Exposure to Hazards Assessment based on 'POP-to-GUF' Methodology

Sequence of Steps

An Operational Manual using QGIS

[Final Draft] Aug 2018

Prepared for Expert Group on Disaster-related Statistics for Asia and the Pacific[1]

Introduction:

This manual provides step-by-step instructions for implementing a methodology for calculating high resolution gridded-population and urban (built-up area) maps and statistics in GIS for overlay with the available maps of hazards (e.g. potential flood inundation maps) for estimating population exposure to hazards – one of the core components of disaster risk measurement. The methodology is implemented using freely available open source GIS software tools, in particular QGIS and related applications or plugins, most notably SAGA tools. Increasingly, computations and production of statistics in GIS can be expedited by write code for the conversions or calculations using Python script. Working with GIS can involve a need for tests or trouble-shooting and there is often more than one way to accomplish the same function. This manual provides the step-by-step instructions for operating within the QGIS or SAGA platforms.

For this manual, we utilize a real-life example and real data for Thailand. The estimates are calculated based on the official national population census data (counts of population by most detailed available geographic areas) along with geospatial data, particularly the Global Urban Footprint (GUF) processed and made available by the German Space Agency (DLR), based on radar satellite imagery. Other relevant geospatial data inputs utilized include a map of land cover types, which was produced by the European Space Agency (ESA) Climate Change Initiative, and hazard area maps accessible online. The methodology integrates information from the population census into the GUF gridded data to estimate population density by grid cell, thus in short-hand it is called the 'Pop-to-GUF' method.

For this manual, QGIS version 2.14 is recommended (more stable than the 2.18 and 3.03 recent releases). The package includes GDAL, GRASS and SAGA processing modules. In some cases, for convenience or efficiency, calculations be done with the native SAGAGis 6.3.0 and re-imported to QGIS or similarly with any other GIS package using a raster format read by QGIS. Thus, it's recommended to also download the SAGA open source package (http://www.saga-gis.org/en/index.html), which is inter-operable with the QGIS.

[1] This manual developed by Jean-Louis Weber (Consultant) with Daniel Clarke (UNESCAP), Trevor Clifford (Consultant), and Gao Xian Peh (Consultant), with valuable feedback from the Disaster Management Agency of Indonesia (BNPB), the Dept. of Disaster Prevention and Mitigation (DDPM) of Thailand and the National Statistical Office (NSO) of Thailand.

Table of Contents

1 Collect data

This first section of the Operational Manual summarizes the basics on the datasets, by file type, that will need to be downloaded and organized in your project folder.

1.1 Create a folder and name it Downloaded Data

1.2 Statistical Tables

Tables are collected for Population census statistics. Optional is to collect tables on other topics relevant to assessing exposure or vulnerability to hazards, where available, such as crop production statistics, for example. Statistical tables can be integrated into GIS formats if they contain a reference to a place or geographic location, such as government administrative regions (e.g. districts).

Tables can be handled and processed with different software packages:

- **Spreadsheets**: e.g. MSExcel (.xls and .xlsx), LibreOffice or OpenOffice (.ods),
- **Data Base Management Systems**: e.g. POSTGRESQL/POSTGIS, MSAccess, SQLite/Spatialite. Common tables formats are .dbf (Note: .dbf headings of 8 digits maximum), .csv ot .txt
- **GIS attribute tables to shapefiles** (and some grid file formats); they can be viewed with spreadsheets as .dbf or .csv file (Note: MSExcel cannot save tables as .dbf; LibreOffice and Open Office can).

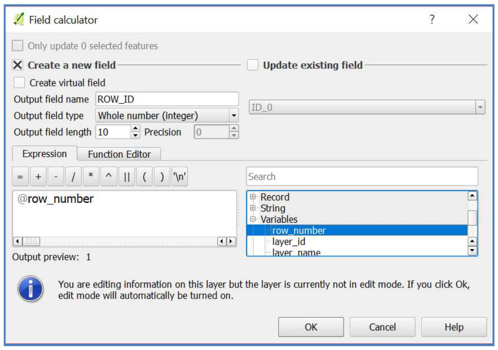
The pollution statistics tables are downloaded in spreadsheet format (.csv or .xlsx). Cases exist where detailed statistic are supplied in .pdf format only. It is possible to convert .pdf to .xls or .csv but it requires careful and long check of the tables as well as elimination of parasitic graphic features.

QGIS reads and saves tables in .dbf fomat. Conversions to .dbf format is done, in principle using Save As... or Import/Export functions within QGIS, but not all packages can read all formats.

Tables used by DBMS and GIS calculations record:

In rows: shapes, features, statistical units (e.g. administrative divisions), objects...
In columns: fields, attributes (e.g. name, codes, variables' values)

IMPORTANT : In rows, shapes' internal IDs are automatically given by the GIS when the shapefile is created. It corresponds to the order of the rows in the initial shapefile. It means that this order should NEVER be modified. GIS package manage this issue. Spreadsheets don't visualize the internal ID. As a security, it is recommended upon importing table into QGIS to first create a new field named **ROW_ID** using QGIS vector calculator, and requesting @row number.



The statistical tables are joined with shapefiles in GIS only if the imported statistics and shapefile have common identifiers, ideally common administrative codes (geographic codes identifying location for the relevant values). Often coding or naming of geographic areas vary, so careful consistency checks are needed to join the statistics into your GIS project.

From the population census, collect statistics for the most detailed administrative divisions available (usually ADM2 or ADM 3) and the hierarchy of levels. ADM0 stands for the whole country. ADM1 usually stand for regions and ADM2 for districts, but the terminology and division patterns vary from country to country. More detailed levels vary even more from country to country. Statistics on Population Censuses by ADM0, ADM1 and ADM2 are available on international websites (e.g. CityPop, World Bank). National statistical offices increasingly provide open access to population census data, including that data at the most detailed level (e.g. villages), with geographic names and codes to use as common identifiers for linking with shapefiles.

1.3 Vector files (shapefiles)

Vector files record points, lines (e.g. roads) and polygons (e.g. administrative divisions) as a drawing to which is attached an attribute table. The commonly used format for vector files is ESRI shapefile (.shp). It is a de facto standard for ArcGis, QGis, SAGAGis. There are other formats used such as for example in MapInfo (.mif), AutoCAD (DWG) or Google Earth (kml, kmz). Conversion between vector files format is done by the various GIS packages.

The shapefiles to download for POPtoGUF methodology are administrative divisions. They can be obtained through mapping agencies, international websites such as *GDAM, GitHub, or OpenStreetMap* or directly from the National Statistical Office or other responsible government agencies. Use of government official shapefiles is preferable for best consistency between shapes and tables (common ADM codes). When shapefiles and tables come from two different sources, careful check of codes is necessary.

Other shapefiles used in our assessment are hazard perimeters or zones (unless in raster format). There are usually national data available based on modelled probabilistic assessments of likelihood of hazards under various scenarios. The national maps of hazard areas are an important resource for disaster reduction, often produced by mapping, environmental, or disaster management agencies at the national level. These resources should be shared with other agencies, e.g. national statistics offices, for statistical purposes. International sources also exist, including hazard area vector or raster files, at the global scale. Unfortunately, currently, the global assessments are not yet available at high levels of resolution for use at national or lower scale assessments, except in pilot or sample exercises like the one used in this manual.

High resolution land cover maps are also needed Land cover maps are usually available as shapefiles when produced at the national level or regionally in the case of ICIMOD, for example.

1.4 Raster files

Raster files (or grid files) are images made of pixels (like digital photos) to which are attached geographical coordinates. Raster files are convenient formats for supporting large amounts of data and image processing algorithms used in GIS.

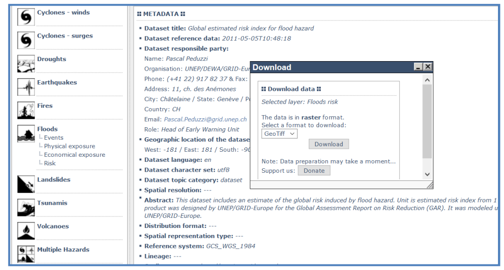
A well known basic raster format is .tif or .tiff which is commonly called GeoTiff when it is used in GIS. A Geotiff is in fact the association of a .tif file (which can be viewed in any image viewer) and a small file providing the information on the geographical referencing of the pixels of the map. Many other raster formats exist. Modern software packages can easily convert maps from one format to another.

Raster files needed for POPtoGUF are:

- Global Urban Footprint (GUF) : the built up area 2012; the standard distribution at 80 m of resolution, obtained from the ESA U-TEP platform run by DLR. Note: GUF is accessed via user agreement with the German Aerospace Agency (DLR). See also, TEP Urban platform (urban-tep.eo.esa.int /
- Land Cover maps**:** needed to map areas with dispersed population not sensed by GUF; In this example we used a publicly available international source: the European Space Agency Climate Change Initiative (CCI) global land cover at 300 m of resolution is an acceptable input, easy to download (https://www.esa-landcover-cci.org/). Higher resolution land cover maps are available in some countries or from regional intuitions such as ICIMOD, often in vector format.
- Maps of hazards or risks perimeters/areas: needed either as shapefile or in raster format. For test purposes, global maps can be accesses as raster files via the international data sharing platforms of UN Environment GRID Portal

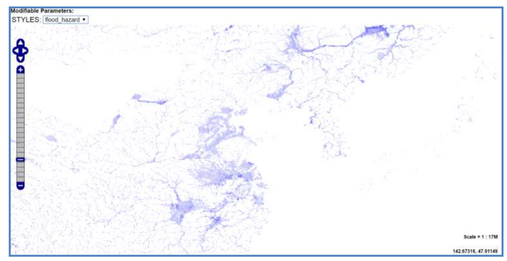
(http://www.grid.unep.ch/index.php?lang=en and the Group on Earth Observations (GEO) (www.geoportal.org).

For the purpose of this manual, data have been downloaded from the UNEP GRID Global Risk Data Platform http://preview.grid.unep.ch/



Also available at: www.geoportal.org

Flood hazard has been retained for the present example. The flood hazard layers of historical flood hazard probability occurrence over 100 and 500 years produced for the GAR2015 assessment have been recovered from http://preview.grid.unep.ch/geoserver/gwc/demo/GAR2015:flood_hazard_500_yrp?gridSet=EPSG:4326&format=image/png ➔



The Darmouth Flood Observatory at University of Colorado (a contributor to UNEP products) is developing its website with new products http://floodobservatory.colorado.edu/

2 Upload layers to QGIS and prepare downloaded data

In the step we upload the datasets into GIS and begin to process the data for the layered analysis. As some uploaded global layers may be very large and cannot be computed easily, some steps are includes with the purpose of downsizing the files or eliminating unnecessary data or layers from the project. This includes clipping global maps to the project extent in order to avoid processing irrelevant data.

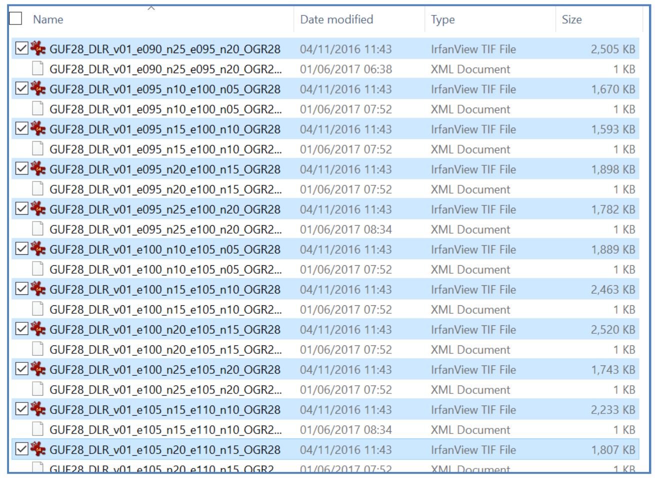
2.1 Upload the national boundaries shapefile (ADM0)

Use:

The same button is available on the vertical set, left of the pane: . The ADM0.shp is the map of the whole country. It will be the extent of the project.

2.2 Upload and pre-process GUF Data

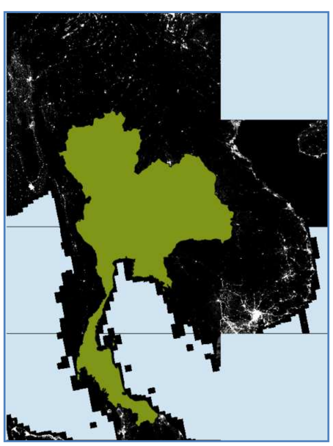
This can be done from the Menu simply as:
Layer /Add Layer /Add Raster Layer



Raster files can be supplied as rectangular tiles representing relevant quadrants of the globe to avoid downloading and processing useless data. GUF tiles are named according to the geographical coordinates of their top-left and bottom-right corners.

For Thailand, the tiles are:

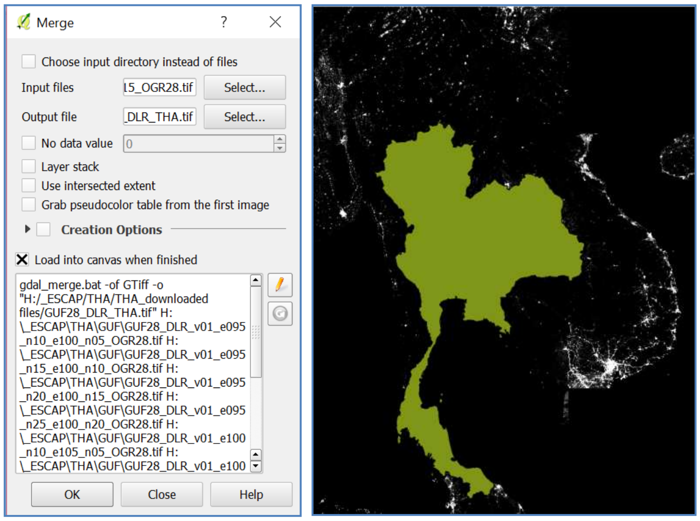
GUF28_DLR_v01_e095_n10_e100_n05_OGR28.tif
GUF28_DLR_v01_e095_n15_e100_n10_OGR28.tif
GUF28_DLR_v01_e095_n20_e100_n15_OGR28.tif
GUF28_DLR_v01_e095_n25_e100_n20_OGR28.tif
GUF28_DLR_v01_e100_n10_e105_n05_OGR28.tif
GUF28_DLR_v01_e100_n15_e105_n10_OGR28.tif
GUF28_DLR_v01_e100_n20_e105_n15_OGR28.tif
GUF28_DLR_v01_e100_n25_e105_n20_OGR28.tif
GUF28_DLR_v01_e105_n15_e110_n10_OGR28.tif
GUF28_DLR_v01_e105_n20_e110_n15_OGR28.tif



Note: pixel values from the GUF .tif files are 0 and 255 (they will be converted later to 0-1)

2.2.1 Merge GUF tiles into one layer

Use QGIS/Raster/Miscellaneous/Merge



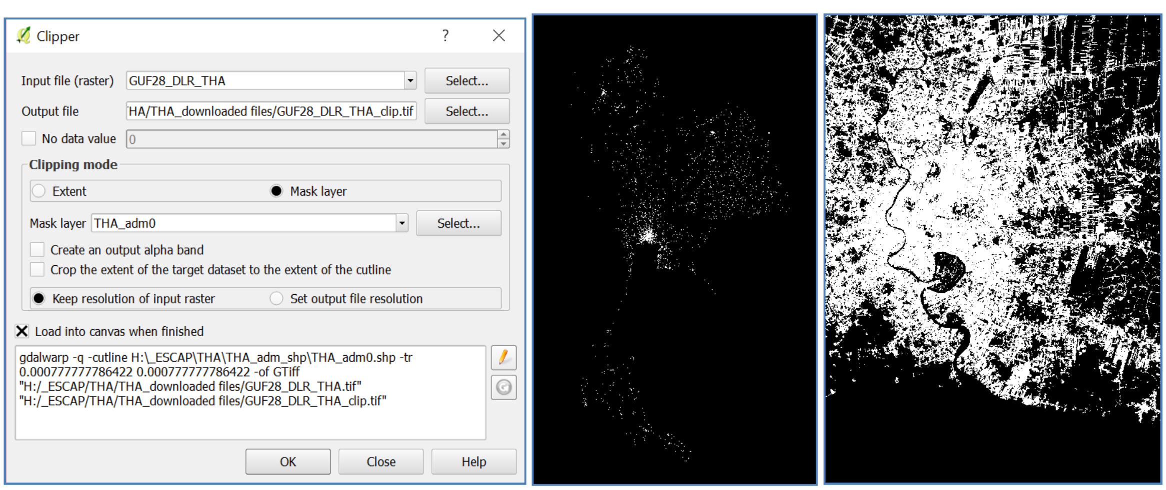
Select all input tiles and save the merged output file as **GUF28_DLR_THA.tif**.

2.2.2 Clip out unused GUF data

The pilot study used in this example produces output for the country of Thailand. Therefore, unused data are data outside the national (ADM0) boundary.

The **QGIS Clipper tool** can be used to combine the GUF merged file (GUF28_DLR_THA.tif) with ADM0 to delete values out of the boundary. In terms of the project/grid system, GUF, as all global datasets, is provided in EPSG:4326. Therefore, we can select ADM0 as the **mask .shp layer** as it is also in EPSG:4326.

Use QGIS/Raster/Extraction/Clipper



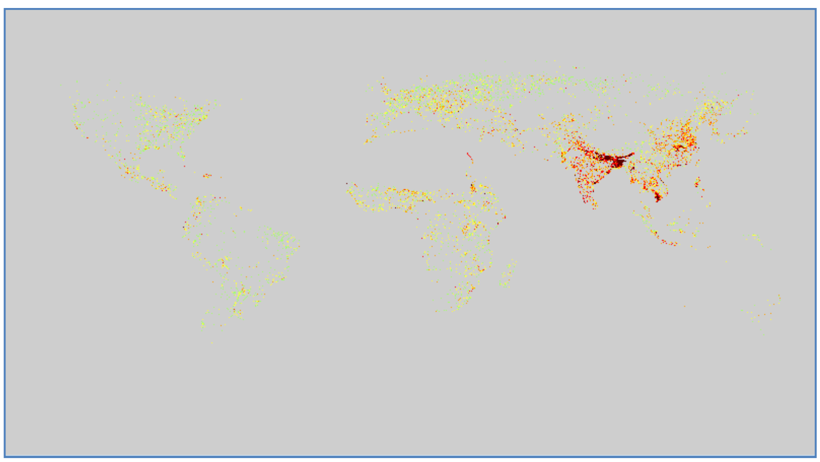
Save as **GUF28_DLR_THA_clip.tif**

2.3 Upload and pre-process land cover and hazard maps

The ESA-CCI land cover map is used in order to assess population not living in buildings/houses detected by GUF pixels. This file is 2015 map with 300m (grid-size) resolution, version 2.07. In short-hand the file is: ESACCI-LC-L4-LCCS-Map-300m-P1Y-2015-v2.0.7. The ESA-CCI land cover map is a single file for the global dataset (in EPSG:4326 coordinate system). We will extract a work area for our project in order to avoid heavy calculations on irrelevant data.

The ESA-CCI land cover contains values with reference to a classification of land cover types following the Land Cover Classification System (LCCS) developed by the United Nations Food and Agriculture Organization (FAO). For more information on this dataset, see the ESCA-CCI User Guide: http://maps.elie.ucl.ac.be/CCI/viewer/index.php.

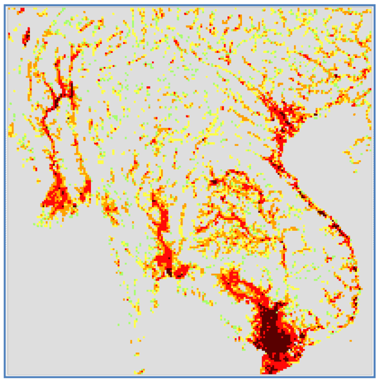
For this example pilot study, download the global flood risk map from UN Environment Grid Portal or GeoPotal. We accessed the flood risk map file called fl1010irmt.tif (from Grid Portal). It displays flood risks in 5 classes of severity with raster cells of 10x10 km size.

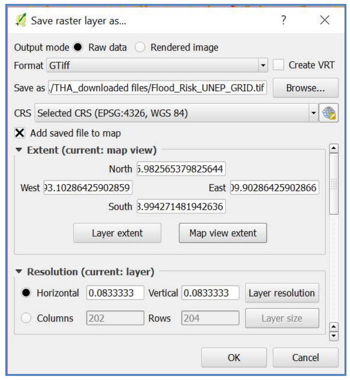


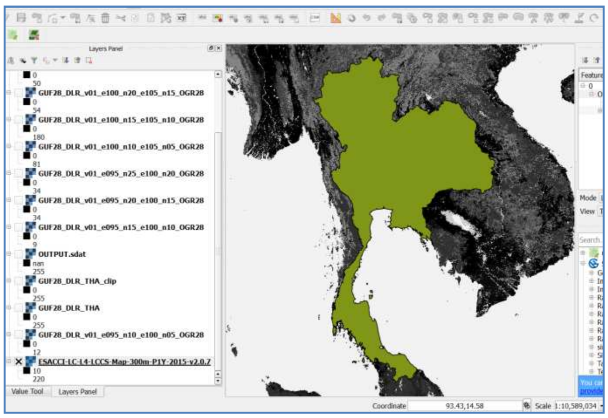
2.3.1 Extract a working window from the global ESACCI land cover and hazad maps

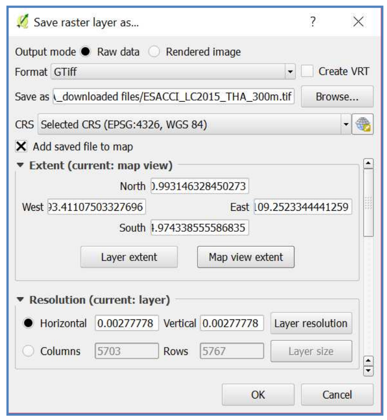
- **Zoom ESACCI-LC** (and adjust manually the QGIS window size) so that the Map View Extent matches the ADM0 map.
- **Right-click on ESACCI-LC** to open "**Save As**"... Select **Map View Extent** and save under **ESACCI_LC2015_THA_300m**.

Repeat these steps for the flood hazard layer





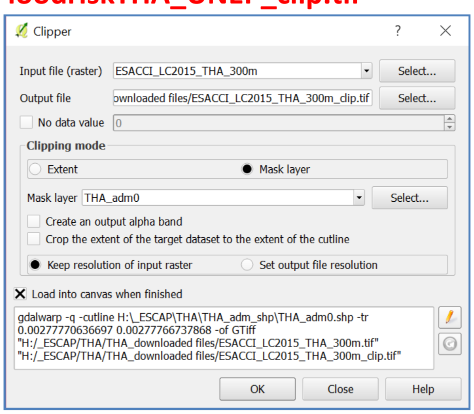


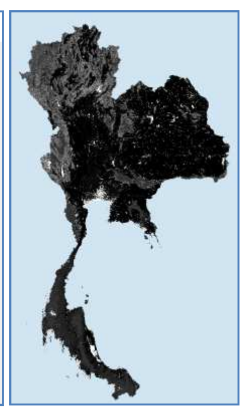


2.3.2 Clip out unused raster data

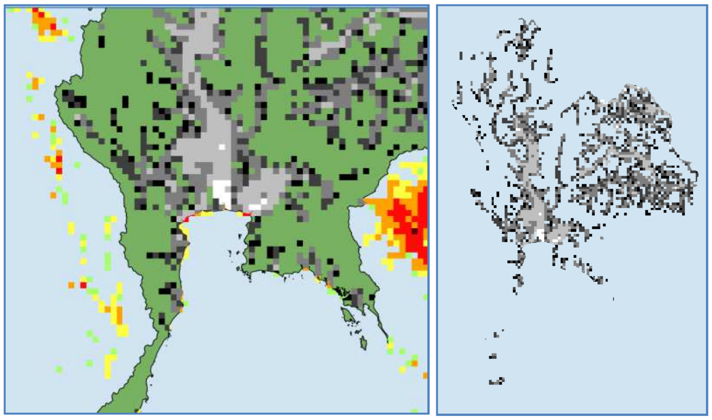
Use the QGIS Clipper – same as in Step 2.3.2

Save as…in THA_Input folder with a new name, e.g.: **ESACCI_LC2015_THA_300m_clip.tif** and **FloodriskTHA_UNEP_clip.tif**





•



Results for the hazard data after initial processing are displayed here in a ramp of greys.

2.4 Save the project

Now we have uploaded and layered the key datasets within our project area and it is a good moment **to save the project** (QGIS project file) so that we don't lose the progress thus far. Use the Save As... button on left-top. This v0 project will be kept as a link to archives. Save the file with a short but descriptive name with indication of the version, e.g. Save as **THA_v0.**

A GIS project is the collection of maps (layers of input files) uploaded and processed, and the results of calculations, with characteristics such as styles (colours...). A project can be saved at any moment of the work. When reopening QGIS, you can go directly to state of the last time that you saved. To avoid problems in the course of the application, it is recommended to define carefully the project at the very start, which includes extent and projection system.

3 Prepare Shapefiles for Assimilation of Population Data

The following steps apply specifically to this example case study of calculating 'Pop to GUF' for Thailand and these specifications can vary and be adapted for other applications of the methodology.

3.1 Define project projection system (CRS: Coordinates Reference System)

By default, QGIS opens in "WGS84 EPSG:4326" which is the global system of geographical coordinates expressed in Arc Degrees, Minutes and Seconds. For calculating values related to areas, the spherical coordinates need be converted to meters, which means projecting them to a flat surface. The projection depends from the latitude and longitude of the country (the Earth is not a perfect sphere!), as well as of the geodic model used; there is no single solution. There are systems of projection leading to acceptable accuracy of areas measurements by zones, the most precise ones being defined by countries (and even regions within countries). For the purpose of the application, the projection system used is called **World Mercator (code name:**

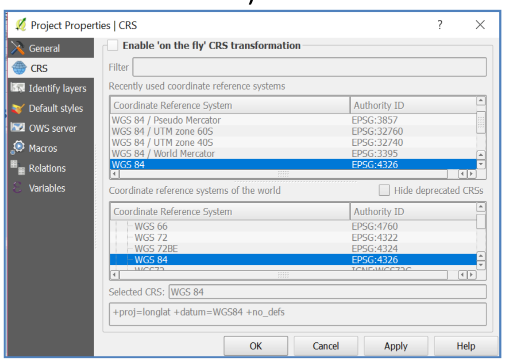
WGS84 EPSG:3395).

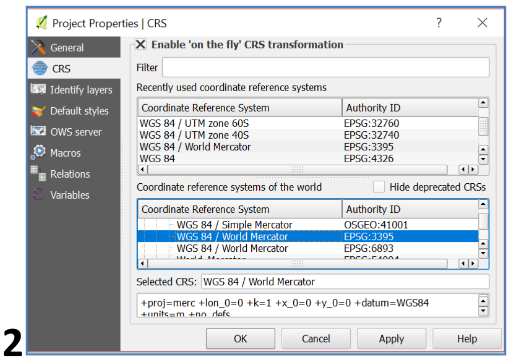
3.1.1 Set the CRS for the project

At the bottom-right of the window, **click o**n 

. You get the first impage below, where CRS (Coordinates Reference System) is EPSG:4326.

Click on Enable 'on the fly' CRS transformation and **select WGS84/World Mercator EPSG:3395** in the pane Coordinate reference systems of the world as in the second image below (you may have to browse the list). **Click OK.**

1 

2 

Save the project as **THA_v1**

Now, when you upload a map to this project (THA_v1), it will be displayed 'on the fly' to EPSG:3395 (the chosen projection system) so that it can be visualized and overlaid with other layers.

Note: if you are having difficulties with projection, it is possible to force this display [see QGIS user assistance]. Calculations can be done as well for each file but original files (raster or vector) will keep their own EPSG. This means that it is not possible to combine 2 layers having different EPSG (e.g. clipping a raster file with a shapefile, merging two shapefiles ...). Therefore, it is safer to convert the input layers used for the project into the project EPSG. It can be done by saving them under a different name and giving them in the saving box the desired EPSG.

In the EPSG3395 project, grid size measurements are in meters.

3.2 Upload the national boundary (ADM0) shapefile to QGIS

3.2.1 Upload the national boundary (ADM0) shapefile to QGIS

Proceed as in Step 2.2., using the  button.

3.2.2 Check the CRS of ADM0 in the Properties/General box

If the map is registered in the project's CRS, there is nothing to do (it may be the case with nationally provided maps...) → **Step 4**

However, If the map is not registered in the project's CRS (e.g. because it has been downloaded from international source such as GDAM and is in WGS84 EPSG:4326, or for any other reason), we must re-project it to the project's CRS → **Step 3.3**

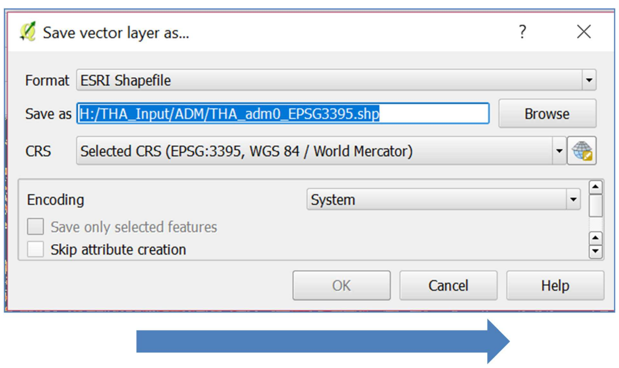
3.3 Project (or re-project) ADM0 to the project CRS

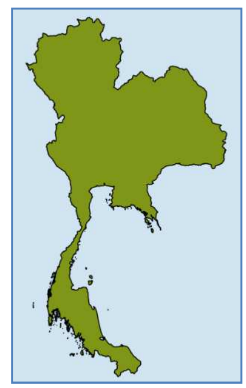
Create a new directory and name it THA_input.

Select the shapefile layer (ADMO) and Save As... a new shapefile named THA_adm0_ EPSG3395 and placed in the new folder THA_Input\ADM.

While saving, **change the displayed CRS** to the Project's new CRS (in this case, the project's CRS is WGS84/ World Mercator EPSG:3395). **"Save as", but don't type the file name in the box. Instead, click on Browse, go the appropriate folder and type the name there (to declare the full path...).**

ADM0.shp

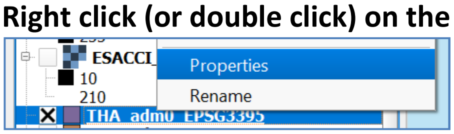




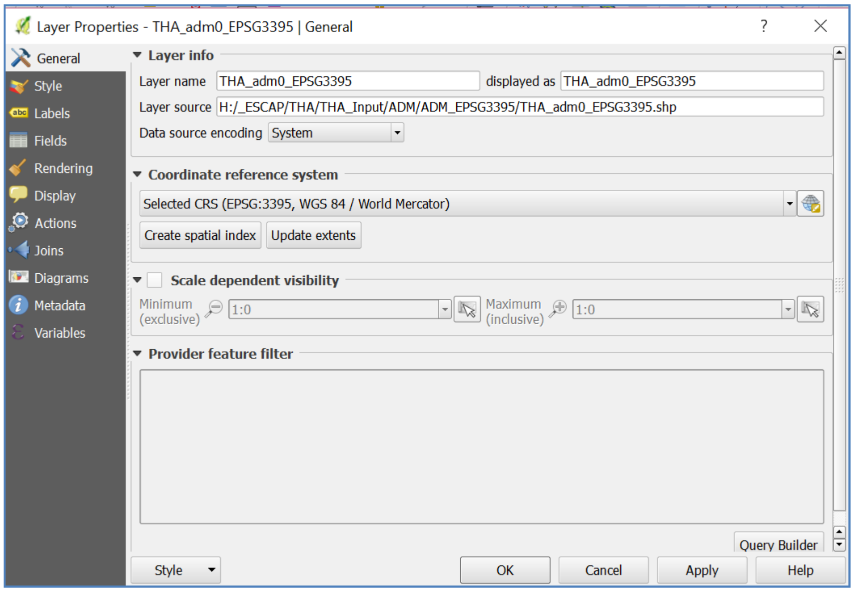
THA_adm0_EPSG3395.shp

3.3.1 Verify the new coordinate reference system (CRS) for ADMO

Right click (or double click) on the layer's name to open the Layer Properties box:



and go to General\Coordinate reference system:



Remove the old ADM0 file from THA_v1

3.4 Project (or re-project) other Shapefiles (as needed)

3.4.1 Upload the ADM1, ADM2, ADM3 shapefiles to THA_v1

Other national shapefiles for Thailand will be needed for integrating the census data into the computations. If these files are in geographical coordinates (WGS84 EPSG:4326) or another CRS, they have to be re-projected into the project's CRS.

Same method as above, i.e.: **Save As... change the displayed CRS** to the Project's new CRS (in this case, the project's CRS is WGS84/ World Mercator EPSG:3395). **"Save as", but don't type the file name in the box. Instead, click on Browse, go the appropriate folder and type the name there (to declare the full path...).**

The destination folder is **THA_Input\ADM**.

The new files (with CRS WGS84/World Mercator, EPSG:3395) are:

THA_adm1_EPSG3395.shp
THA_adm2_EPSG3395.shp
THA_adm3_EPSG3395.shp

4 Create a raster mask layer with properties for data assimilation

Raster calculations involving several layers usually require identical pixels size and alignment. Some software packages require that raster files have the same extent. QGIS can accommodate various extents in one calculation, except when using the SAGA and GRASS tools, which require files to have exactly the same extent.

A common data assimilation grid chosen for convenience for raster files is 100m x 100m pixels.

For convenience in computations, the raster files are resampled to this format (100mX100m grid) when they have smaller (e.g. GUF) or larger (e.g., in this case, land cover, hazard map...) pixels. This is accomplished with what's called a mask layer -essentially a reference template for the project.

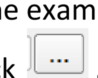
For this methodology, we will actually develop two common reference mask raster layers. The first one will be made of 0 and NoData cells and is used for additions. The second mask layer contains values of 1 and NoData cells and is used for multiplications. Note: addition to or multiplication by a cell with value of NoData results in NoData.

These masks will be produced by rasterizing the **THA_adm0_ EPSG3395** created at the previous step. They will be used as exact references when resampling raster files, when rasterizing new shapefiles, or when clipping out results for Thailand.

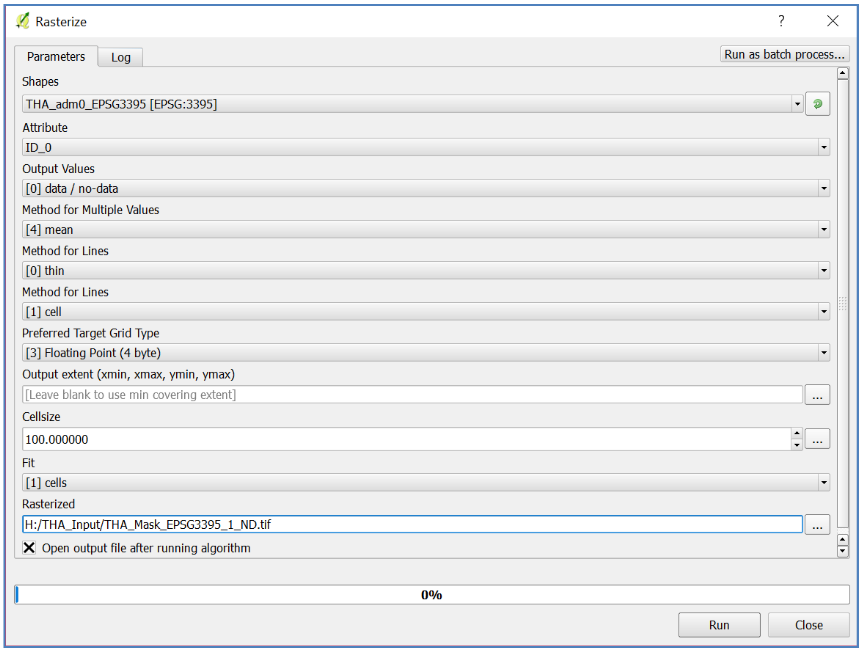
4.1 Creation of the 1_No Data Mask with 100m pixels

Go to QGIS Processing Tool Box.
Menu: Processing: Toolbox. Open SAGA/Raster creation tools/Rasterize.

In **the command box, select the THA_adm0_EPSG3395.shp vector layer.**
Keep all default values for all parameters in the Rasterize function are kept EXCEPT for:
Output Values: change from default ([2] attributes) to [0] data/no data
Fit: change from Nodes to Cells
Cellsize: set to 100 (in case when default value is not 100.00000 as in the example)

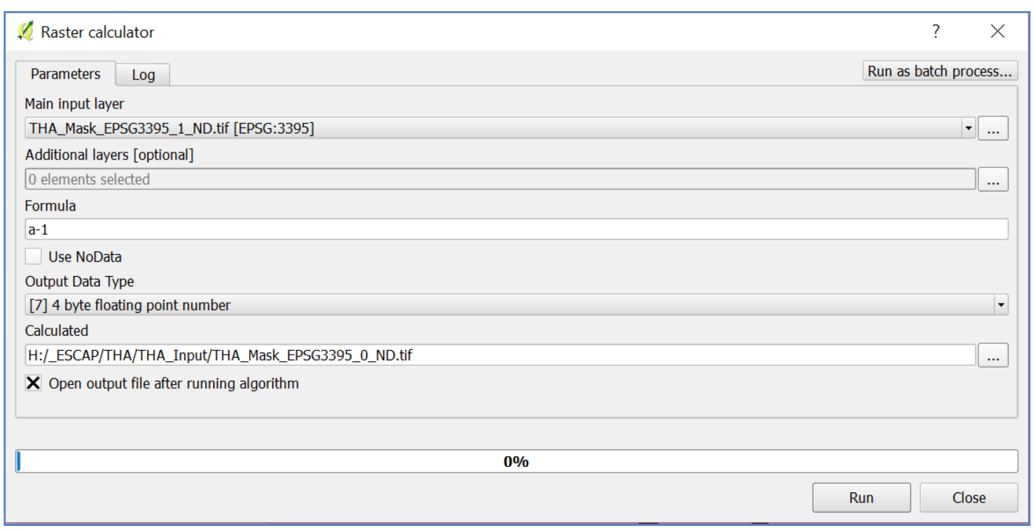
Change from "Save to a temporary file" to **"Save to file..."** and then click  **,** go to the **THA_Input\ folder** and give the name **THA_Mask_EPSG3395_1_ND.**

Notation note: "1_ND" l indicates that this mask has 1 and No Data values.



4.2 Creation of the 0_No Data Mask with 100m pixels

Go to QGIS/Processing Tool Box/SAGA/Raster calculus/Raster calculator
Select the new mask raster layer created above (THA_Mask_EPSG3395_1_ND) as the input layer
Formula: a-1
Save Calculation as: THA_Input/THA_Mask_EPSG3395_0_ND.tif



Save the Project.

5 Assimilation of Input data

5.1 Prepare Population Census tables

The required data is population from the census, by administrative regions. Generally, the lower level of the groupings (e.g. districts or villages) gives more accurate outcome but the methodology can be implemented at different scales, in this case with data by districts.

For the THA exercise, the Population Census table were downloaded from the CityPop database (https://www.citypopulation.de/php/thailand-admin.php). From this download, a single file was created with 2000 and 2010 statistics: **CityPop2000_2010_Districts.csv**.

Districts codes corresponding to shapefiles used have been inserted in the table. NOTE: This requires a careful check as common attributes used were names, which sometimes are subject to spelling variants.

5.2 Introduce Population Census statistics into the project shapefile

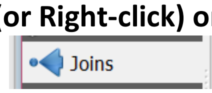
5.2.1 Upload CityPop2000_2010_Districts to the project.

If the format is .csv, upload with the “Add vector layer” button 

If the format is .xlsx or .ods, upload with the “Add Spreadsheet” button 

5.2.2 Join to the administrative region shapefile

For this THA pilot, the census data is associated with administrative region level 2 (districts or ADM2), i.e: THA_adm2_EPSG3395

- **Double-click (or Right-click) on THA_adm2_EPSG3395 and go to Properties.**
- **Click on Joins** 

and then on the + button  **[bottom left]**

For this example, we utilized a file with population data by districts for 2000 and 2010 and a common join field called CCA_2:

Join Field: CCA_2 ;
Target Field: CCA_2;
Check “Choose which fields are joined” and select your fields, e.g: CTYP_NAM, Pop_2000 and Pop_2010 ;
Check “Custom field prefix name” and delete the default text **(keep it blank)**
Click OK and then **OK**

Save the file As... THA_adm2_POP_EPSG3395 to a new folder named THA_CALC

This file will be used in further steps for calculations.

NOTE: If you forget to save this layer, the "join" will disappear when closing the project.

5.3 (Re)Project and Resample raster layers

The raster layers also need to be projected from WGS84 EPSG:4326 to the CRS WGS84/World Mercator EPSG:3395 and then to be resampled to the 100m x 100m pixels resolution chosen for this sample application for Thailand. The files to re-project and then resample are:

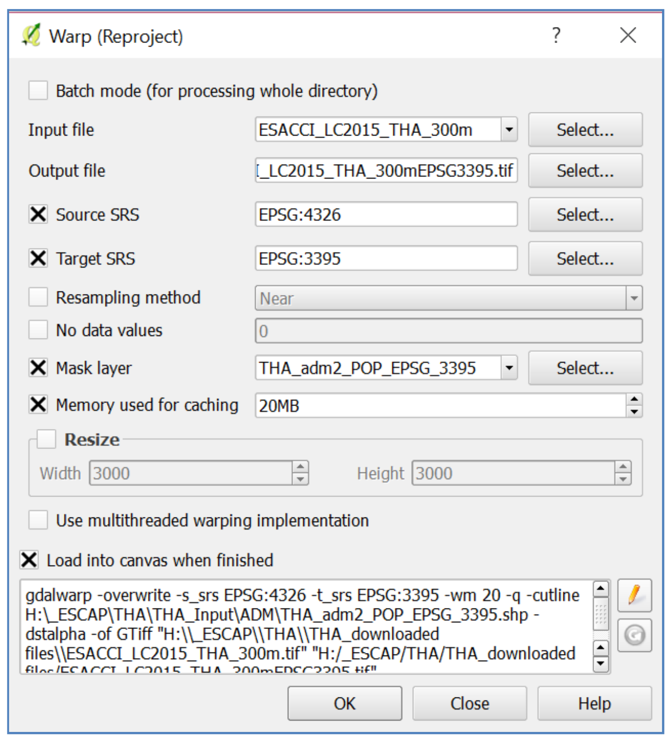
GUF28_DLR_THA_clip.tif
ESACCI_LC2015_THA_300m_clip.tif
FloodriskTHA_UNEP_clip.tif

Therefore, **conduct step 5.3.2 for each of these raster layers, as needed**

5.3.1 Re-project (convert, warp...) each of the raster layers to EPSG:3395

Use QGIS/Raster/Conversion/Warp

- **Set Input file to be converted**
- **Output files: set (save to) INPUT DATA** and the same name as in input file augmented **with EPSG3395**
- **Source SRS:** in principle, **use default value**, which in this case is EPSG:4326
- **Target CRS:** check and **set to EPSG:3395**
- **Mask layer:** use for example THA_adm2_POP_EPSG_3395



5.3.2 Remove the old (prev. projection) files from the THA_v1 project

Remaining raster files are:

GUF28_DLR_THA_clip3395..tif
ESACCI_LC2015_THA_300m_clip3395.tif
FloodriskTHA_UNEP_clip3395.tif

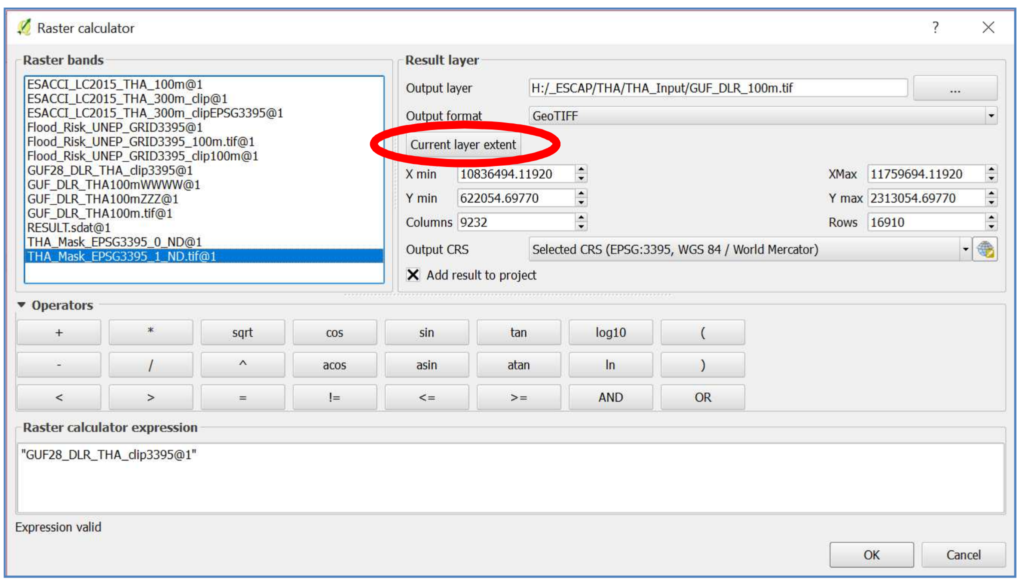
5.4 Resample raster data to common grid

Resampling raster data to 100mx100m grid is done with **QGIS/Raster/Raster calculator** where change in "current layer extent" means change in extent AND in pixel size. This function uses the nearest neighbourhood algorithm. In this case, the the aggrdation/down-scaling does not noticeably change the distribution of values. Similar resampling can be carried out with the SAGA toolbox.

5.4.1 Resample GUF data to 100x100m grid

GUF28 pixels have a size of approximately 80m in Thailand. To resample the data with the nearest neighbourhood algorithm to 100x100m cell size:

Use QGIS/Raster/Raster calculator
Raster calculator expression (formula): "GUF28_DLR_THA_clip3395@1" (introduced by **double clicking** the name in Raster bands)
IMPORTANT: In Raster bands**, SINGLE click** now on **THA_Mask_EPSG3395_1_ND.tif** AND THEN click on the **"Current layer extent" button.** The raster calculator will resample the input map to the resolution set for Current layer extent (100m), using the nearest Nnighbourhood formula.
Save output layer as **THA_Input/GUFDLR_100m.tif**

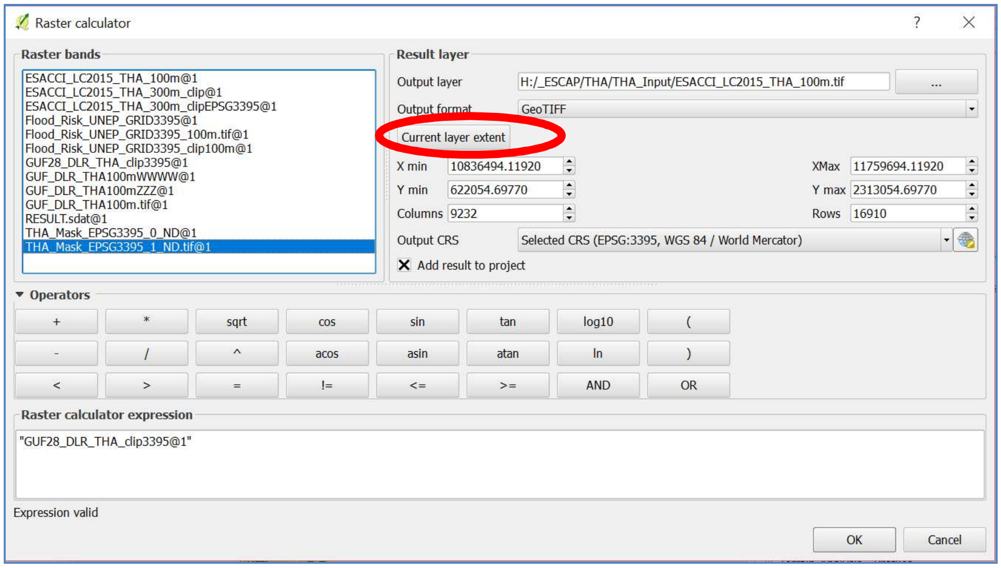


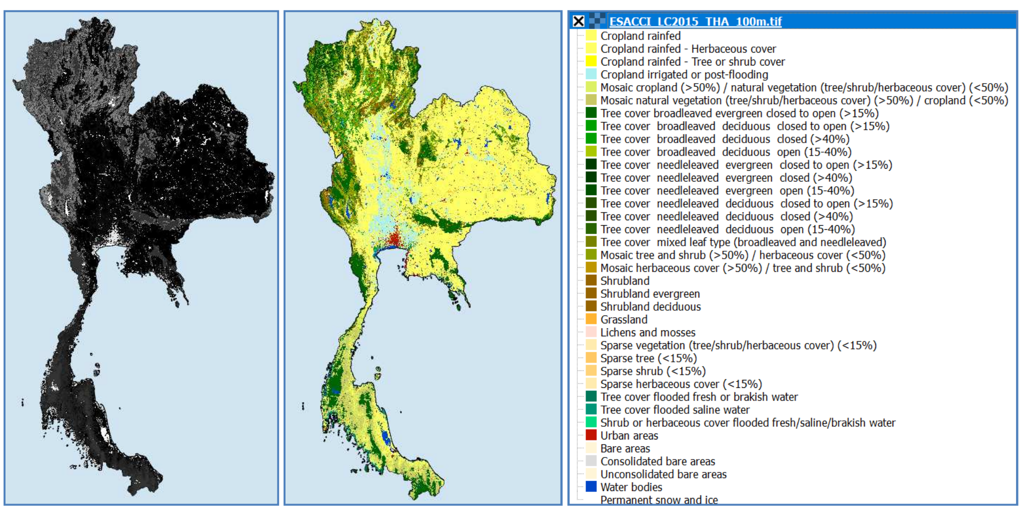
5.4.2 Resample ESACCI land cover data 100x100m grid

Pixels of the downloaded map are of circa 300 m x 300 m. For assimilation in this project, the ESACCI map is converted to a 100 m x 100 m raster.

Use QGIS/Raster/Raster calculator

> **Raster calculator expression (formula): “ESACCI_LC2015_THA_300m_clip3395@1" (**by
> **Use same procedure as in 5.4.1 and Save** output layer as
> **THA_Input/ESACCI_LC2015_THA_100m.tif**





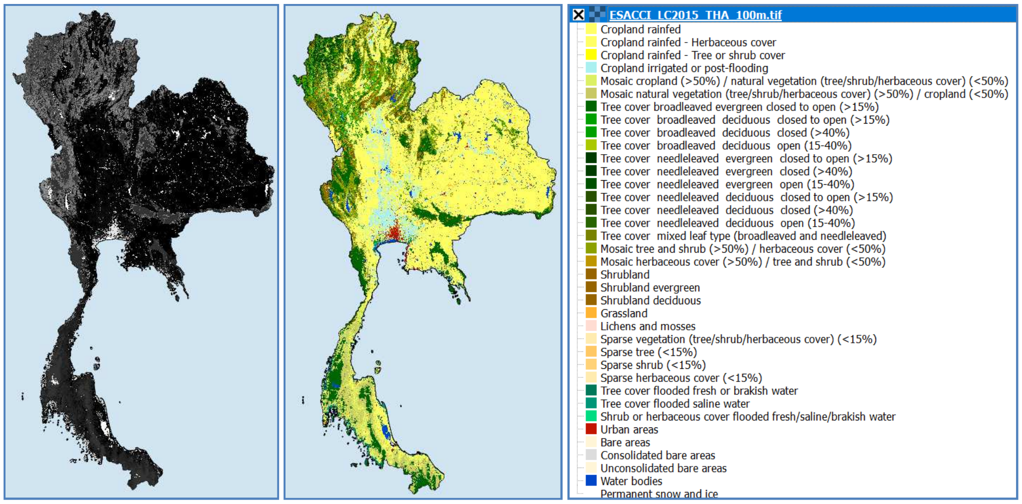
5.4.2.1 Visualize land cover classes

Included with ESACCI land cover raster data for download are metadata for the land cover classes used and a GIS (.qml) file for integrating the legend and visualization of the classes (colors) in your GIS package.

The procedure will vary depending on the version of QGIS, but generally this step is accomplished by selecting the layer and going to:

Properties/Style/Load

Select ESACCI-LCMapsColorLegend.qml

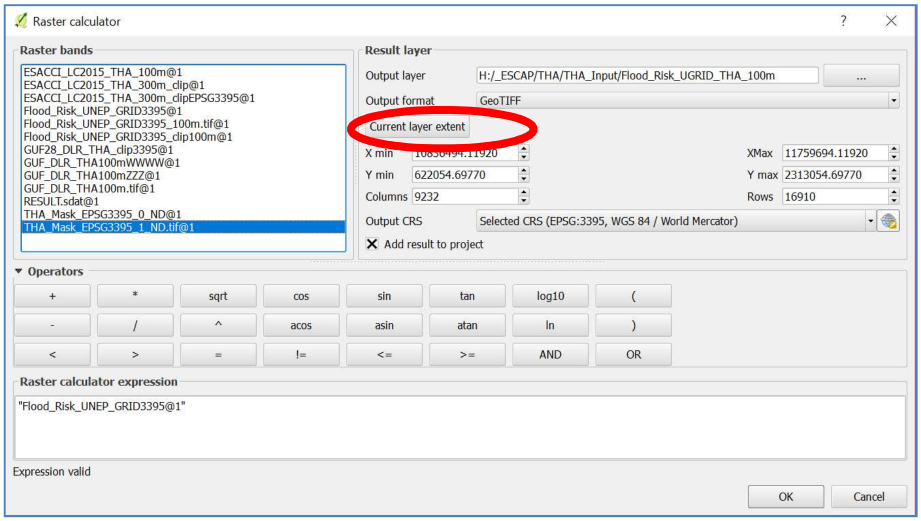


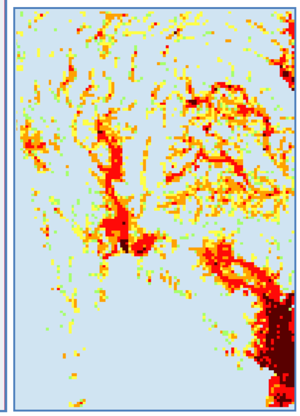
5.4.3 Resample Hazard/Risk data (Flood_Risk_UNEP_GRID3395.tif)

Pixels of the downloaded map are of circa 10 km x 10 km. For assimilation in this project, again the map is converted to a 100 m x 100 m raster.
When the nearest neighbourhood algorithm is used, it does not change the distribution of values, which are in this case categories (1 to 5).

Use QGIS/Raster/Raster calculator

Use same procedure as in 5.4.1 and Save output layer as
THA_Input/Flood_Risk_UNEP_GRID3395_100m.tif



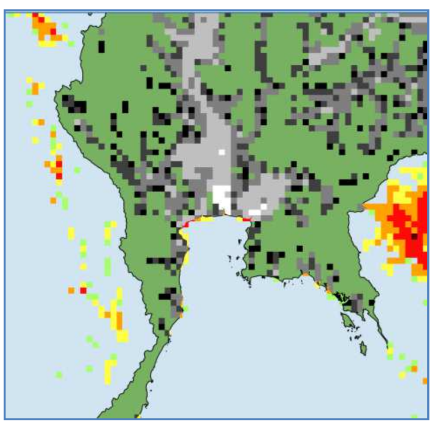
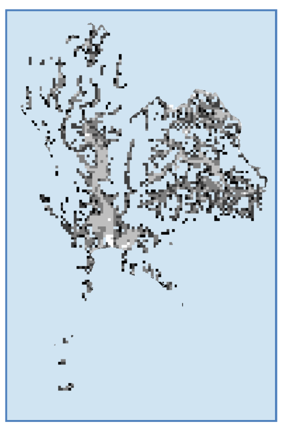


The input file includes No Data pixels and the result is ADM0 rectangular extent, not the ADM0 polygon as previously. So, it is necessary now to Clip out Flood_Risk_UNEP_GRID3395_100m.tif with ADM0.

Use QGIS/Processing Tool Box/SAGA/Vector<>Raster/Clip Raster with Polygon.

Shapefile: Adm0_EPSG3395

Save Rasterized result as **THA_Input/ FloodRiskTHA_UNEPGRID_100m_3395.tif**

Results from final processing are displayed here in a ramp of greys.

6 Weighting GUF pixels according to their agglomeration

A basic objective for the Pop to GuF methodology is to downscale population data by utilizing the GUF, knowing that an assumption of an even density of population per GUF pixel within an administrative division is not realistic. Obviously, population density is much higher in cities where land is scarce and houses and building are attached than in villages where land is abundant and individual houses are surrounded by gardens. To capture this uneven density, POPtoGUF starts from a very simple model where isolated GUF pixels are down-weighted as compared to GUF pixels in cities (or in areas of agglomerations of built-up areas) which keep the maximum values in the center and are moderately deflated in the outskirts. This model is based on Gaussian Filtering or Smoothing. The algorithm is easy to implement in GIS (e.g. with SAGA) and can be tuned (regarding smoothing radius and intensity) according to the context of the study. It is a model that can be combined with other models, for example models needed to assess the dispersed population in habitats that could not be observed by satellite (seet Step 7). Moreove, POP to GUF can be implemented at various scales, from Regional to National and local assessments, depending on the detail of population statistics available.

How does Gaussian smoothing (filtering, blurring...) work?

Smoothing is a methodology which transforms crisp data into values taking into account their neighbourhood. It gives more weight at the central pixels and less weights to the neighbours. The farther away the neighbours, the smaller the weight. The process is repeated all over the map (or the image) with a mobile window. In this simplified example of a 5x5 kernel, one cell of a value of 100 (on a 0-100 scale) is surrounded by cells with zero value. Once smoothed, the central value drops to 15.02, the total of the whole kernel remaining at 100. If this cell had been surrounded by cells with values greater than zero, it would have received value in return. In the case where all neighbouring cell have a 100 value, the result of these exchanges would have been 100 for the central cell.

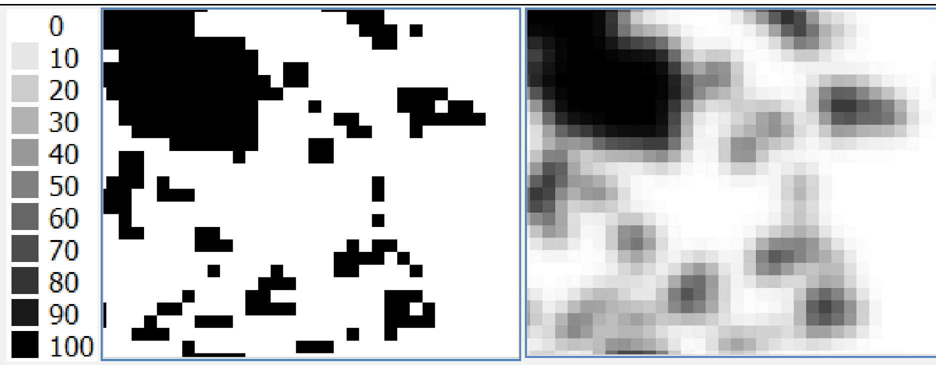
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	100.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

➔

0.37	1.47	2.56	1.47	0.37
1.47	5.86	9.52	5.86	1.47
2.56	9.52	15.02	9.52	2.56
1.47	5.86	9.52	5.86	1.47
0.37	1.47	2.56	1.47	0.37

The program computes (and re-computes successively) the values of all pixels using a mobile window.

Regarding the issue of weighting GUF pixels according to their agglomeration, the smoothing methodology keeps the value of towns or agglomerated clusters of urban areas (still in black on the right hand picture) while it reduces that of small cities and villages (in grades of grey).



This property is used to weight the average population density (per administrative divisions) according to the size of human settlements (higher density in large towns than in small cities and then in villages...). At the end, total population data (net of dispersed rural population) by administrative division is distributed in proportion to the value of the smoothed pixels. The total population by administrative divisions is unchanged.

6.1 Clip-out the sea and convert GUF pixels to 1 value

GUF pixels in **THA_Input/GUFDLR_100m.tif** are given the conventional value of 255 as many raster files, or 0 (0 meaning no GUF). For calculation purposes, this 255 value will be converted to 1 (with a Real number format with decimals).

Also, a smoothed pixel exchanges values with its neighbours. This poses a problem on the shoreline as (in principle) no built-up and related population should be sent to the sea. If neighbouring cells have 0 values, coastal built-up pixels will lose value to the sea. Fortunately, the algorithm ignores NoData, so coastal pixels will lose value only on the inland side, not to the sea if we clip pixels out of THA_ADM0.

Note: This simple solution has a bias for inland borders but because we don't process data out of the project area (for this sample application) no correction is done for simplicity reasons. But an adjustment could be done for transnational projects.

These two operations will be done in one step with the SAGA raster calculator.

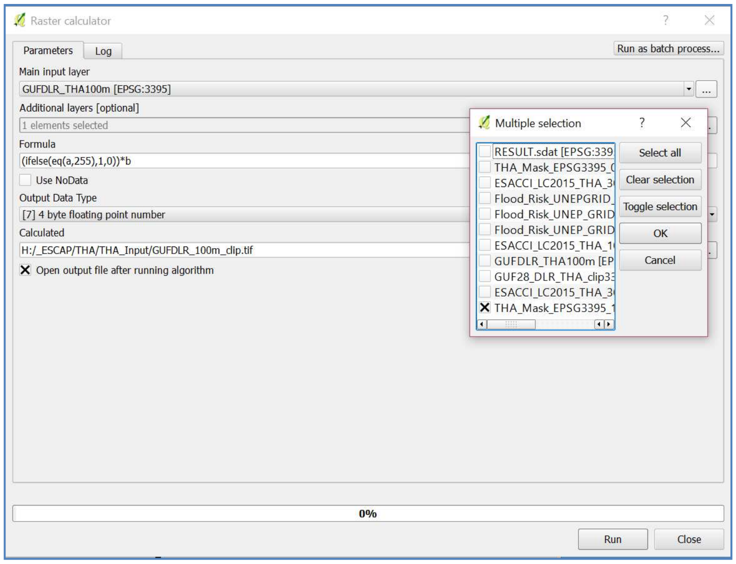
Multiply **THA_Input/GUFDLR_100m.tif** by the 1-No Data raster Mask (created in Step 4.1) and divide the result by 255 in order to set GUF pixels value at 1.

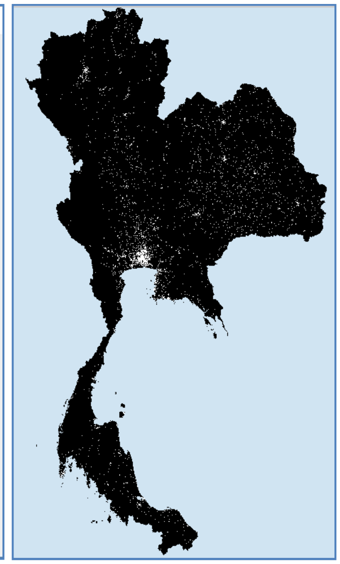
Use QGIS/Processing Toolbox/SAGA/Raster calculus/Raster calculator

Main input layer: GUFDLR_THA100m.tif [is "a" in the Formula]
Additional layers (optional): THA_Mask_EPSG3395_1_ND [is "b" in the Formula]
Formula: (ifelse(eq(a,255),1,0))*b [It reads: if "a" equals 255, the return 1; if not return 0; and lastly multiply by "b".]

Save calculated as: **THA_Input/GUFDLR_THA100m_clip.tif**

Now GUF values have been clipped to the mask layer and pixel values are converted to 0 and 1 (and NoData, outside the project area – e.g. in the sea).





6.2 Smoothing Clipped GUF data

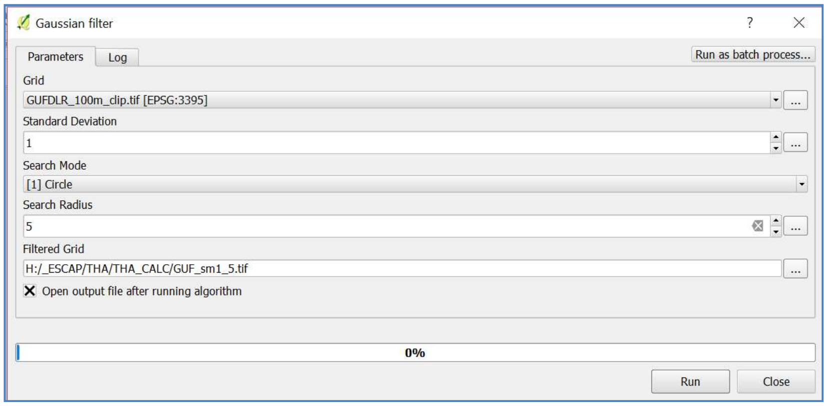
As mentioned above, the parameters (smoothing radius and standard deviation) for the Gaussian smoothing operation can be adjusted to fit with the expected reality. This is a matter of calibration and different options may be tested according to *a priori* knowledge of the way that populations are distributed across space and urban agglomerations for the specific country or project area.

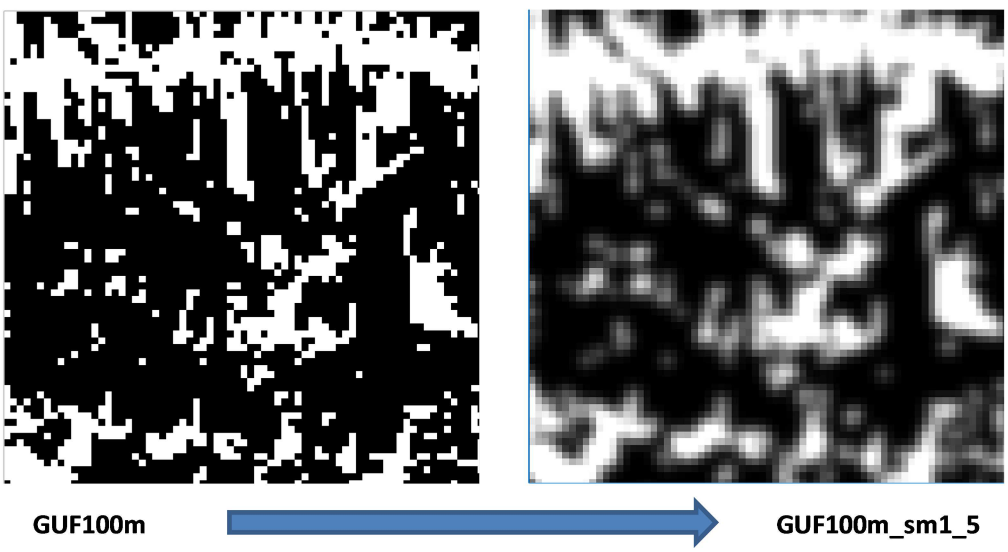
For this pilot test case for Thailand, the parameters selected are:

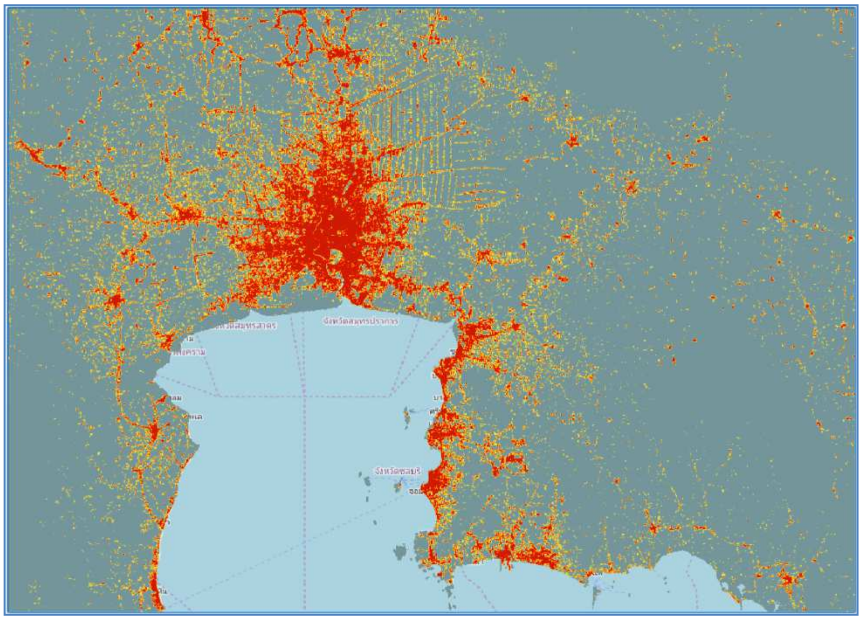
- **Smoothing radius**: a multiple of the cell size. **5** is a commonly used value and corresponds in the case of 1 ha cells to a little bit more than 1km2 in the case of a square search and a little bit less for a circular search. Higher values for the radius can be considered or tested, noting that since the values are decreasing in proportion to the square of the distance from the center pixel, they rapidly approach zero anyway.
- **Standard deviation (std):** it is the measurement of the quantity of value which will be spread out over the pixel. It is expressed in number of standard deviations in relation to the chosen radius. The larger the radius, the higher should be the standard deviation to get meaningful results. For the THA pilot project, with radius of 5**, std is 1**. This is an empirical choice, resulting from various experiments in several countries. It can be easily changed in the SAGA toolbox.

Use QGIS/Processing Toolbox/SAGA/Raster filter/Gaussian filter

> **Input: THA_Input/GUFDLR_100m_clip.tif**
> **Standard deviation: 1**
> **Search mode: Circle**
> **Radius: 5**
> **Calculated result saved as: THA_CALC/GUF_sm1_5.tif**







THA_CALC/GUF_sm1_5.tif

6.3 Clip out smoothed values outside of GUF cells (optional)

Smoothed values range from 0.0001 to 1.0000.

An effect of the smoothing is that smoothed values have been generated in the grid outside of the cells with original GUF values (cells previously with 0 value have acquired coefficient values from smoothing by cells within 5 cell radius). These smoothed values could be potentially analytical useful for other Gaussian Smoothing applications – e.g. examining the effect of the smoothing on peripherals areas of urban conglomerations and for studying nexus areas along the boundaries of urban and rural areas.

For the POPtoGUF methodology, we choose to eliminate these values in order to keep the native footprint of the map of GUF pixels. These unwanted non-native values are easily eliminated with a simple trick: multiplying the result of the Gaussian Smoothing (**GUF_sm1_5.tif**) by the clipped **GUFDLR_100m_clip.tif**. Since the non-GUF grid cells in the unsmoothed input layer (GUFDLR_100m) are equal to zero, and GUF grid cells are equal to 1, the effect of this multiplication is to convert values outside of GUF cells to zero (multiply by zero) and maintain the values within the GUF cells only (multiply by one).

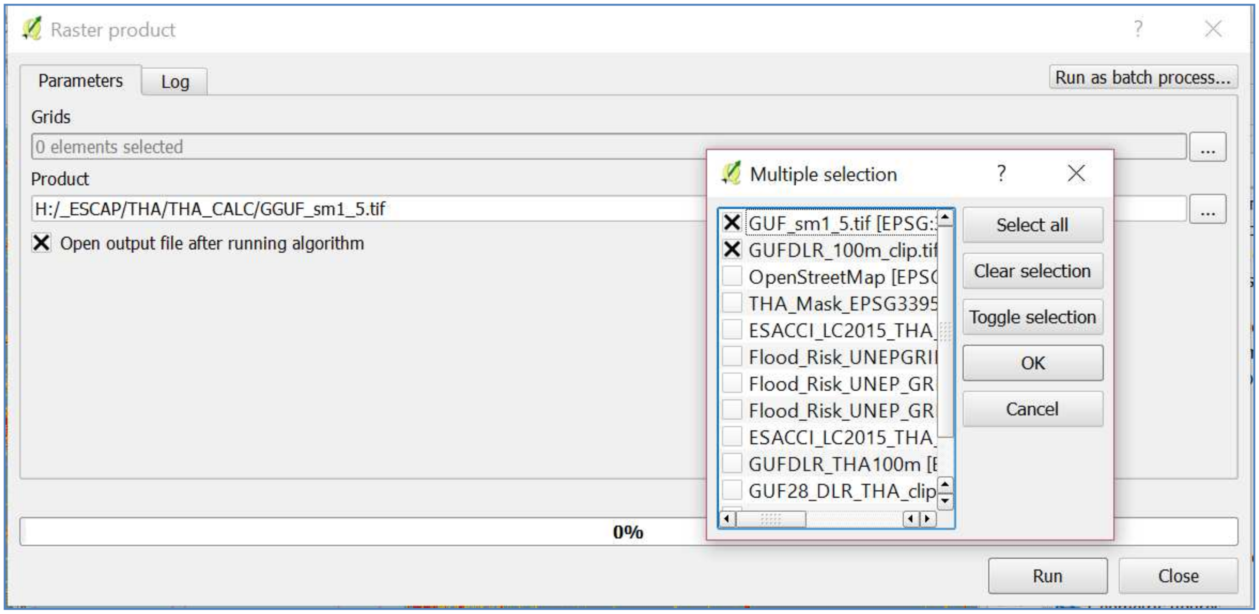
Use QGIS/Processing Toolbox/SAGA/Raster calculus/Raster product

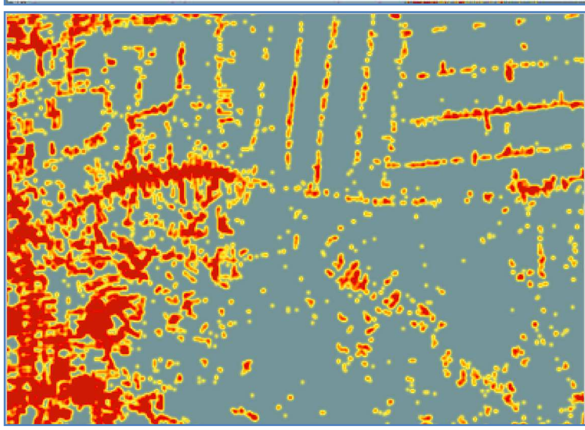
Grids : Multiple selections/ GUF_sm1_5.tif , GUFDLR_100m_clip.tif

Run

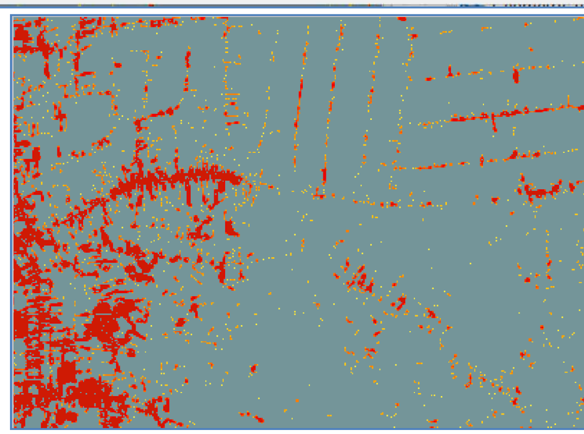
Save product as: THA_CALC/GGUF_sm1_5.tif

Notional note: We introduce the additional G in GGUF to indicated that the smoothed values have been clipped to the GUF native area.





GUF sm 1_5



GGUF sm 1_5

7 Estimate population living out of GUF pixels

At this stage we have completed all the necessary preparations and calculations for applying a gridded-population density estimation for Thailand using the GUF, under the assumption that population density, per district of the country, can be estimated by smoothed values for agglomerations (and location within the agglomerations) of built-up areas, as identified by GUF.

A core principle for the POP to GUF methodology is simplicity, which ensures that the outputs are relatively easy to understand and consistent with the population census input data.

Another advantage of simplicity in the model is that it more easily replicable and can be used as a baseline upon which more complex calibrations can be designed or tested (or identified via machine learning techniques), building on the opportunity of accessing the new high resolution GUF products.

POP to GUF can be customised to supplement standard national outcomes with tailor-made applications on areas of specific interest, such as regions with specific risks which can be covered with higher resolution data and more detailed statistics.

For Pop-to-GUF methodology, there is one such customization or calibration to the modelling which is necessary, and described in this Step 7, utilizing the ESACCI land cover raster data.

The GUF is calculated based on radar sensors from satellites. Remote sensing of houses and buildings, even with radar sensors, meets limitations. For example, isolated houses or farms with thatched roofs or under trees (plantations or forest), tents of nomadic people etc, all may be missing from the GUF cells. Therefore, the core and crucial supplemental adjustment made to the baseline assumption in Pop to GUF is to estimate locations of dispersed population in such situations, living in areas that are not mapped by GUF.

There are multiple possibilities for using land cover data to make an adjustment for dispersed populations that are hidden from the satellite imagery. For the POP-to-GUF methodology, we use the simple strategy of assigning an estimated average population density for selected land cover classed in non-GUF areas of Thailand (e.g. cropland and cropland, mosaic tree cover, and grass lands). This step is one of the simplest possible approaches based on land cover data and may be refined over time with additional information about these populations.

For the non-GUF areas, we assume an average density - e.g. 30 inhabitants per km2, or 0.3 inhabitants per hectare. This average density value can be modified according to additional information available for these populations**.**

7.1 Define area where dispersed population is expected to live outside of GUF

The area where dispersed population is expected to live is defined from a selection of classes of the ESACCI land cover maps and a correction to eliminate GUF pixels from this area in order to avoid double counts.

7.1.1 Selection of classes of the ESACCI land cover maps

The classes selected for non-GUF population estimation in the THA study are in **red** in the legend. These are the types of areas that tend to have populations not visible from the GUF directly.

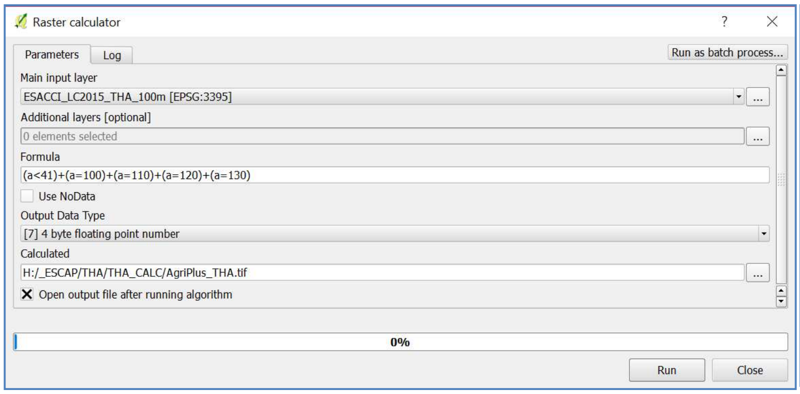
10 Cropland, rainfed

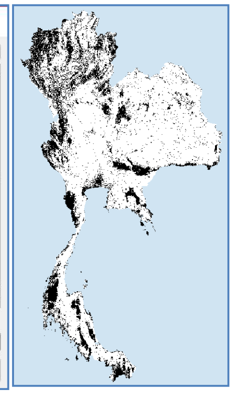
11 Herbaceous cover (cropland)
12 Tree or shrub cover (cropland)
20 Cropland, irrigated or post-flooding
30 Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous
40 Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland
50 Tree cover, broadleaved, evergreen, closed to open (>15%)
60 Tree cover, broadleaved, deciduous, closed to open (>15%)
61 Tree cover, broadleaved, deciduous, closed (>40%)
62 Tree cover, broadleaved, deciduous, open (15-40%)
70 Tree cover, needleleaved, evergreen, closed to open (>15%)
71 Tree cover, needleleaved, evergreen, closed (>40%)
72 Tree cover, needleleaved, evergreen, open (15-40%)
80 Tree cover, needleleaved, deciduous, closed to open (>15%)
81 Tree cover, needleleaved, deciduous, closed (>40%)
82 Tree cover, needleleaved, deciduous, open (15-40%)
90 Tree cover, mixed leaf type (broadleaved and needleleaved)
100 Mosaic tree and shrub (>50%) / herbaceous cover (<50%)
110 Mosaic herbaceous cover (>50%) / tree and shrub (<50%)
120 Shrubland
121 Evergreen shrubland
122 Deciduous shrubland
130 Grassland
140 Lichens and mosses
150 Sparse vegetation (tree, shrub, herbaceous cover) (<15%)
152 Sparse shrub (<15%)
153 Sparse herbaceous cover (<15%)
160 Tree cover, flooded, fresh or brakish water
170 Tree cover, flooded, saline water
180 Shrub or herbaceous cover, flooded, fresh/saline/brakish water
190 Urban areas
200 Bare areas
201 Consolidated bare areas
202 Unconsolidated bare areas
210 Water bodies
220 Permanent snow and ice

To extract the selected classes**, use QGIS/Processing Toolbox/SAGA/Raster calculus/Raster calculator**

Formula: (a<41)+(a=100)+(a=110)+(a=120)+(a=130)
Output calculated layer: THA_CALC/AgriPlus_THA

This new output layer is a combined presentation of the relevant land cover classes selected in the previous step.





7.1.2 Elimination of dispersed population overlay (duplication) with GUF pixels

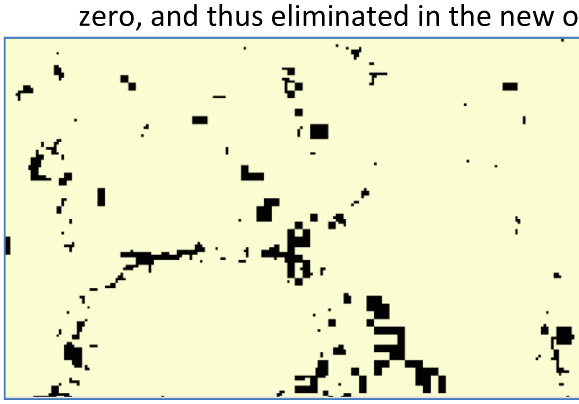
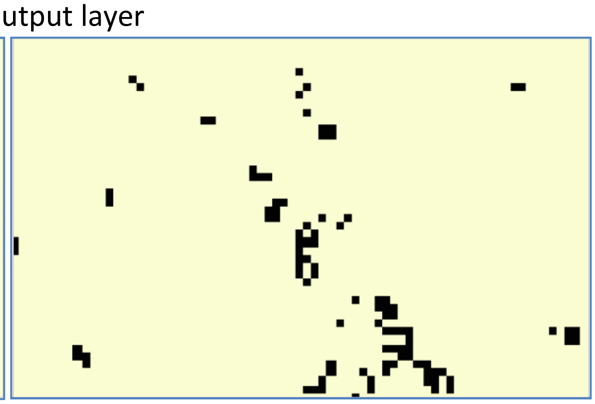
In this step we eliminate pixels where GUF may have incidentally overlapped with the land cover classes selected in 7.1.1 (these 2 raster files are independent datasets from different sources that have been resampled to the 100x100 m. grid and its possible that some of the GUF values overlap with some of the pixels in THA_CALC/AgriPlus_THA). These pixels must be eliminated to avoid double counting.

Subtract the raster data **THA_Input/GUFDLR_THA100m_clip.tif** **from** **THA_CALC/AgriPlus_THA.tif** **,** keeping only values >0.

Use QGIS/Toolbox/SAGA/Raster calaculus/ Raster calculator

Main input layer: **AgriPlus_THA.tif** (is "a" in the formula)
Additional layer: **THA_Input/GUFDLR_THA100m_clip.tif** (is "b" in the formula)
Formula: ifelse(eq(a-b,0),0,a)
(it reads: if a-b=0, keep 0, if not keep the a value which is 0 or 1)
Output: THA_CALC/AgriPlusNoGUF_THA.tif

The effect of this arithmetic for the output layer is that overlapping values would be less than zero, and thus eliminated in the new output layer

AgriPlus_THA.tif **AgriPlusNoGUF_THA.tif**

7.2 Estimation of population in Non-GUF pixels

Multiply the new file **AgriPlusNoGUF_THA.tif** **by 0.3** (default mean density)

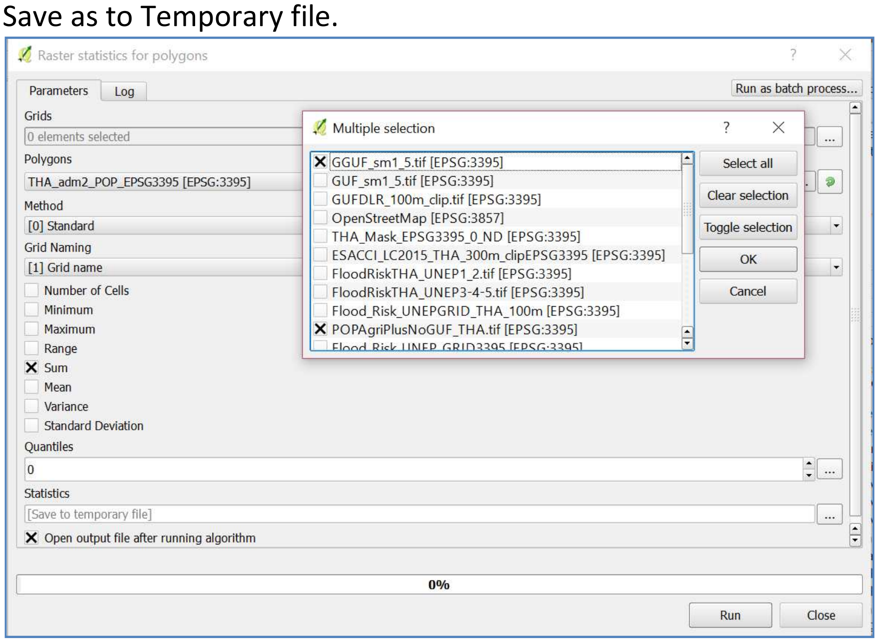
Use QGIS or SAGA Raster calculator
Formula: AgriPlusNoGUF_THA.tif *0.3 (in SAGA: a*0.3)
Output: THA_CALC/POPAgriPlusNoGUF_THA.tif

NOTE: in the absence of additional information, at this time we use the same assumption for 2000 and 2010, i.e. 0.3 density of dispersed population in the selected non-GUF areas.

7.3 Extract values from pixels to Administrative Regions (e.g. districts)

Now we can link back with the population census statistics joined to the administrative regions (THA_adm2_POP_EPSG3395.shp) and use this data to assign portions of the actual population (according to the census) to our GGUF and non-GUF dispersed population areas identified in the previous steps, according to the assigned density value (.3 per hectare). First, we sum across the grid for the **GGUF_sm** values and for **POPAgriPlusNoGUF** population (as calculated in 7.2) by polygon (administrative region).

Use QGIS/Toolbox/SAGA/Vector<>Raster/Raster statistics for Polygons
Grids: Multiple selections / **GGUF_sm1_5** and **POPAgriPlusNoGUF_THA**.
Polygons: THA_adm2_POP_EPSG3395
Method: [0] Standard
And then, uncheck all output options, **CHECK ONLY SUM**
Save as to Temporary file.



In order to keep the same file name, check that it's a temporary file. Once the program is run and the Temporary fie is checked OK, remove **THA_adm2_POP_EPSG3395**. Then Save the Temporary file with the same name **THA_adm2_POP_EPSG3395** (overwrite). (This is because QGIS does not allow for direct overwriting of files by using the same file name).

The new file includes the values for the non-GUF areas for each administrative region (district). Calculate mean population density by administrative division and GGUF pixels points

In step 6, we have assigned a weight to GUF pixels, in the range of 0.01 to 1.00. This weight is a way of adjusting population density in GUF pixels in order to take into account the agglomeration factor.

In step 7, we have calculated in THA_adm2_POP_EPSG3395.shp the SUM of GGUF pixels points by ADM2 divisions. The field is named **GGUF_sm**.

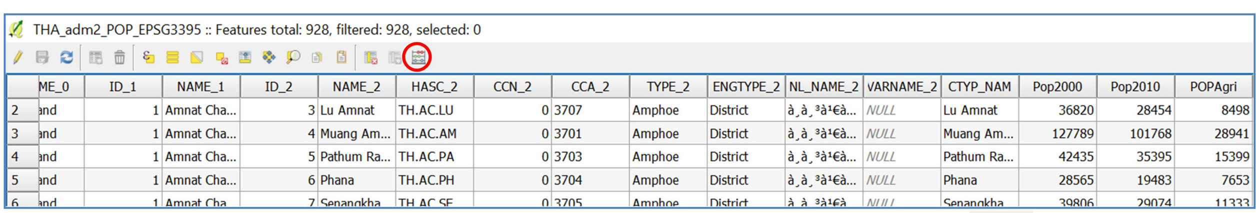
In step 5.2, we introduced population data in the attributes table of **THA_adm2_POP_EPSG3395.shp**. They are the population census data 2000 and 2010 (fields: resp. Pop2000 and Pop2010) and our estimation of population living out of GUF pixels carried out in **POPAgriPlusNoGUF_THA.tif** (field: **POPAgri**).

The purpose of step 8 is to calculate for each administrative division the population average value for GUF weighted pixels. The formula is **[Population in GUF pixels by ADM2]/ [SUM of GGUF pixels points].**

7.4 Subtract estimated Non-GUF population from

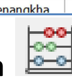
Prior to making this calculation, we firstly need to subtract from the total population in each administrative region to population living outside of GUF – the dispersed population labelled **POPAgri**. We name these new fields **PGUF00v0** and **PGUF10v0.**

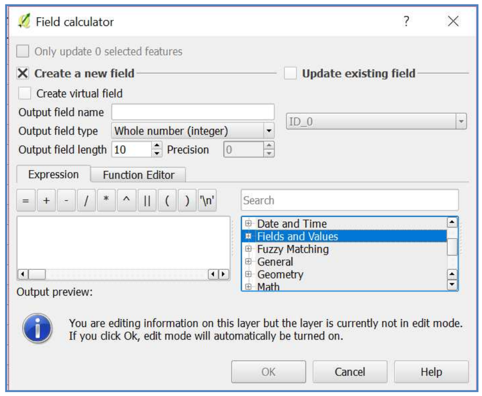
Open THA_adm2_POP_EPSG3395.shp **Attributes Table** by clicking  on the top window pane.



THA_adm2_POP_EPSG3395 :: Features total: 928, filtered: 928, selected: 0

	ME_0	ID_1	NAME_1	ID_2	NAME_2	HASC_2	CCN_2	CCA_2	TYPE_2	ENGTYPE_2	NL_NAME_2	VARNAME_2	CTYP_NAM	Pop2000	Pop2010	POPAgri
2	and	1	Amnat Cha...	3	Lu Amnat	TH.AC.LU	0	3707	Amphoe	District	à¸à¸³à¹€à...	NULL	Lu Amnat	36820	28454	8498
3	and	1	Amnat Cha...	4	Muang Am...	TH.AC.AM	0	3701	Amphoe	District	à¸à¸³à¹€à...	NULL	Muang Am...	127789	101768	28941
4	and	1	Amnat Cha...	5	Pathum Ra...	TH.AC.PA	0	3703	Amphoe	District	à¸à¸³à¹€à...	NULL	Pathum Ra...	42435	35395	15399
5	and	1	Amnat Cha...	6	Phana	TH.AC.PH	0	3704	Amphoe	District	à¸à¸³à¹€à...	NULL	Phana	28565	19483	7653
6	and	1	Amnat Cha	7	Senangkha	TH.AC.SE	0	3705	Amphoe	District	à¸à¸³à¹€à	NULL	Senangkha	39806	29074	11333

Open the Field Calculator by clicking on the **abacus** (counting frame) **icon**  (circled in red above).

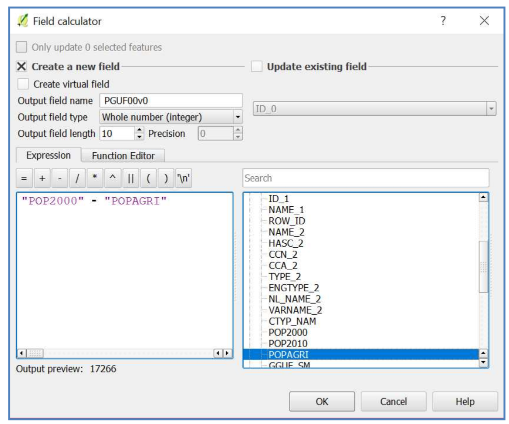


Population in GUF 2000:

Output field name: **PGUF00v0**

In the bottom right box, click on **Fields and Values** to open the fields list.

Double click on **Pop2000**



Click on the “Expression” “minus” icon

Double click on **POPAgri**

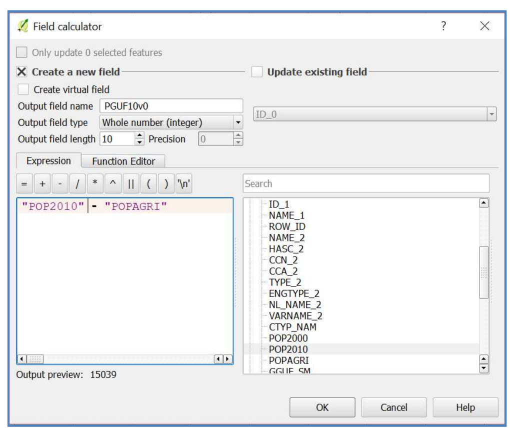
Click OK

Population in GUF 2010:

Output field name: **PGUF10v0**

click on **Fields and Values**

Double click on **Pop2010**



Click on the “Expression” “minus” icon

Double click on **POPAgri**

Click OK

7.5 Check results and Fix anomalies

At this point, you may find anomalies - shown as negative values, which are irrational. In the THA test this happended for 26 cases out of 928 divisions. It shows that the formula used for estimating non-GUF population does not work in all cases. A probable bias may come from landscape patterns with very large agriculture and grassland areas, which leads to irrational results when using a fixed coefficient. It could also mean that the default coefficient chosen (0.3 inhab. per ha.) is to high, in general or in some places.

Ideally, the estimation should be improved on the basis of other sources of information on rural population. In absence of such information for this test, we will modify the previous formula used for estimation by introducing the condition that when AgriPlus population estimation is obvioulsly overestimated, we will replace it by another formula. For the example, when estimated **AgriPlus** is more than 34% of the population census statistics for an administrtative area, we force instead a new coefficient for estimation of Total Population in non-GUF area = 66% of census statistics. [other similar formulas can also be tested…]

With the double condition formula, in ADM2 divisions with high population density, estimations depend from agriculture/pasture area while in other divisions with low population density and large agriculture/pasture area, the estimation depends on statistics only.

This calculation cannot be done in the QGIS vector calculator, but it can done using a spreadsheet such as MSExcel, OpenOffice or LibreOffice (or other…). MSExcel opens the attribute table as a .csv file. OpenOffice or LibreOffice open the attribute table as a .dbf file

In either case, use the IF function and parameterise it as such:

- For 2000: **Name field** (column) **POPGUF00**
  If we use OpenOffice or LibreOffice (.dbf format), add to the name the caracteristics of the field which are ,N,19,0 → **POPGUF00,N,19,0**

  Formula (use Function IF): =IF(R2>P2*0.34,P2*0.66,P2-R2)
  Where R2 is **POPAGRI**
  P2 is **POP2000**
- For 2010: **Name field** (column) **POPGUF10**
  If we use OpenOffice or LibreOffice (.dbf format), add to the name the caracteristics of the field which are ,N,19,0 → **POPGUF10,N,19,0**

  Formula (use Function IF): =IF(R2>Q2*0.34,Q2*0.66,Q2-R2)
  Where R2 is **POPAGRI**
  Q2 is **POP2000**

The output of the IF-function formulats in the spreadshet will is as follows: t if R2 is > than P2*0.33 (a large value), then **POPGUF**= P2*0.66, if not, take P2-R2 for **POPGUF**. Again, the purpose of this additional calculation is simply to force an alternative coefficient for cases (by district) where the estimated non-GUF populations are irattionaly large.

Save the results:

- **With MSExcel: Save as .csv** file (accept the warning).

- **With OpenOffice or LibreOffice: Save as .dbf** file.

Open **THA_adm2_POP_EPSG3395.shp in QGIS**

- If we used **OpenOffice or LibreOffice, nothing more has to be done.**
- If we used **MSExcel we have still to join the results of the .csv table to the shapefile.**

Upload THA_adm2_POP_EPSG3395.csv with the “Add vector layer” button 

Go to THA_adm2_POP_EPSG3395.shp and double-click Properties and the Joins**.** Follow the procedure described in **Step 6.3. Keep only** the POPGUF00 and POPGUF10 fields.

7.6 Calculate population density per GGUF point

Open THA_adm2_POP_EPSG3395 Attributes Table and use QGIS **Vector Calculator**.

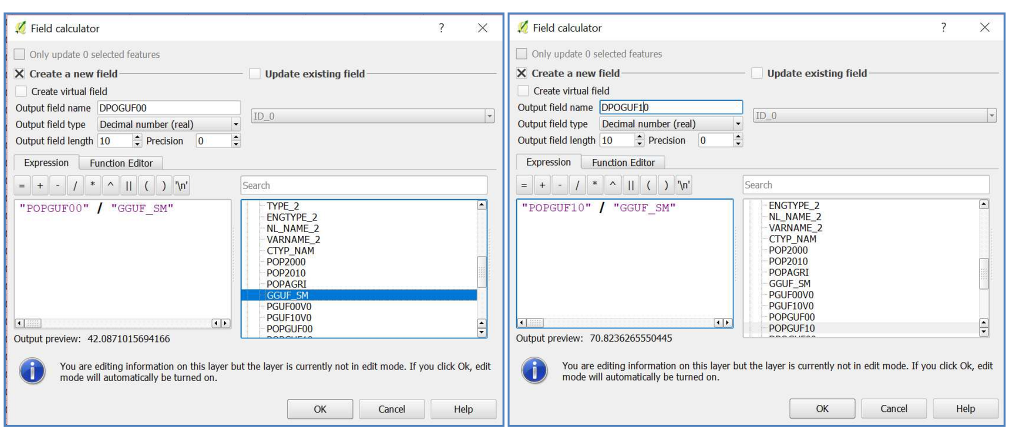
Note: the default number format in the box is “Integer” (no decimals). It is better having values with decimals, so we have to declare “**Decimal number (real)**”

Formula: **POPGUF00** / **GGUF_SM** and Output field name: **DPGGUF00**

Recall that POPGUF00 is population by adminsitrative region excluding the estimated non-GUF (dispersed) population.

Notation note: DPGGUF stands for Density of Population per GGUF point

and **POPGUF00 / GGUF_SM** and Output field name: **DPGGUF10**



Save the results **.**

8 Produce final gridded population density maps

This final part of the gridded population estimation is accomplished in two steps:

1/ rasterizing **THA_adm2_POP_EPSG3395.shp** fields: GGUF points' population density and
2/ multiplying the raster result by **GGUFsm1_5.tif** (which are our GGUF coefficients)

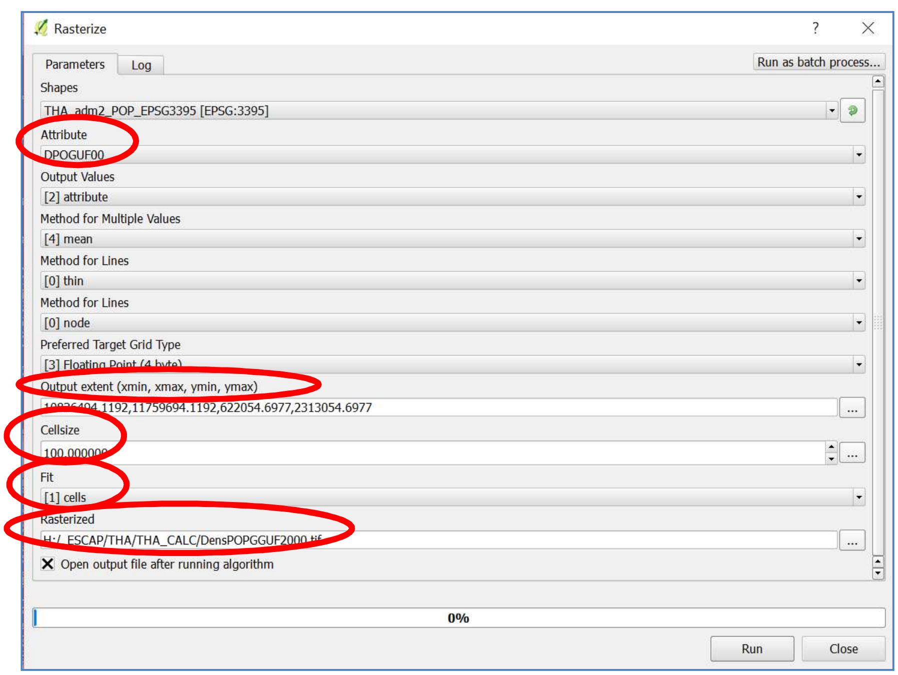
8.1 Rasterisation of GGUF points' population density

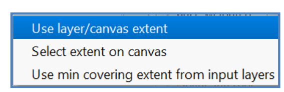
Use QGIS/Toolbox/SAGA/Raster creation/Rasterize

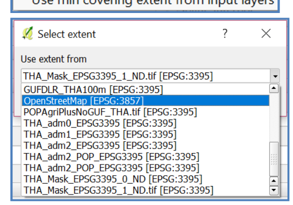
Shapes: THA_adm2_POP_EPSG3395.shp
Attribute: Firstly DPOGUF00 and then DPOGUF10
Leave default values except:

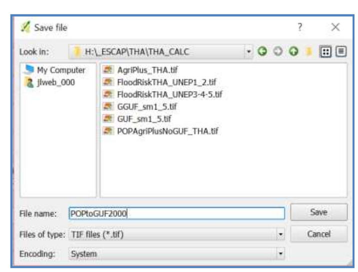
> **Output extent:** click [...] and **Use layer/canvas extent** and then **Select extent THA_Mask_EPSG3395_1_ND.tif**
> **Cellsize:** confirmat value is **100.0000**
> **Fit**: change to **[1] cells**

Rasetrized: click [...], go to the THA_CALC folder and give the name **DensPOPGGUF2000** (and then, **DensPOPGGUF2010** for the second file)









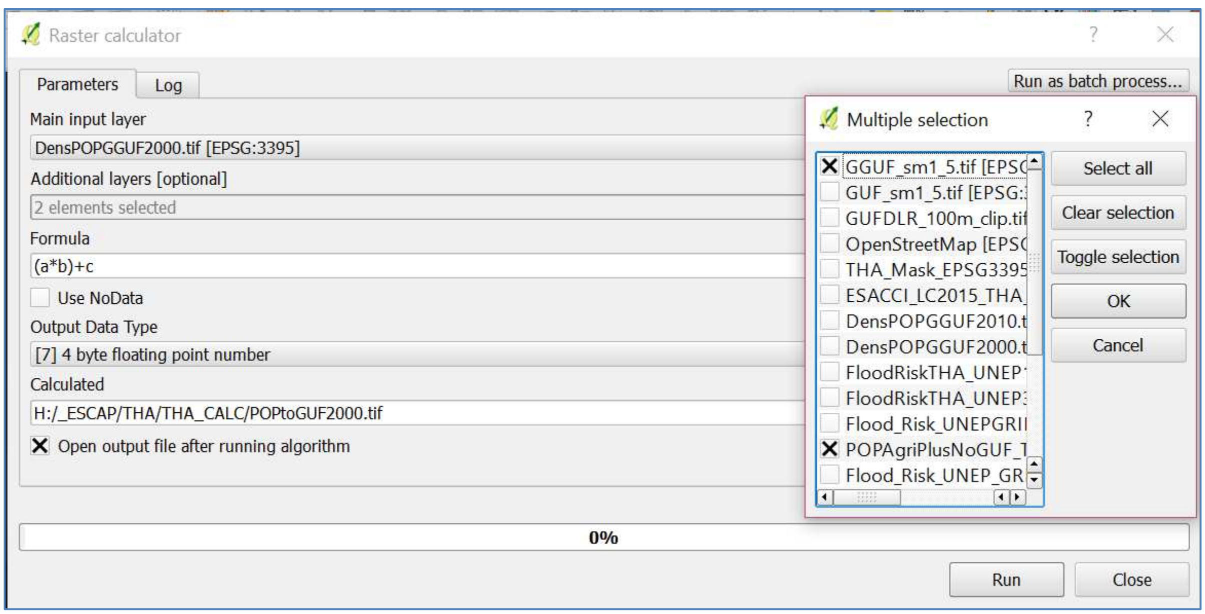
8.2 Calculation of final POPtoGUF Rasterized Results

Multiply the raster layers obtained in the previous step by **GGUFsm1_5.tif** and add **POPAgriPlusNoGUF** for the final result.

Use **QGIS/Toolbox/SAGA/Raster calculus/Raster calculator**

For POPtoGUF2000

Main input layer**: DensPOPGGUF2000** (it will be “a” in the formula)
Additional layers**: GGUF_sm1_5** (it will be “b”)
POPAgriPlusNoGUF (it will be “c”)
OK
Formula: (a*b)+c
Calculated: click ..., go to the THA_CALC folder and type **POPtoGUF2000**



For POPtoGUF2010

Main input layer**: DensPOPGGUF2010** (it will be “a” in the formula)
Additional layers**: GGUF_sm1_5** (it will be “b”)
POPAgriPlusNoGUF (it will be “c”)
OK
Formula: (a*b)+c
Calculated: click ..., go to the THA_CALC folder and type **POPtoGUF2010**

These are the final results for the Pop-to-GUF estimation, at this point in gridded (raster) format.

9 Estimation of population exposure to flood hazard areas

Now, in the final step, we integrate our population density estimations with hazard areas to calculate the population exposure to hazard and reintegrate with the administrative region shapefile to produce the final statistics and present the results.

9.1 Creation of two risk intensity levels: Low risk and High risk

Recall that the UNEP GRID raster data set has been pre-processed in previous steps (resampled and clipped to THA extent - recall steps Step 2.3.2 and 5.4) to make it compatible with other layers. The file that we will use now is **FloodRiskTHA_UNEPGRID_100m_3395.tif.**

As an additional processing step for analysis, we also now convert the data from 5 classes to 2 classes with the sole purpose of simplifying the analysis:

5 classes of risk intensity as distinguished in the UNEP GRID file, from min (1) to max (5). For simplicity, we will group them into 2 classes only: Low (1,2) and High (3,4,5).
Use QGIS/Toolbox/SGA/Raster calculus/ Raster calculator
Main input layer: **FloodRiskTHA_UNEPGRID_100m_3395.tif**
Formula: ifelse(lt(a,3),1,0)
Rasterised: FloodRiskTHA_UNEP1_2.tif

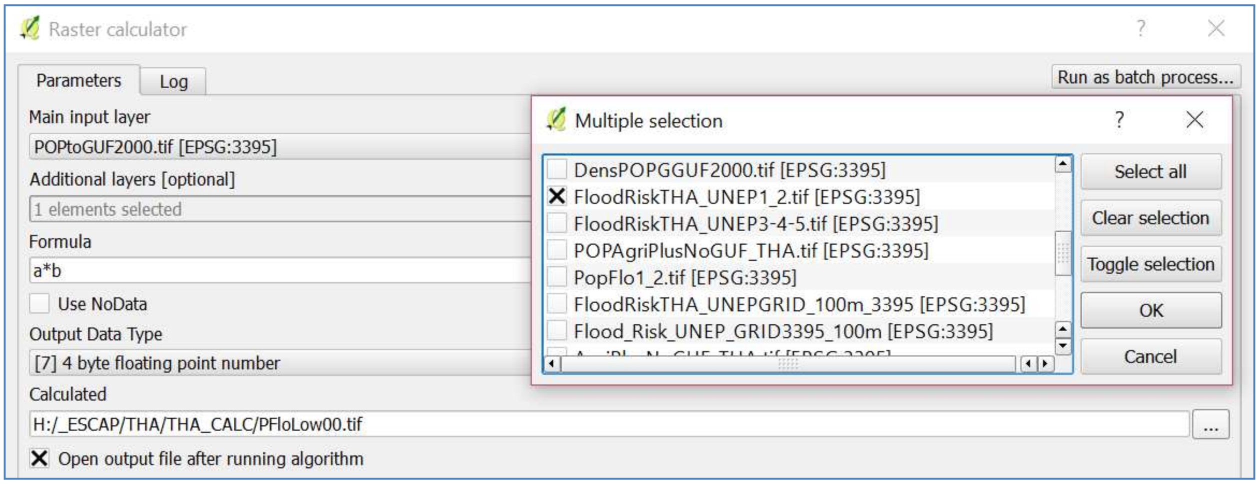
Use QGIS/Toolbox/SGA/Raster calculus/ Raster calculator
Main input layer: **FloodRiskTHA_UNEPGRID_100m_3395.tif**
Formula: ifelse(gt(a,3),1,0)
Rasterised: FloodRiskTHA_UNEP3_4_5.tif

9.2 Calculation of population in Low risk and High risk exposure areas

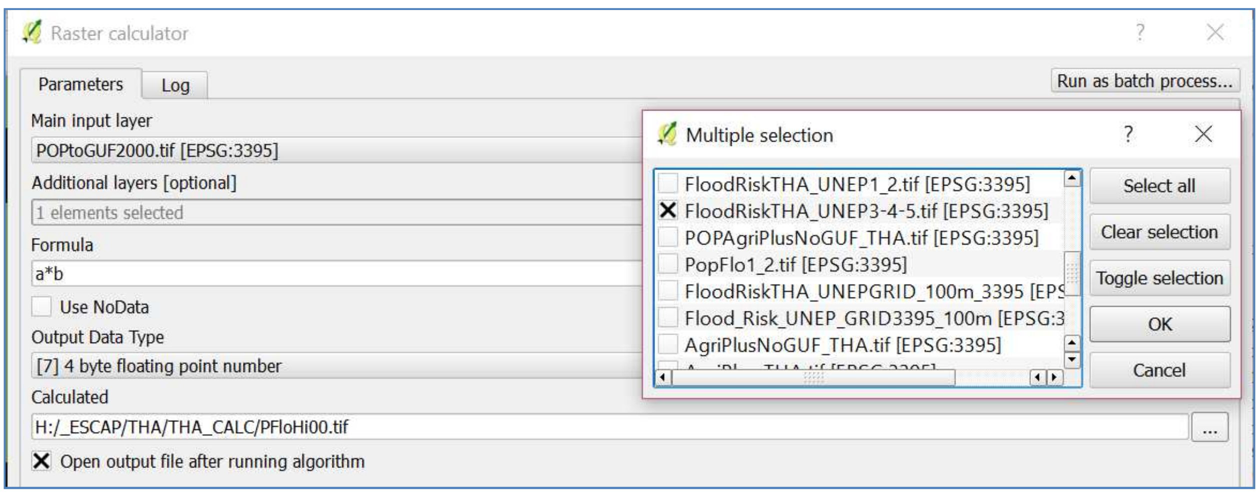
Use QGIS/Toolbox/SAGA/Raster calculus/ Raster calculator

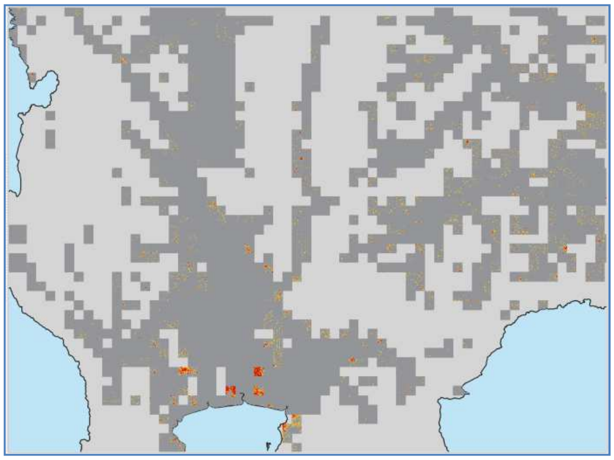
<u>For 2000</u>

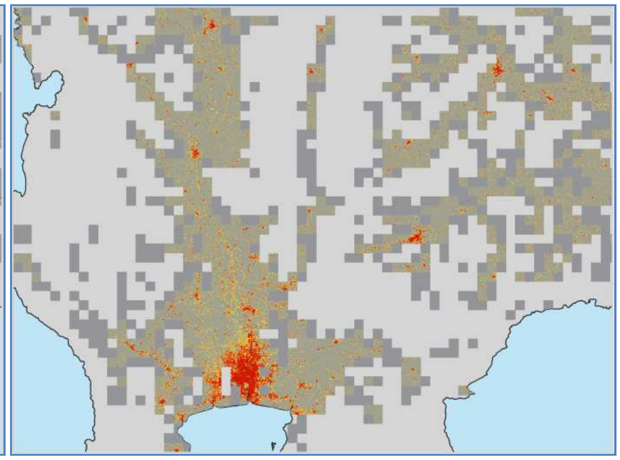
- Low risk (1_2)
  - **Main input layer: POPtoGUF2000.tif** (for 2000) or **POPtoGUF2010.tif** (for 2010)
  - **Additional layer : FloodRiskTHA_UNEP1_2.tif**
  - **Formula: a*b**
  - **Calculated outcome: PFloLow00.tif** or **PFloLow10.tif**



- High risk (3, 4, 5)
  - **Main input layer: POPtoGUF2000.tif** or **POPtoGUF2010.tif**
  - **Additional layer: FloodRiskTHA_UNEP3-4-5.tif**
  - **Formula: a*b**
  - **Calculated outcome: PFloHi00.tif** or **PFloLow10.tif**



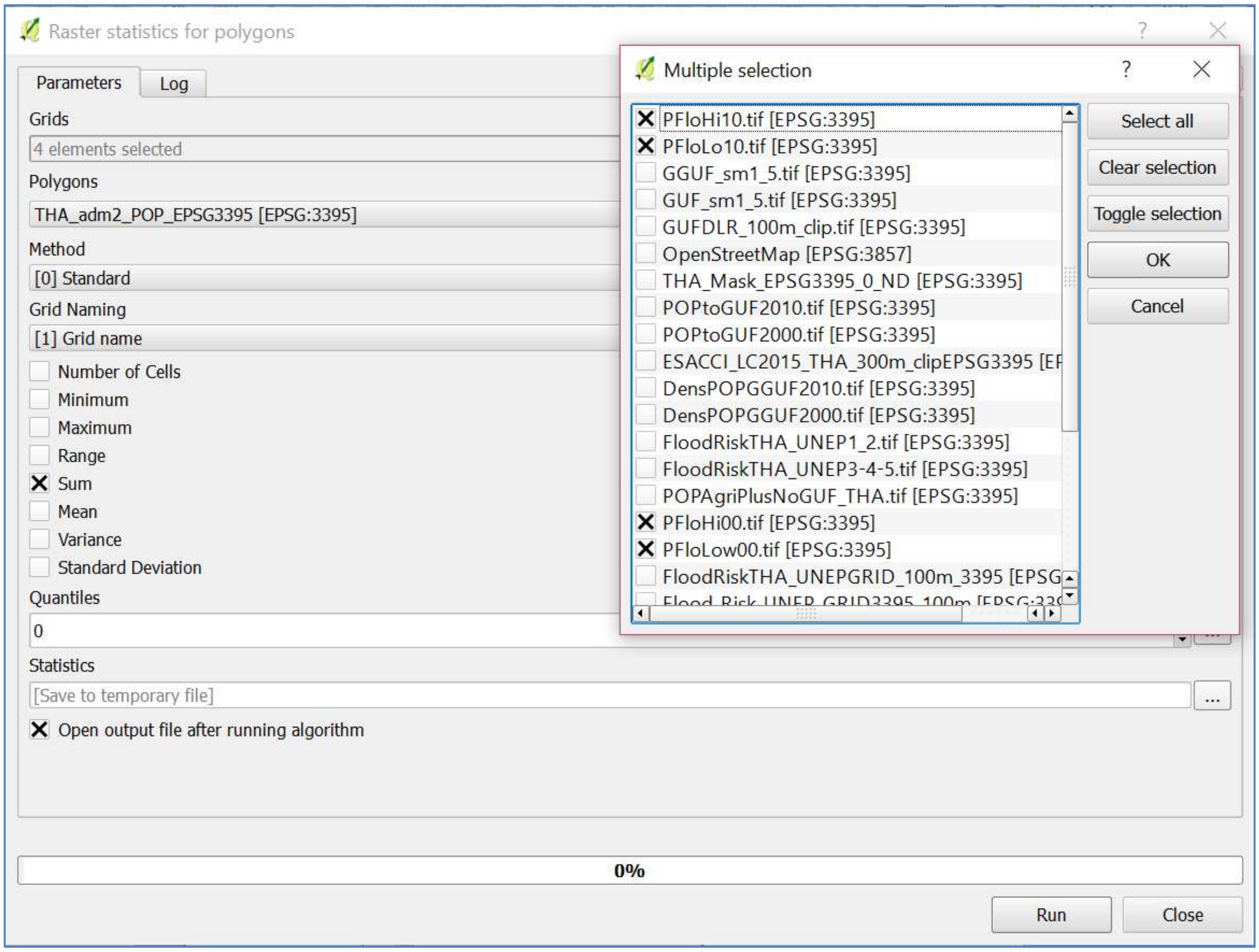

Population 2010 living in areas with **low risk of flood.**


Population 2010 living in areas with **high risk of flood.**

9.3 Extraction of statistics

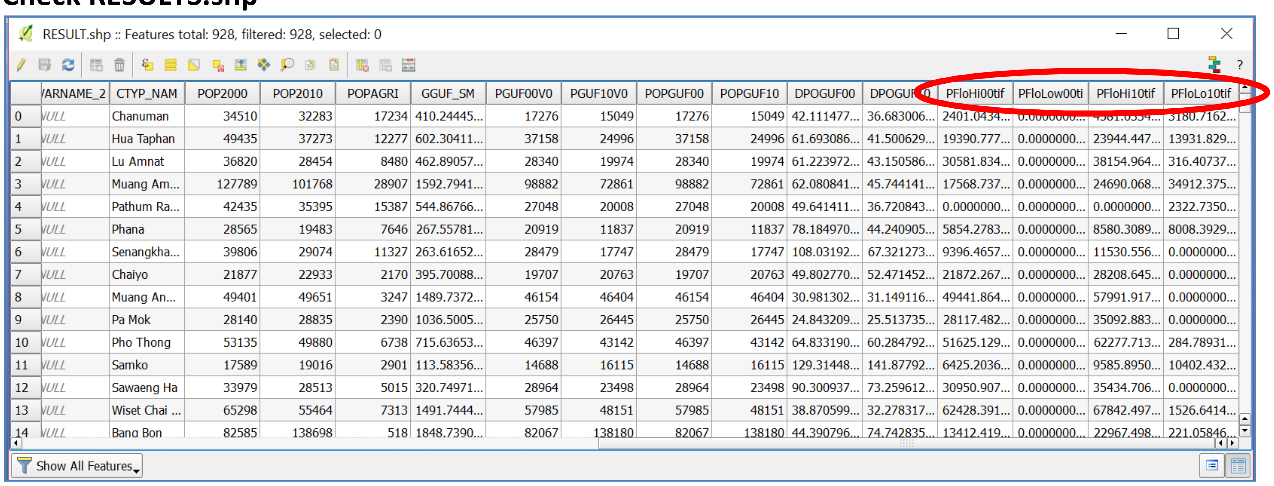
The outputs from the previous step are still in raster file format. To produce statistics, e.g. summaries by administrative regions, we must aggregate the information back to groupings by regions (polygons or vector file). We are interested only in the Sum (i.e. total estimated number of individuals within the criteria).

Use **QGIS/Toolbox/SAGA/Vector<>Raster/Raster statistics for polygons**
Grids**: select 4 elements: PFloLow00.tif, PFloHi00.tif, PFloLow10.tif and PFloHi10.tif**
Polygons**: THA_adm2_POP_EPSG3395**
Method**: [0] Standard**
Uncheck all options, **keep only SUM**
Statistics**: [Save to temporary file]**
Run



If any problems, process raster layers 1 by 1.

Check RESULTS.shp



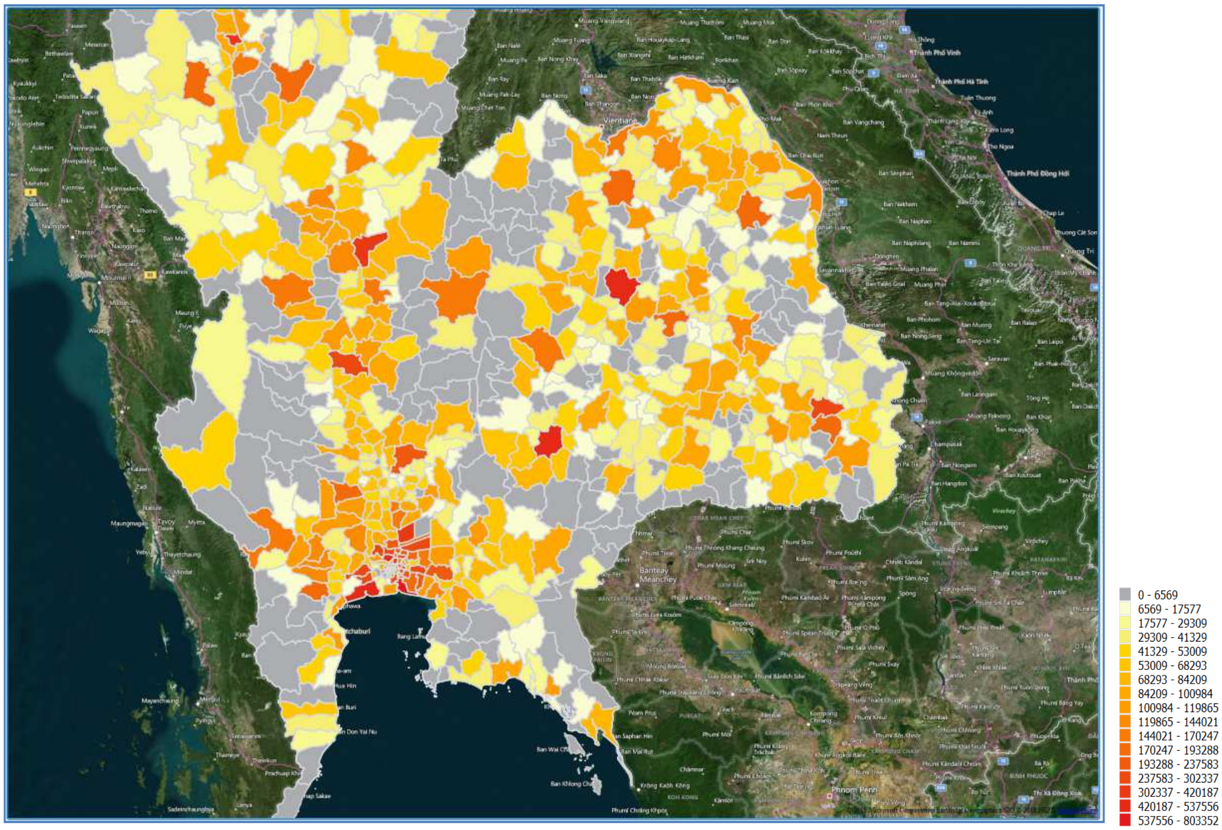
	VARNAME_2	CTYP_NAM	POP2000	POP2010	POPAGRI	GGUF_SM	PGUF00V0	PGUF10V0	POPGUF00	POPGUF10	DPOGUF00	DPOGUF10	PFloHi00tif	PFloLow00ti	PFloHi10tif	PFloLo10tif
0	NULL	Chanuman	34510	32283	17234	410.24445...	17276	15049	17276	15049	42.111477...	36.683006...	2401.0434...	[illegible]	[illegible]	3180.7162...
1	NULL	Hua Taphan	49435	37273	12277	602.30411...	37158	24996	37158	24996	61.693086...	41.500629...	19390.777...	0.0000000...	23944.447...	13931.829...
2	NULL	Lu Amnat	36820	28454	8480	462.89057...	28340	19974	28340	19974	61.223972...	43.150586...	30581.834...	0.0000000...	38154.964...	316.40737...
3	NULL	Muang Am...	127789	101768	28907	1592.7941...	98882	72861	98882	72861	62.080841...	45.744141...	17568.737...	0.0000000...	24690.068...	34912.375...
4	NULL	Pathum Ra...	42435	35395	15387	544.86766...	27048	20008	27048	20008	49.641411...	36.720843...	0.0000000...	0.0000000...	0.0000000...	2322.7350...
5	NULL	Phana	28565	19483	7646	267.55781...	20919	11837	20919	11837	78.184970...	44.240905...	5854.2783...	0.0000000...	8580.3089...	8008.3929...
6	NULL	Senangkha...	39806	29074	11327	263.61652...	28479	17747	28479	17747	108.03192...	67.321273...	9396.4657...	0.0000000...	11530.556...	0.0000000...
7	NULL	Chaiyo	21877	22933	2170	395.70088...	19707	20763	19707	20763	49.802770...	52.471452...	21872.267...	0.0000000...	28208.645...	0.0000000...
8	NULL	Muang An...	49401	49651	3247	1489.7372...	46154	46404	46154	46404	30.981302...	31.149116...	49441.864...	0.0000000...	57991.917...	0.0000000...
9	NULL	Pa Mok	28140	28835	2390	1036.5005...	25750	26445	25750	26445	24.843209...	25.513735...	28117.482...	0.0000000...	35092.883...	0.0000000...
10	NULL	Pho Thong	53135	49880	6738	715.63653...	46397	43142	46397	43142	64.833190...	60.284792...	51625.129...	0.0000000...	62277.713...	284.78931...
11	NULL	Samko	17589	19016	2901	113.58356...	14688	16115	14688	16115	129.31448...	141.87792...	6425.2036...	0.0000000...	9585.8950...	10402.432...
12	NULL	Sawaeng Ha	33979	28513	5015	320.74971...	28964	23498	28964	23498	90.300937...	73.259612...	30950.907...	0.0000000...	35434.706...	0.0000000...
13	NULL	Wiset Chai ...	65298	55464	7313	1491.7444...	57985	48151	57985	48151	38.870599...	32.278317...	62428.391...	0.0000000...	67842.497...	1526.6414...
14	NULL	Bang Bon	82585	138698	518	1848.7390...	82067	138180	82067	138180	44.390796...	74.742835...	13412.419...	0.0000000...	22967.498...	221.05846...

Remove **THA_adm2_POP_EPSG3395.shp** from QGIS

Save **RESULTS.shp** as **THA_adm2_POP_EPSG3395.shp** and open the file.

Remove RESULTS.shp

9.4 Presentation of results



Population 2010 exposed to high risk of flood, by Districts (ADM2)