build an up-to-date sampling frame for Somali Health and Demographic Survey (SHDS) and subsequent household surveys in Somalia.

3. Study Area

Somalia; officially the Federal Republic of Somalia is a country located in the Horn of Africa. It is bordered by Ethiopia to the west, Djibouti to the northwest, the Gulf of Aden to the north, and Kenya to the southwest. Around 85% of its residents are ethnic Somalis who have historically inhabited the northern part of the country.

Figure 1: Basemap of Somalia

4. Methodology

The sampling frame that was used in PESS comprised of defined clusters of households (primary sampling units-PSUs) within the urban areas, rural areas, and camps for the Internally Displaced Persons (IDPs) and water-points for nomadic areas within the 18 pre-war regions. A total of 7,505 Enumeration Areas (EAs) were digitized in urban areas. These EAs were delineated within the lowest administrative sub-division available. The rural frame was derived from a settlement list from the census conducted by UNDP in 2005/2006 which was updated to obtain a master list of settlements. Although, the PESS sampling frame PESS has proved to be a useful resource for international agencies conducting surveys in Somalia, it had numerous limitations hence the need to build a new sampling frame to be used in the subsequent household surveys.

The sampling frame geo-database generated from PESS had the following limitations that needed to be addressed:

- The existing sampling frame geo-database consists of urban EAs represented as polygons and the rural frame; settlements, represented as points. As a result, there is no continuity in the EAs between the two frames or geographic layers.
- The digitizing of small urban towns and large rural settlements was not done at an Enumeration Area (EA) level.
- Some major urban towns in South/Central Somalia were not digitized at an EA level due to insecurity.
- A large number of EAs in the rural frame are missing coordinate/locational data due to inaccessibility of the settlements during the time of the survey. Some of these areas are accessible now and high spatial resolution imagery available.

This paper focus on the methodology used to design the Urban/IDP stratum comprised of 1988 pre-war district capitals and the Rural/IDP stratum comprised of a list of Somali settlements that was updated in 2017, obtained from United Nations Support Office for Somalia (UNSOS) of the SHDS sampling frame.

The main source of data used was high resolution satellite imageries below:

Table 1: List of satellite imageries used and their spatial resolution

Imagery	Resolution
WorldView-1	0.5 M
WorldView-2	0.5 M
WorldView-3	0.31 M
Geo-Eye-1	0.5 M
Geo-Eye-2	0.31 M

Through the interpretation of up-to-date high-resolution satellite imagery and field knowledge, all structures on the ground were digitized in urban and accessible rural areas. This included a first-level verification of the PESS rural settlement list; where settlements in each region were authenticated by Somali staff from the statistical offices. Thereafter, settlements that appeared to have more than 50 dwelling structures were digitized.

A 10 percent ground validation of the digitized structures was done for the large cities and rural settlements to make the necessary adjustments. The digitized structures were categorized into three types:

Type 1 - Dwelling Structures: - these are structures that were interpreted as residential structures

Type 2 - Non-Dwelling Structures: - these are non-residential structures (non-dwelling) e.g. commercial buildings, government buildings, mosques, hospitals, schools etc.

Type 3 - IDP shelters: - these are residential structures that house internally displaced persons; particularly those in camps.

All structures categorized as Type 1 and 3 constituted a dwelling structures database. Enumeration areas were created based on the count of the dwelling structures whilst ensuring their boundaries followed recognizable features such as roads and rivers. Each Enumeration area had a minimum of 50 and a maximum of 149 dwelling structures. A spatial join was carried out between enumeration areas and the dwelling structures. The spatial join creates an additional field in the attribute table showing the number of dwelling structures contained in each enumeration area.

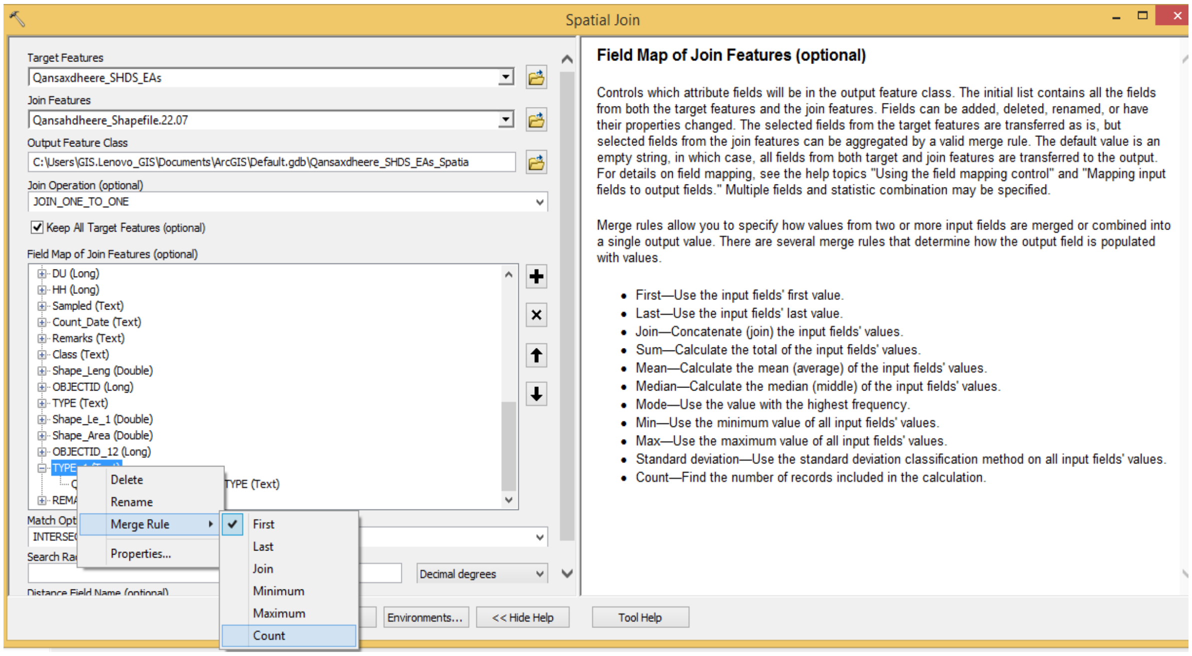

Figure 2: The spatial join tool in ArcGIS software used to assign dwelling structures to the EAs based on location

Figure 3 below shows the output feature class attribute table showing the number of dwelling structures as the join count field. This was done for all the digitized towns and settlements Somalia

Table

Qansaxdheere_EAs

New_EA_C	Join_Count	Region_N	Region_C	District_N	District_C	Town_N
0013	114	Bay	24	Qansax Dheere	2404	Qansax Dheere
0001	147	Bay	24	Qansax Dheere	2404	Qansax Dheere
0003	111	Bay	24	Qansax Dheere	2404	Qansax Dheere
0005	101	Bay	24	Qansax Dheere	2404	Qansax Dheere
0014	104	Bay	24	Qansax Dheere	2404	Qansax Dheere
0012	104	Bay	24	Qansax Dheere	2404	Qansax Dheere
0010	149	Bay	24	Qansax Dheere	2404	Qansax Dheere
0002	116	Bay	24	Qansax Dheere	2404	Qansax Dheere
0009	115	Bay	24	Qansax Dheere	2404	Qansax Dheere
0011	106	Bay	24	Qansax Dheere	2404	Qansax Dheere
0007	124	Bay	24	Qansax Dheere	2404	Qansax Dheere
0006	120	Bay	24	Qansax Dheere	2404	Qansax Dheere
0004	98	Bay	24	Qansax Dheere	2404	Qansax Dheere
0008	128	Bay	24	Qansax Dheere	2404	Qansax Dheere

Figure 3: Output feature class attribute table after the spatial join

Once the spatial analysis was complete, the attribute tables were exported to MS Excel to allow for sampling of the EAs for the survey. All EAs were selected systematically with known and non-zero probabilities to obtain unbiased estimation and to be able to measure the precision of the survey indicators by estimating the sampling errors. Enumeration Area maps were then created for the sampled areas and Google Earth (KMZ) files of their boundaries generated. These digital files were uploaded to the GPS Essentials mobile application during field work for navigation ensuring that enumerators collected data from within the boundary of correctly sampled EAs.

5. Results

A total of 11,255 Enumeration Areas (EAs) were digitized; 7,968 in urban areas and 3,287 in rural areas as shown in the table below. The whole of Banadir region is urban; hence no rural EAs. Rural areas in Middle Juba region was not digitized, since the rural list of settlements in area could not be verified by the field team due to the high level of insecurity in the area.

Table 2: Number of EAs digitized by region

Region	**No. of** Urban **Enumeration Areas**	**No. of** Rural **Enumeration Areas**
Awdal	153	172
Woqooyi Galbeed	1079	443
Togdheer	344	216
Sool	141	118
Sanaag	152	106
Bari	342	254
Nugaal	193	66
Mudug	249	183
Galgaduud	208	215
Hiraan	212	172
Middle Shabelle	140	270
Banadir	3015	-
Lower Shabelle	373	215

Bay	588	172
Bakool	74	270
Gedo	330	215
Middle Juba	102	-
Lower Juba	273	76
Total	**7,968**	**3287**

The figures below show a sample of enumeration areas that were digitized for two major urban areas; Mogadishu and Baidoa towns, and one rural area; Guriceel town.

Figure 4: Mogadishu town EA boundaries; 3,021 EAs were digitized.

Figure 5: Baidoa town EA boundaries; 448 EAs were digitized.

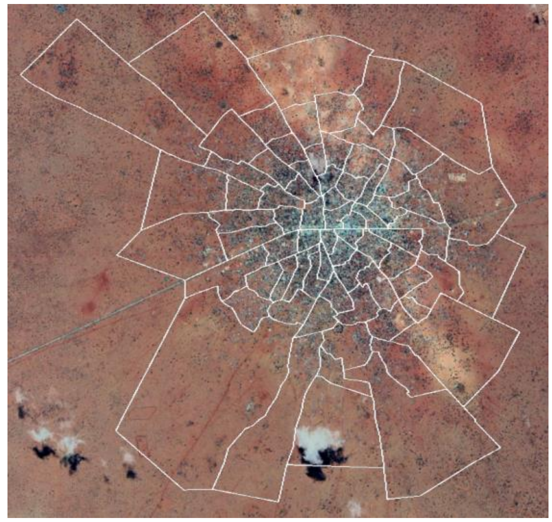

Figure 6: Guriceel rural settlement EA boundaries; 73 EAs were digitized.

6. Conclusions and Recommendations

Somalia is a country where there is need for spatial demographic data like an up-to-date sampling frame which is crucial for continued development planning and humanitarian responses. This study has proven that it is possible to use GIS and remote sensing to build a sampling frame of Somalia particularly the inaccessible areas.

Due to the dynamic nature of the Somali context; future research will aim to using a more automated approach coupled with field knowledge to continuously maintain an up-to-date sampling frame for future surveys

References

Elangovan A. et. al. (2016). Geospatial Techniques in Health Survey to Overcome the Lack of Sampling Frame. *International Journal of Advanced Remote Sensing and GIS*, *5*(1), 1908–1914. https://doi.org/10.23953/cloud.ijarsg.78

Iglesias, L. M. (2013). *IMPROVING THE USE OF GPS , GIS AND RS FOR SETTING UP A MASTER SAMPLING FRAME Luis Iglesias Martínez July 2013 Summary*. (July).

Nusser, S. (2007). Using geospatial information resources in sample surveys. *Journal of Official Statistics*, *23*(3), 285–289.

UNFPA/PESS. (2013). *Population Estimation Survey (PESS)*. (October), 1–2.