

WORKFLOW DATA PREPARATION

IGARSS 2022

COLLECTING DATA

SURFACE TEMPERATURE

The surface temperature is measured by Landsat 8 and can be downloaded using this link (<http://landsatlst.appspot.com/>).

It is important that the satellite image meets following conditions:

- Minimal cloud cover since clouds will result in missing data.
- High surface temperatures are measured.
- The image must be collected during the time of the year when the vegetation is most apparent.

On this site following inputs are expected:

- Period of interest: This period must be less than 6 months. This period is chosen in spring and summer so that the vegetation is as clear as possible. In doing so, care should be taken that the vegetation is as well developed as possible, without already being less visible due to drought and/or heat stress. To determine the optimal period, the Sentinel-2 images can be used.
- Area of interest: Designate the area around the city. This area should be chosen as small as possible to allow for downloading the images, but these boundaries should not be too close to the city limits since the census data may differ from the city limits.
- Satellite: Choose Landsat8.
- Emissivity: Choose ASTER for the best estimation of the surface temperature. If the ASTER emissivity results in large areas of missing data in the satellite image, independently of the , NDVI-based emissivity can be used.

SENTINEL-2

Sentinel images can be downloaded using Google Earth Engine. The scripts to access the data can be found in this repository. Consulting the Sentinel-2 images consists of two steps. First, the images in the area of interest are queried for a certain period of time. This period is usually the summer months in the year for which a suitable LST image was found. The Sentinel image and the Landsat image should follow each other as closely as possible. Here it is also important to check that all vegetation in the city is clearly visible on the Sentinel image and that there are as few clouds present as possible. Based on these criteria, different options for Sentinel and Landsat images are determined. Finally, from these possibilities, the most suitable images are chosen. This phase is performed in the script 'Exploration_Demo'.

Once the satellite images are chosen, they are downloaded using the script 'Download_Demo'. Both scripts contain detailed instructions.

CENSUS DATA

The census data can be found at the following link:

<https://data.census.gov/cedsci/>

This involves entering the city as the search query and selecting the census tracts in the broad vicinity of the city by filter. The files which are downloaded in the demo can be found in the folder '*Data preparation\Raw data\United States\Social_data_Atlanta*'. The shapefiles associated with this data can be downloaded via the following link:

<https://www.census.gov/cgi-bin/geo/shapefiles/index.php>

Here both the census tracts and the counties can be downloaded. Both shapefiles are needed since the census tract IDs overlap over several counties.

DATA PREPARATION

Once you have downloaded all the data, it must be preprocessed. This is always done according to the following steps:

THE CENSUS DATA:

- Reading the different datasets into Python or another data processor.
- Adjusting the variables to a usable format.
- Merge the datasets.
- Merge with geographic data.
- Final adjustment of variables and column names.
- Export data to shapefile.

Table 1: variables used in the workshop for Atlanta.

Variable	Description	year
tract	Identifier of the census tract	2021
county	County in which the census tract is located	2021
area	Area of the census tract (km ²)	2021
tot_pop	Population of the census tract	2021
pop_dens	Population density of the census tract	2021
med_inc	Median income of the population of the census tract (\$)	2021
unempl	Ratio of population of the census tract that is unemployed (%)	2021
under_5_y	Ratio of population of the census tract that is less than 5 years old (%)	2021
under_10_y	Ratio of population of the census tract that is less than 10 years old (%)	2021
under_15_y	Ratio of population of the census tract that is less than 15 years old (%)	2021
over_65_y	Ratio of population of the census tract that is more than 65 years old (%)	2021
over_70_y	Ratio of population of the census tract that is more than 70 years old (%)	2021
over_80_y	Ratio of population of the census tract that is more than 80 years old (%)	2021
alone	Ratio of population of the census tract that lives alone (%)	2021
no_dipl	Ratio of population of the census tract that has no high school diploma (%)	2021

Variable	Description	year
high_sch	Ratio of population of the census tract that has only a high school diploma (%)	2021
degree	Ratio of population of the census tract that has more than a high school diploma (%)	2021
renter	Ratio of population of the census tract that rents a house or apartment (%)	2021
pre_1960	Ratio of the dwellings in the census tract that is build before 1960 (%)	2021
pre_1980	Ratio of the dwellings in the census tract that is build before 1980 (%)	2021
after_2014	Ratio of the dwellings in the census tract that is build after 2014 (%)	2021
apartments	Ratio of the dwellings in the census tract that is an apartment (%)	2021
attached	Ratio of the dwellings in the census tract that is an attached house (%)	2021
detached	Ratio of the dwellings in the census tract that is a (semi-) detached house (%)	2021

SATELLITE DATA AND CROPPING

- Open files in QGIS.
- Set as project CRS the CRS of the census file.
- Find a shapefile with the city boundaries online and load it.
- Merge the LST images, if necessary.
- Open the 'Preprocessing satellite data' model via the processing toolbox. The file location of this model is: 'Data preparation Satellite images Preprocessing satellite data.model3'. The structure of this model is shown schematically in Figure 1.
- If this model does not work on your version of QGIS, all steps must be performed by yourself:
 - Reproject all files to the project CRS.
 - Select the Census tracts that have sufficient overlap with the city limits.
 - Clip the satellite images based on the selected census tracts.

Note: A frequent problem with QGIS models is saving the files. However, this can be solved by running it not as Batch processes but as a single process and by opening the outputs as temporary layers and exporting them manually.

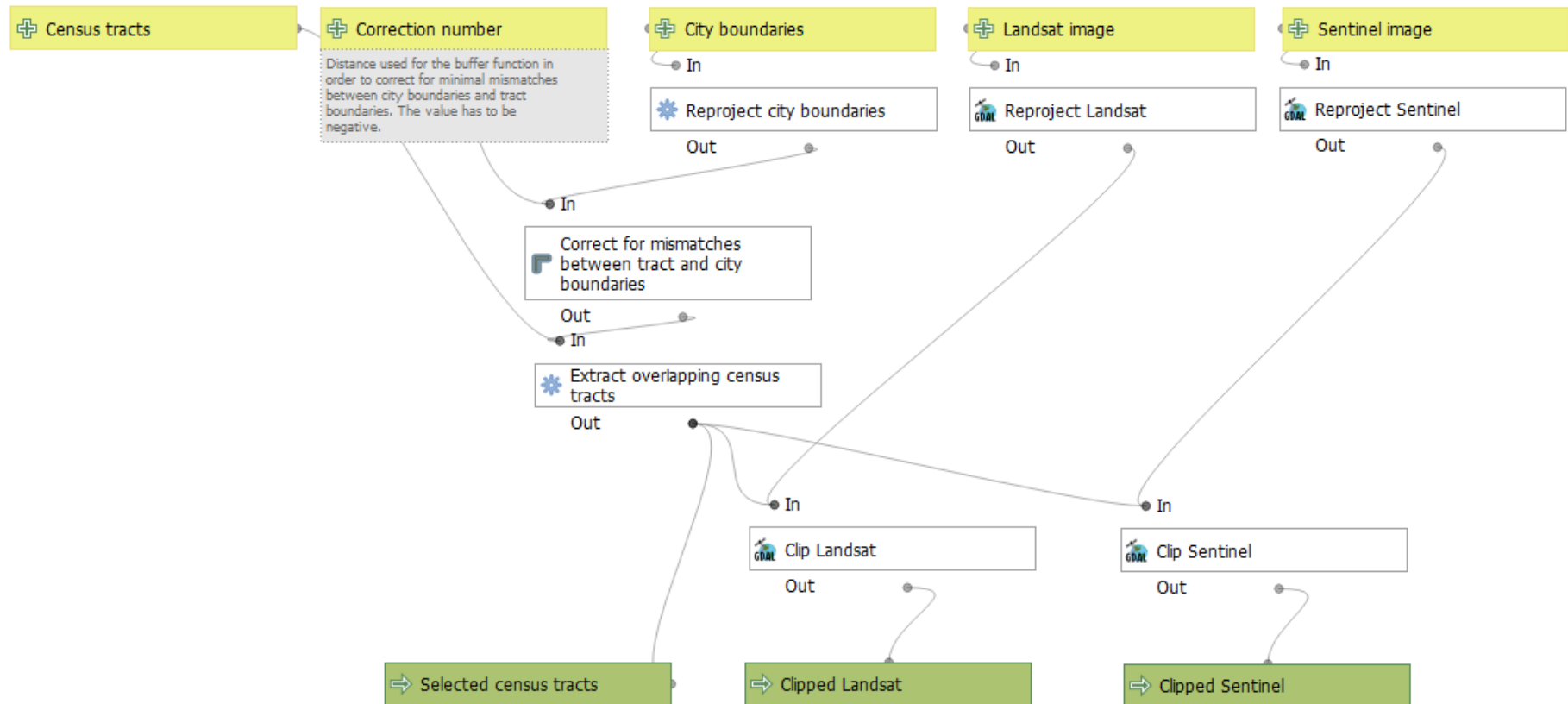


Figure 1: Schematic representation of the preprocessing steps of the satellite images and selecting the census tracts, using the model 'Preprocessing satellite Data'.