

Population Estimation Survey (PESS)

Estimation, methodology, application and validation

Table of Contents

1.	Introduction	1
2.	Djibouti Agreement of February 2014	1
3.	Theoretical background	1
4.	Adjustment for nonresponse	3
5.	Corrections on data file.....	5
6.	Nomadic data	5
7.	Inaccessible areas	5
8.	Illustrations of computations of urban and rural population estimates	7
9.	Population estimation of nomads at water-points.....	9
10.	Illustrations of computations of nomadic population using the adjusted weights	10
11.	Use of high resolution imagery as a validation tool	11
12.	Conclusion.....	18
13.	Data Sets (Attachments in excel spreadsheet)	18

1. Introduction

This document is as a result of a request made by the Somali partners on the need for an elaboration of the Population Estimation Survey (PESS) methodology and its application. The request was made on 23 July 2014 during a conference call meeting. The long process of the PESS greatly benefitted from a series of consultations between the UNFPA Technical Support Unit (TSU) and Somali experts. International standards were observed during all the stages of the implementation of the survey.

2. Djibouti Agreement of February 2014

The agreement stipulated the selection, weighting and the implementation of the data collection at water-points. Refer to the attached Djibouti Agreement under item 3), 4) and 10).

3. Theoretical background

Sampling weights are required to compensate for unequal selection probabilities of EAs or settlements. They may also assist in adjusting for non-response, and for known differences between the sample and the reference sub-population units and populations. The weights were used in the estimation population size and characteristics of interest from the samples.

The base weight of a sampled unit, i.e. EA or settlement is the number of units that represent the population units (EAs or settlements) not included in the sample. In estimating the PESS population value we assigned a weight to each sample unit, such as EAs, Settlements, Water points, and IDPs or each of the responding units. In addition to design weights; response weights were applied where necessary.

PESS is a one-stage stratified cluster sample design. The design weight for a unit in the sample is the inverse of its probability of inclusion. Note that in any multi-stage selection, account is taken of selection probabilities of segment(s) (Primary Sampling Units (PSUs)), non-response or non-coverage. The probability of selection is the combined probability of selection at each stage; the weight takes into account adjustment if done. The non-response adjustment factor is defined as the ratio of the sum of the weights in the original sample to the sum of weights of the responding units.

Weighting: Urban/rural sedentary

The urban and rural sedentary population was weighted using the inverse of the selection probabilities determined using the following result

$$p_1 = \frac{a \cdot m_i}{m \sum_i^M m_i}$$

Where p_1 is the probability of selecting a given EA or settlement, a is the number of EAs or settlements selected in a region. m_i is the number of households in the i th EA or settlement while

$$\sum_{i=1}^M m_i$$

is the total number of households in region. In the spreadsheets, p_1 has been computed. At the second stage, sampling probability for selection has been computed. This applies both to the rural settlements and EAs in the urban areas that were segmented due to large sizes exceeding one measure of size. The second stage of the selection of the sample is given by

$$p_2 = \frac{n_s}{n_r}$$

Final probability of selection $p_i = p_1 \times p_2$ resulting in the weight for the EA/Settlement being

$$w_i = \frac{1}{p_i}$$

The PESS data appears in a file, with a record for each household in the sample EA or settlement. As stated earlier, with probability sampling, each EA/settlement in the frame has **a known and non-zero probability**. Using a hypothetical example of one in 30 which implies that each selected EA/settlement represents on average 30 EAs/settlements of the survey population, the weight in this case is 30. Also assume that a selected EA is larger than the measure of size adopted for the sample survey and it is segmented into 7 segments. The probability of selecting a given segment is $1/7$. Consequently, the application of result gives the overall selection probability is given as $\frac{1}{30} \times \frac{1}{7}$

The corresponding weight is the reciprocal of the probability of selection of the PSU which is

$$w_{hi} = \frac{7}{p_{hi}}$$

4. Adjustment for nonresponse

Consider that in a given settlement, 200 households were selected and of these, only 150 responded. The adjustment factor in this case will be $\frac{200}{150} = 1.33 \approx 1.3$.

The overall selection probability being
$$p_{hi} \times \frac{1}{7} = \frac{p_{hi}}{7}$$

The corresponding weight is the reciprocal of probability of selection of the PSU,

which is $w_{hi} = \frac{7}{p_{hi}}$.

Adjustments for non-response made on the data

Some EAs or settlements were not covered even though selected for the survey. Equally, some households listed in the EAs were not fully covered in a number of cases. There was need, therefore, to make adjustments for these two problems in conformity with the explanation provided above. In this case, considering the number of selected EAs to be n_s and the number responding to be n_r , the adjustment for the non-response involving EAs/settlements was $\frac{n_s}{n_r}$. Following the same approach, the adjustment for the households missed in the enumeration was given as $\frac{n_{sh}}{n_{rh}}$. These two adjustments were carried out on the sedentary urban and rural data. The number of the PSUs selected for the urban and rural is provided in Table 1, while the number of EAs that responded are provided in Table 2. The two tables provide a snapshot view of the extent of the non-response which consequently required adjustment.

Table 1: Allocated Primary Sampling Units for the PESS

Region	Urban	Rural	Total
Awdal	42	22	64
Woqooyi Galbeed	173	37	210
Togdheer	94	27	121
Sanaag	38	37	75
Sool	25	15	40
Bari	23	123	146
Nugaal	27	25	52
Mudug	77	112	189
Galgadud	38	48	86
Hiran	16	59	75
Middle Shabelle	8	53	61
Lower Shabelle	30	136	166
Banadir	193	0	193
Bay	17	219	236
Bakool	5	67	72
Gedo	19	72	91
Middle Jubb—a	8	31	39
Lower Jubba	35	21	56
Total	868	1104	1972

Table 2: The Number of Responding EAs/Settlements

Region	Rural	Urban	Total
Awdal	22	38	60
Wogooyi Galbeed	38	170	208
Togdheer	30	92	122
Sanaag	41	39	80
Sool	19	25	44
Bari	38	87	125
Nugaal	11	24	35
Mudug	30	71	101
Galgaduud	31	31	62
Hiraan	45	6	51
Middle Shabelle	18	8	26
Banadir	0	183	183
Lower Shabelle	99	22	121
Bay	102	16	118
Bakool	33	5	38
Gedo	43	12	55
Middle Juba	11	5	16
Lower Juba	16	33	49
Total	627	867	1494

5. Corrections on data file

As mentioned during the presentation, there were various duplicates in the data file involving records of individuals in the households. Within one household, there would be an individual having records duplicated. Such anomalies were detected by scrutiny of the records in the households. The members would be having same values for the variables of the data. These were cleaned out by deleting one of the duplicates whenever identified.

6. Nomadic data

The estimation was based on the weights based on the number of water points allocated to a given region and the mean number of watering episodes. Further adjustment was made in consideration of the duration of the watering episodes. One observation is that the mean watering time was not in conformity to the documented time for the animal types. For example it was noted that the camels were watering at closer intervals of time contrary to the 12 day period assumed in the study. This means that within the 12 day period there would be more camels appearing at the water points than expected. Thus if the watering interval was earlier understood to be 12 and yet in the interviewing period it is found to be 4 days, then there will be more camels at the water points during the interviewing period than expected in the theoretical formulation. Where such an observation was observed, an adjustment based on the ratio of the number of days n to 12 was used.

7. Inaccessible areas

The inaccessible areas were estimated independently using the satellite imageries. The satellite images were used to count the number of structures in the boundaries of each of the inaccessible areas. The information on mean sizes of structures, dwelling units and households were key in the application of the technique. From the listing information and the mapping exercise, the expected number of dwelling units per structure was established. In addition the average number of dwelling units in the structures was determined and using the mean household size of 6 the population for areas that could not be accessed was estimated. The use of this data was based on the fact that the averages for the variables under consideration do not vary widely particularly in the rural areas.

An illustration of the computations carried out in the estimations

In this report, our example is based on the urban component of Wooqoyi Galbeed. Spreadsheets for the computations are provided showing the stages followed in the computations. This applies to all the regions in the survey.

In the attached spreadsheet, the computations were carried out in conformity to the formulations in this document. The probability for selection of the first segment was segment was carried out as follows:

$$a= 173 \quad bi=60 \quad \sum bi=128,979$$

$$p1=0.080478 \quad wi=12.42572$$

In this report, we shall provide the estimates for one of the urban regions in Somaliland, viz. Wooqoy Galbeed. An example for the rural areas will be drawn from Puntland while the Nomadic population will be represented by an example from South Central.

Table 1: Population distribution by region

RC	Region	Urban	Rural	Nomads	IDP	Total
11	Awdal	287,821	143,743	233,709	7,990	673,263
12	Wooqoy Galbeed	802,740	138,912	255,761	44,590	1,242,003
13	Togdheere	483,724	57,356	154,523	25,760	721,363
14	Sool	120,993	13,983	187,632	4,820	327,428
15	Sanaag	59,717	30,804	352,692	910	544,123
16	Bari	471,785	65,483	133,234	49,010	719,512
17	Nugaal	138,929	31,047	213,227	9,495	392,698
18	Mudug	381,493	79,752	185,736	70,882	717,863
19	Galgaduud	183,553	52,089	214,024	119,768	569,435
20	Hiran	81,379	135,537	252,609	51,160	520,685
21	Middle Shabelle	114,348	249,326	100,402	51,960	516,036
22	Banadir	1,280,939		-	369,288	1,650,227
23	Lower Shabelle	215,752	723,682	159,815	102,970	1,202,219
24	Bay	93,046	463,330	195,986	39,820	792,182
25	Bakool	61,928	134,050	147,248	24,000	367,226
26	Gedo	109,142	177,742	144,793	76,728	508,406
27	Middle Juba	56,242	148,439	131,240	27,000	362,921
28	Lower Juba	172,861	161,511	124,335	30,600	489,306
	Total	5,216,392	2,806,787	3,186,965	1,106,751	12,316,894

8. Illustrations of computations of urban and rural population estimates

Table 2: Population Estimates for the Urban Woqooyi Galbeed (Complete EAs provided in Excel Spreadsheet separately)

REGION	DISTRICT	OLD EA IDENTIFICATION	NEW EA IDENTIFICATION	DESIGN WEIGHT	COMPOSITE WEIGHT	NUMBER PERSONS IN EA	ESTIMATED POPULATION
A	B	E	F	G	H	I	J=HxI
12	1	07b	12010010022	9.9405780	9.4465048	534	5,044.43
12	1	11c	12010010033	14.3373720	13.6247665	496	6,757.88
12	1	15d	12010010045	12.8541960	12.2153077	526	6,425.25
12	1	19b1	12010010057	7.7660770	7.3800818	1068	7,881.93
12	1	22b	12010010068	14.9108670	14.1697573	637	9,026.14
12	1	25a	12010010080	13.0797080	12.4296117	646	8,029.53
..
..
..
..
..
..
12	2	14A	12020010072	5.6480560	5.3673318	890	4,776.93
12	2	19b	12020010078	6.6566370	6.3827730	846	5,399.83
12	2	002	12020060001	3.6149908	3.6149908	691	2,497.96
12	3	6c	12030010015	5.9170110	5.6229199	777	4,369.01
12	3	14b	12030010035	5.2875420	5.0247363	765	3,843.92
12	3	019	12030010049	5.1339700	5.1339700	755	3,876.15
12	3	19b	12030010050	5.4024880	5.1339700	755	3,876.15
12	3	34	12030010096	7.6075850	7.2294678	547	3,954.52
12	3	041	12030010122	6.1110110	5.8072780	560	3,252.08
12	3	47c	12030010151	5.1416780	4.9521523	540	2,674.16
Total urban population for Wogooyi Galbeed							802,740

Table 3: Population Estimates for the Rural Mudug (Excel Spreadsheet provided separately)

REGION	DISTRICT	OLD EA IDENTIFICATION	NEW EA IDENTIFICATION	DESIGN WEIGHT	COMPOSITE WEIGHT	NUMBER PERSONS IN EA	ESTIMATED POPULATION
A	B	E	F	G	H	I	J=HxI
18	1	2	18010010001	0.638117	1.257867	1088	1368.56
18	1	5	18010040001	2.576875	5.079585	422	2143.58
18	1	1	18010080001	2.642884	5.209703	796	4146.92
18	1	10	18010160001	1.737951	3.425882	476	1630.72
18	1	12	18010260001	2.538919	5.004766	539	2697.57
18	2	6	18020020001	2.573661	5.07325	420	2130.76
18	2	1	18020170001	2.56947	5.064988	1103	5586.68
18	2	1	18020400001	2.606926	5.138822	268	1377.20
18	2	1	18020560001	4.645069	9.343313	221	2064.87
18	2	1	18020720001	4.569285	3.033295	760	2305.30
18	2	1	18020730001	1.726208	3.402734	462	1572.06
18	2	1	18030520001	2.537869	5.002695	331	1655.89
18	3	1	18030070001	0.860198	1.695638	1829	3101.32
18	3	2	18030070001	0.860198	1.695638	1829	3101.32
18	3	3	18030070001	0.860198	1.695638	1829	3101.32
18	3	1	18030110001	2.472375	4.939452	449	2217.81
18	3	1	18030140001	1.277394	2.518023	555	1397.50
18	3	2	18030310001	2.747084	5.415103	707	3828.48
18	3	9	18030360001	2.550287	5.293767	1205	6378.99
18	3	6	18030520001	2.537869	5.002695	1008	5042.72
18	4	1	18040040001	2.485739	4.934202	794	3917.76
18	4	5	18040150003	0.948221	0.91674	545	499.62
18	4	7	18040230001	2.580594	5.086915	871	4430.70
18	4	6	18040410001	2.574093	5.074101	798	4049.13
18	4	3	18040580001	2.322534	4.578223	1104	5054.36
18	4	19	18040970001	1.266139	2.495838	622	1552.41
18	5	1	18050030001	2.49857	4.925228	690	3398.41
Total rural population for Mudug							79,752

9. Population estimation of nomads at water-points

Design assumptions vs sample data

The estimation of population of nomads was based on the weights derived from the number of water points allocated to a given region and the mean number of watering episodes.

As agreed during the Djibouti meeting, the single day model was to be used. However, the review for nomadic population estimates done indicated that the time duration for watering did not conform to the assumption of the design i.e. watering of camels – 12 days, cattle – 2 days and shoats – 4 days. Consequently, further adjustment was made in consideration of the actual duration of the watering episodes.

For example it was noted that the camels were watering at closer intervals of time contrary to the 12 day period assumed in the design of the survey. This means that within the 12 day period there would be more camels appearing at the water points than expected.

Thus if the watering interval was earlier understood to be 12 and yet in the interviewing period it is found to be 4 days, then there will be more camels at the water points during the interviewing period than expected in the theoretical formulation. In such cases, the adjustment was based on the ratio of the number of days, n to 12.

Adjustment of population estimates of pure nomads was therefore based on adjustment in the time interval of watering to conform to the mean time intervals observed in the sample data.

Water-point types

There were cases where water-point types were missing. In such instances, the water-point types earlier indicated in the geo-file /water-point frame used for sample selection were used.

10. Illustrations of computations of nomadic population using the adjusted weights

Nugaal Region

REGIO N	Identification for water points	Waterin g interval s in Days	Weighting for water points in the region	Water Point Type	Number of Households watering at the water- point	Number of persons watering at the water- points	Design Weight	Ad. Factor for the Weights	Weights	Estimate of the Population
A	B	C=Mean at each water point	D=Number of water points in the frame divided by number selected	E	F	G	H=C*D	I=Based on the mean watering intervals for adjusting watering intervals	J=I*H	K=F*J
17	170102200224	3	6.40	1	45	201	19.20	0.68	12.98	584
17	170107000704	2	6.40	1	75	521	12.80	0.68	8.66	649
17	170203800384	6	5.00	2	87	513	30.00	0.68	20.29	1,765
17	170108800884	3	5.00	2	45	316	15.00	0.68	10.14	456
17	170107300734	2	5.00	2	28	211	10.00	0.68	6.76	189
..
..
..
..
..
..
..
..
..
..
..
17	170203800384	2	5.00		87	513	10.00	0.68	6.76	588
17	170205800584	2	5.00		54	338	10.00	0.68	6.76	365
17	170105900594	1	9.57		11	42	9.57	0.68	6.47	71
17	170201800184	1	9.57		34	228	9.57	0.68	6.47	220
17	170202600264	1	9.57		20	132	9.57	0.68	6.47	129
17	170102200224	1	6.40		45	201	6.40	0.68	4.33	195
17	170102600264	1	5.67		70	430	5.67	0.68	3.83	268
17	170108300834	1	5.67		14	97	5.67	0.68	3.83	54
17	170108800884	1	5.00		45	316	5.00	0.68	3.38	152
17	170204100414	1	5.00		85	500	5.00	0.68	3.38	287
17	170205800584	1	5.00		54	338	5.00	0.68	3.38	183
Total for Nugaal Region										213,227

11. Use of high resolution imagery as a validation tool

Recent high resolution satellite imagery was used in PESS as a method of obtaining population estimates for inaccessible sampled areas and also as a tool for validation and quality control.

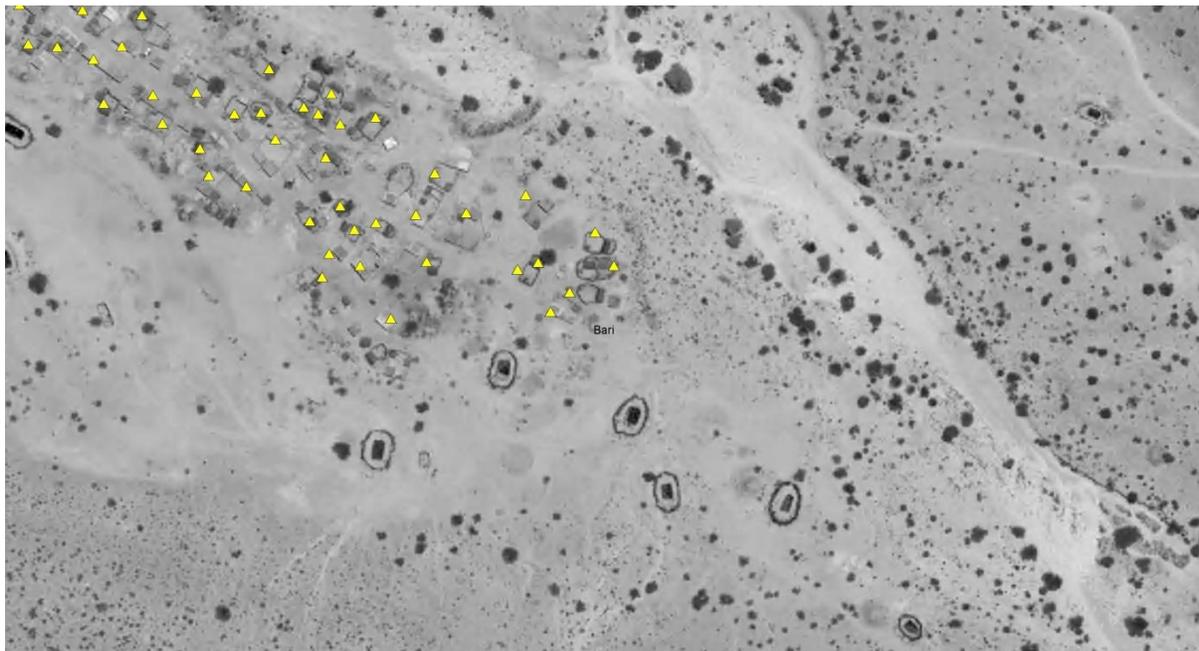
Methodology

A 1x1km grid was over-layed with high resolution satellite imagery and roof tops were manually identified and counted.

In rural settlements, one roof top was considered equivalent to one household.

In urban areas EA's were validated against the geo-file and sparsely populated EAs identified and necessary corrective measures taken.

Example of manual counting of roof tops (in yellow) in a rural settlement



KEY FINDINGS

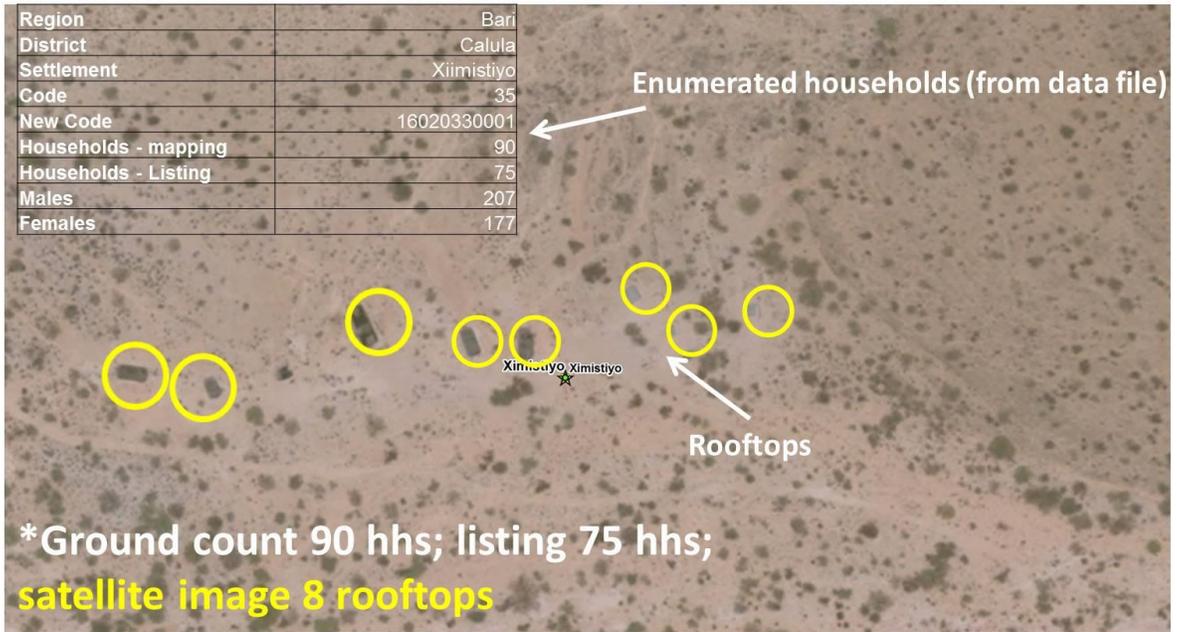
FINDING 1: Generally in urban areas there was convergence among the ground count data, survey data and, satellite imagery

Case of Beletweyne – Hiraan Region

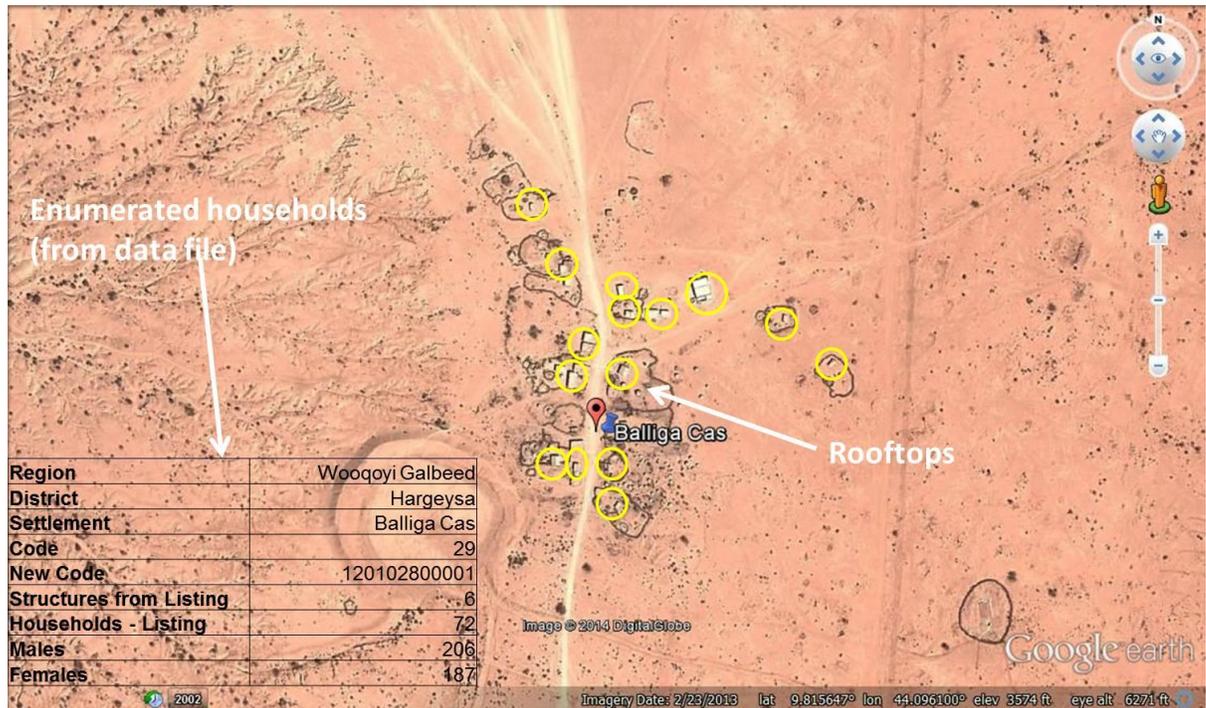


FINDING 2: In rural areas, ground counts were over-estimated compared to satellite images.

Case of Ximistiyo – Bari Region



Case of Baliga Cas – Wooqoyi Galbeed Region



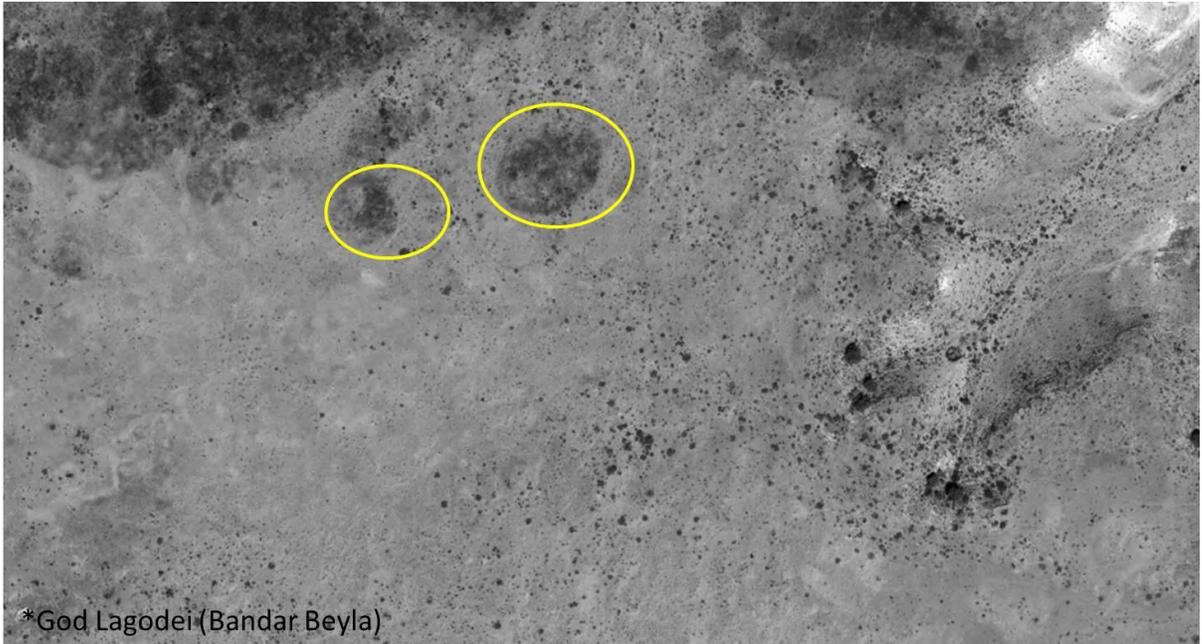
FINDING 3: Observed changes in nomadic living patterns

- Visible evidence of nomadic homesteads no longer in use

Case of Hananley – Gedo Region



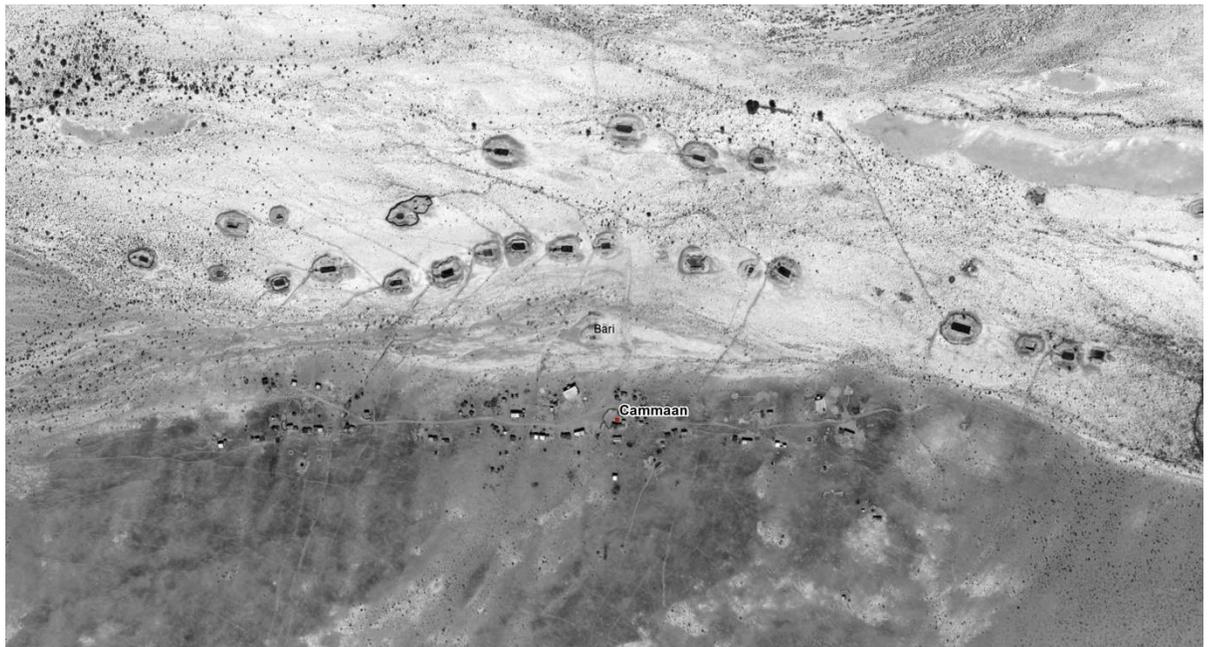
Case of God Lagodei – Bari Region



*God Lagodei (Bandar Beyla)

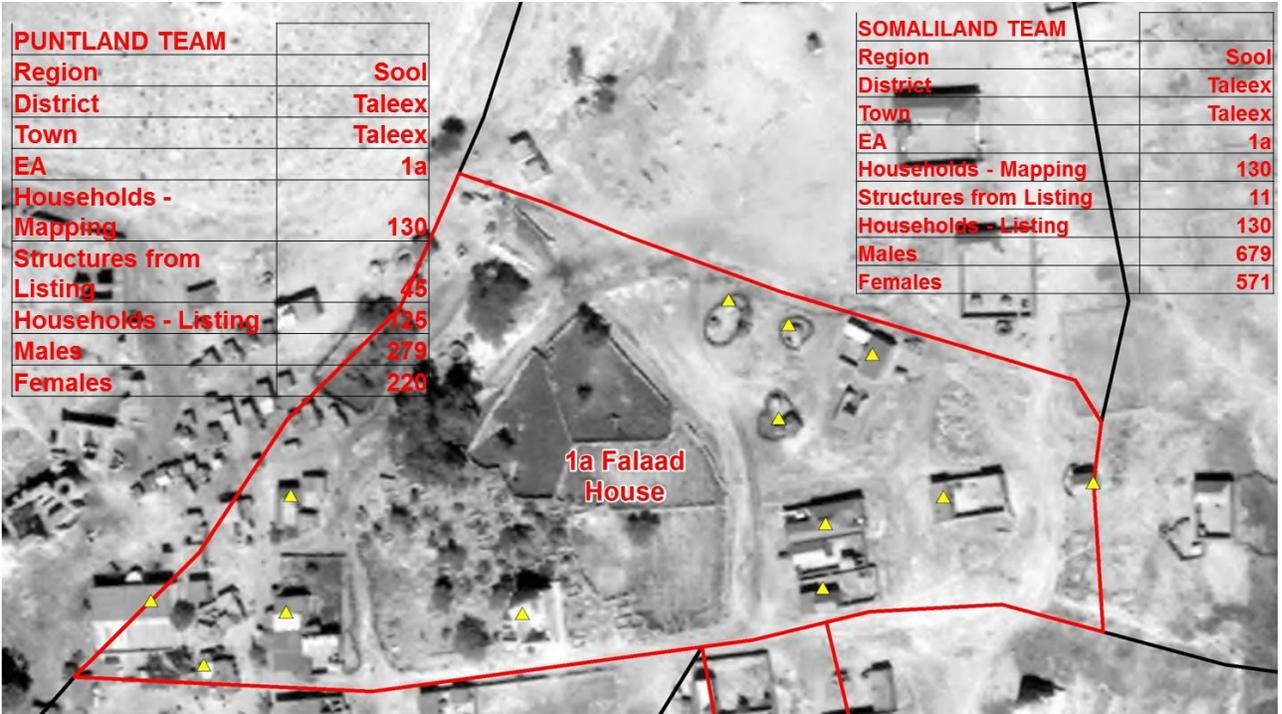
- Found clusters of settlements with man-made water points.

Case of Camaan – Bari Region



FINDING 4: In areas where the same population was counted by different teams, there were huge variations which required the use of the satellite images to validate

Case of Taleex – Sool Region



PUNTLAND TEAM	
Region	Sool
District	Taleex
Town	Taleex
EA	1a
Households - Mapping	130
Structures from Listing	45
Households - Listing	125
Males	279
Females	220

SOMALILAND TEAM	
Region	Sool
District	Taleex
Town	Taleex
EA	1a
Households - Mapping	130
Structures from Listing	11
Households - Listing	130
Males	679
Females	571

<p>PL Team listing 125 hhs; Total Population: 499</p>	<p>SL Team listing 130 hhs; Total Population: 1,250</p>
<p>Satellite Imagery: 11 Rooftops</p>	

Case of Badhan – Sanaag Region



PL Team

listing 120 hhs;

Total Population: 1,106

Satellite Imagery: 15 Rooftops

SL Team

listing 30 hhs;

Total Population: 614

FINDING 5: While digitizing the roof tops, we found some settlements as reported by the respondents were located outside the district and regional boundaries and did not tally with the shape-files for the pre-war regional and district boundaries.

12. Conclusion

The foregoing information provides the methodology and its validation mechanism applied to come up with the population estimates. The methodology was developed with the full involvement of the Somali experts at every stage of the PESS process. UNFPA Technical Support Unit (TSU) wishes to acknowledge the high commitment and professional demonstrated throughout the entire process by the Somali partners.

13. Data Sets (Attachments in excel spreadsheet)

1. Computation of weights for Wogooyi Galbeed Urban
2. Estimation of population for Wogooyi Galbeed Urban
3. Estimation of population for Mudug
4. Computation of weights and estimation of nomad's population of Awdal, Nugaal and Galgaduud regions.