Working area:

Canada and part of Greenland

Datasets Resources:

GHS POP:
This spatial raster dataset depicts the distribution and density of population, expressed as the number of people per cell, that represent the whole world.

World pop:
Spatial raster dataset with the estimated total number of people per grid-cell, that represent the whole world.

Intercomparison:

is a comparison of the two datasets.

Data set pre-processing:

- Clip raster maps.
- Aligned World-Pop raster to the GHS-POP.
- Adjust null values.

Outputs:

In this step we generate three rasters, GHS-POP and World-POP aligned and clipped to the extent of our tiles, and raster of difference between datasets after aligned and clipped.

Statistics:

We computed the basics statistics of the difference map (GHS-POP-World-POP) by **r.univar** function and calculated the correlation matrix between (GHS-POP-World-POP) by **r.covar** function for the complete raster, also calculated for each tile.

- **Univar statistics for differences rasters:**

Maximum value	-3026
Minimum value	3839
Mean	-0.0012
Mean of absolute value	0.3944
Standers deviation	7.4284
Variance	55.1807
Coefficient of variance	-587786.7781

- **Correlation of GHS-POP and World-POP:**

Number of elements compared(N)	22196784
Correlation	0.981640

Correlation for each tile:

Tile ID	Correlation
5	1
6	1
7	1
8	0.972812
36	0.232142
37	0.925651
38	0.912044
39	0.907061
40	0.959147
41	0.909506
42	0.835254
43	0.926008
44	0.904549
72	0.934738
73	0.791802
74	0.572592
75	0.928283
76	0.934104
77	0.967805
78	0.969716
79	0.911005
80	0.902448
108	0.499241
109	0.843254
112	0.943614
113	0.983941
114	0.981225
115	0.978466
116	0.981269

Validation:

We are using cross validation for our data because we have a large data, for applying validation for our datasets, we pre-process it to:

- obtain a thematic map with the value 2 (populated) and value zero (not populated). Then compute the difference between GHS-POP and World-Pop that get following scenario:
- Value 0 means GHS not populated (built), but Word-Pop populated (built).
- Value 1 means GHS and Word-Pop either both populated (built) or unpopulated (not built).
- Value 2 means Word-Pop not populated (not built), but GHS populated (built).
- Generate 600 samples with QGIS AcATaMa plugin were classified by QuickMapService plugin (Google satellite, Bing satellite, i.e.) used as reference of validation.

Finally, we generate accuracy statistics for both GHS POP and World-POP classified and the error matrix by using a python script.

Accuracy indexes are:

- **Overall accuracy (OA)** is proportion of correctly classified pixels in total number of pixels.
- **User's accuracy (UA)** of a class shows how often a user of thematic raster can expect to find the class on the ground.
- **Producer's accuracy (PA)** of a class is probability that the class present on the ground is also captured by classification in the thematic raster.

Sampling:

Preformed by QGIS AcATaMa plugin, we Stratified random sampling scheme using difference raster for stratification, distance between sample points 0.08 deg , and generate 600 samples that distributed as follow:

- 33% of points where GHS not populated, but Word-Pop populated
- 34% of points where GHS and Word-Pop either both populated or unpopulated.
- 33% of points where Word-Pop not populated, but GHS populated.

Outputs:

- Reclassified GHS-POP and World-Pop in GeoTiff format.
- Map of differences (GHS-POP – World-Pop).
- Error matrices and associated accuracy indexes for GHS-POP and for World-Pop validation.

- Vector layer of classified sampling points.

Statistics:

Error matrices and accuracy indices for GHS-POP and World-POP thematic rasters.

- GHS-POP validation:

Overall Accuracy: 68.15

- World-POP validation:

Error matrix:

Class	1(classified)	2(classified)
1(World-POP class)	306	88
2(World-POP class)	83	123

Accuracy indices:

Class	User`s accuracy	Producer accuracy
1/ not populated	0.78	0.79
2/ populated	0.6	0.58

Overall Accuracy: 71.5