Data Configuration Guide for OnSSET

Preparing the population data

OnSSET requires a table that specifies a list of records of human population clusters in the area of interest. The population clusters are polygons in which the human population density is assumed to be constant.

Each record of the table describes the coordinates of the centroid (center point) of the polygon and the total population within the polygon.

Population clusters are available to download from these websites and can be used to skip this step and start working on preparing the input data for the model.

Population clusters ready to download:

https://data.mendeley.com/datasets/z9zfhzk8cr/6

https://energydata.info/ and https://electrifynow.energydata.info/

Alternatively, we can create population clusters by using the population raster data and the clustering plug in for QGis installed during the previous step. Please note that there also alternative methods that can be used such as this Python Jupyter Notebook https://github.com/babakkhavari/Clustering and the code developed for ArcGIS Pro.

Before creating the population clusters we need to define the boundaries of our areas of interest, coordinate reference system (CRS) and download the raster population data and the night light raster (required by the QGis population cluster plug in).

Population raster data sources:

https://data.humdata.org/dataset/highresolutionpopulationdensitymaps

https://data.humdata.org

https://landscan.ornl.gov/

Administrative boundaries:

https://gadm.org/download_country.html

https://data.humdata.org/

Coordinate reference system

https://epsg.io/

Select a projected coordinate system that is valid for your whole area of interest.

Nightime lights

https://eogdata.mines.edu/products/vnl/

The annual VNL 2 product is currently the latest version. The average global annual nighttime lights raster for 2023 (9.1 GB) can be downloaded from:

https://eogdata.mines.edu/nighttime_light/annual/v22/2023/VNL_npp_2023_global_vcmslcfg_v2_c202402081600.average.dat.tif.gz

Before using the plug in or the ArcGIS Pro code, we eventually need to clip the population and night lights raster by using the selected boundaries. This operation can be performed by the "Raster Extraction" processing tool in QGis or Extract by Mask tool if ArcGIS Pro with Spatial Analyst is available.

In QGis we can load the boundaries, clipped population raster and clipped night lights raster and then click on the HRKS_clustering plugin button in the toolbars.

The interface requires to indicate the Workspace (a temporary working folder) the country code (https://en.wikipedia.org/wiki/ISO_3166-2) , country name, the projected CRS, the input layers and threshold for you minimum population in each cell.

Detailed instructions are described here:

https://github.com/OnSSET/PopCluster/blob/master/Instructions/How%20to%20run%20the%20clustering%20plugin.docx

ArcGIS Pro users can use the code available as Python Notebook. To run the code you need the select the Analysis meny, Python button and select Python notebook. Before running the code you need to specify the parameters in the parameter section.

Please note that running time could take from few minutes to 2-3 hours according to the hardware and software configuration, size and spatial resolution of the input data and threshold parameter.

Preparing the Input Data

OnSSET requires several **spatial data layers** to conduct electrification modeling. These layers are typically provided as **shapefiles** or **rasters**, and they represent key factors such as population distribution, energy demand, infrastructure, and renewable energy potentials. The following data layers are required:

1. **Population Density & Distribution**: This raster dataset provides spatial identification and quantification of the population within the area. It is crucial for determining the energy demand and electrification priorities. You can find this data from the WorldPop dataset or national census data.
2. **Administrative Boundaries**: A polygon shapefile that delineates the boundaries of the region being analyzed. This dataset is important for defining the study area and ensuring that all spatial analysis is performed within the correct limits.
3. **Existing High-Voltage (HV) Network** (Optional): A line shapefile representing the existing high-voltage grid infrastructure. This is used to calibrate the model and identify already electrified areas. If the data is unavailable, the model can proceed without it.
4. **Planned High-Voltage (HV) Network** (Optional): A point shapefile showing future high-voltage network extensions. This helps the model plan for future grid expansion.
5. **Road Network** (Optional): A line shapefile of existing roads, which helps model the feasibility of grid extension by considering transport infrastructure.
6. **GHI (Global Horizontal Irradiation)**: A raster dataset representing the solar potential in kWh/m^2/day. This data is used to estimate the feasibility of solar power generation in each grid cell.
7. **Wind Speed**: A raster dataset that provides wind velocity (m/s) for assessing wind energy potential.
8. **Nighttime Lights**: A raster dataset that helps identify electrified versus non-electrified areas based on nighttime light emissions, which are often available from satellite imagery (e.g., from the **VIIRS** dataset).
9. **Elevation and Slope**: Raster data providing the **elevation** of each area, which is essential for determining energy potentials and infrastructure suitability, especially for renewable energy sources.

For finding the data sources start from here:

https://onsset.readthedocs.io/en/latest/data_acquisition.html

Dataset	Description	Source Link
Population Density	High-resolution raster data representing population	https://data.subak.org/dataset/bd8bae5d-e1d3-4e1d-bbba-556e15ce55c0, https://data.worldpop.org/v1/reports/HRSL/BEN/BEN_population_v1_0.zip
Population clusters		https://data.mendeley.com/datasets/z9zfhzk8cr/3
Administrative Boundaries	Shapefiles delineating national and subnational boundaries	https://data.humdata.org/dataset/cod-ab-ben
Existing HV Network	Line shapefiles of current high-voltage grid infrastructure	https://osmtoday.com/africa/benin.html http://africagrid.energydata.info/
Planned HV Network (Optional)	Future high-voltage network extensions	http://africagrid.energydata.info/
MV Lines	Spatial distribution of medium voltage transmission network.	https://zenodo.org/record/3628142#.XxhXF55KhPY
Substations and transformers	Locations of substations and transformers	http://download.geofabrik.de/
Road Network	Line shapefiles of existing roads	https://osmtoday.com/africa/benin.html http://download.geofabrik.de/
GHI (Solar Potential)	Global Horizontal Irradiation in kWh/m²/day	https://globalsolaratlas.info/downloads
Small Scale Hydropower	Locations of small hydropower potential sites	https://www.hydropower.org/hydropower-database
Power plants	Locations of power generation facilities	https://datasets.wri.org/dataset/globalpowerplantdatabase
Wind Speed	Wind velocity data in m/s	https://globalwindatlas.info

Land Cover	Land cover classification maps	https://lpdaac.usgs.gov/products/mcd12q1v006/
Nighttime Lights	Raster data indicating electrified areas based on light emissions	https://eogdata.mines.edu/nighttime_light/
Elevation and Slope	Raster data providing elevation and slope	https://www2.jpl.nasa.gov/srtm/ http://www.cgiar-csi.org/data
Travel Time	Estimated travel time to nearest town of 50k people.	https://malariaatlas.org/research-project/accessibility_to_cities/

Configuring the Excel Input File

Instructions for Configuring and Using the Parameters in the Excel File for OnSSET (with Example Default Values for Benin)

To configure the CountrySpecs.xlsx file for OnSSET, follow these steps:

Open your spreadsheet editor (e.g., Microsoft Excel) and create a new file named, for example, BeninSpecs.xlsx.

Define the Columns

Ensure the column names in your file match exactly with the Parameter names listed below.

Enter the data under the corresponding column headers. Use the example values for Benin below for training purposes.

Parameter	Description	Unit	Default Value (Benin)
Country	Name of the study area		Benin
Pop2015	Population in the base year	People	11,175,000
UrbanRatio2015	Ratio of urban population in the base year	Ratio	0.45
Pop2030	Projected population	People	15,500,000
UrbanRatio2030	Projected urban population ratio	Ratio	0.55
NumPeoplePerHHRural	Number of people per rural household	People	5.5
NumPeoplePerHHUrban	Number of people per urban household	People	4.5

DieselPriceLow	Low diesel price (used in cost calculations for diesel generators)	$/liter	1.0
DieselPriceHigh	High diesel price (used in cost calculations for diesel generators)	$/liter	1.3
GridPrice	National grid electricity generation cost	$/kWh	0.15
GridCapacityInvestmentCost	Investment cost per additional grid capacity	$/kWh	500
GridLosses	Average technical losses on transmission and distribution	Ratio	0.15
BaseToPeak	Ratio of base to peak load	Ratio	0.6
ExistingGridCostRatio	Incremental cost increase for grid extension	Ratio	0.1
MaxGridExtensionDist	Maximum distance for grid extension	km	50
ElecActual	Electrification rate in the base year	Ratio	0.4
MinNightLights	Minimum night light intensity to consider a settlement as electrified	[0-63]	5
MaxGridDist	Maximum distance from grid to consider a settlement as electrified	km	10
MaxRoadDist	Maximum distance from road to consider a settlement as electrified	km	5
PopCutOffRoundOne	Minimum population to consider a settlement as electrified (Round 1)	People	200
PopCutOffRoundTwo	Minimum population to consider a settlement as electrified (Round 2)	People	500
UrbanCutOff	Minimum population to consider a settlement as urban	People	5,000

The file should be saved in **Excel .xlsx format**.

Storage Location:
Store the CountrySpecs.xlsx file in the **working directory** of the **OnSSET** model or in the same directory where the other **input data files** are stored

Step 5: Python Input Parameters

In addition to the Excel file, certain parameters are directly configured in the Python scripts. These parameters are defined in the OnSSET code files (e.g., runner.py and onsset.py). Below is a detailed table listing these parameters, their descriptions, units, default/example values, and their locations in the code

Example of Python Input Parameters:

Parameter	Description	Unit	Example Value	Location in Code
BASE_YEAR	The base year of the analysis.	Year	2020	runner.py (Scenario)

END_YEAR	The final year of the analysis.	Year	2030	runner.py (Scenario)
sa_diesel_capital_cost	Capital cost for standalone diesel generators	$/kW	938	runner.py (Technology)
sa_pv_capital_cost	Capital cost for standalone PV modules	$/kW	6950	runner.py (Technology)
mg_diesel_capital_cost	Capital cost for mini-grid diesel generators	$/kW	721	runner.py (Technology)
mg_pv_capital_cost	Capital cost for mini-grid PV systems	$/kW	2950	runner.py (Technology)
mg_wind_capital_cost	Capital cost for mini-grid wind systems	$/kW	3750	runner.py (Technology)
mg_hydro_capital_cost	Capital cost for mini-grid hydropower systems	$/kW	3000	runner.py (Technology)
discount_rate	Discount rate applied to cost calculations	Ratio	0.08	runner.py (Technology)
existing_grid_cost_ratio	Incremental cost increase for grid extension	Ratio	0.1	runner.py (Technology)
grid_cell_area	Size of each grid cell in the study area	km²	Varies	onsset.py (Settlement)
distribution_losses	Transmission and distribution losses	Ratio	0.05	runner.py (Technology)
diesel_price	Price of diesel for fuel cost calculations	$/liter	0.72	runner.py (Scenario)
LHV_DIESEL	Lower heating value for diesel	kWh/l	9.95	onsset.py (Constant)