gep$_{o}nsset$

Release 15-11-2021

KTH dESA

Apr 24, 2023

Contents

1 Contents 3

The extant documentation serves as a guide to gep_onsset. The latter is a modified version of OnSSET that was developed to support the functionalities of the Global electrification Platform.

The following pages, aim to cover aspects related to the repricability and/or reproducibility of the **gep_onsset** model. That is, the following sections provide a guide on how to set up, prepare, customize and run the model at different levels andt/or entry points, but they **do not** provide an exhaustive description of the model's core functionality (for this you may refer to here).

Note: Although the documentation is updated frequently, discrepancies might occur due to the evolving nature of the GEP project. Please do not hesitate to get in touch with us in case you observe any inconsistencies or you have any suggestions.

CHAPTER 1

Contents

1.1 Software installation

gep_onsset.py contains all the functions needed to prepare and run a GIS based electrification analysis. It is therefore important that the module is installed and set up properly on your local machine. The following paragraphs provide all information needed to get you started!

1.1.1 Install from GitHub

Access the gep_onsset repository on GitHub. Download the zipped repository directly or clone it to you designated local directory. The latter requires that git is installed in your machine. To clone the repo you may use the following command:

```
git clone https://github.com/global-electrification-platform/gep-onsset.git
```

1.1.2 Requirements & Working environment

gep_onsset.py (as well as all supporting scripts in this repo) have been developed in Python 3. We recommend installing Anaconda free distribution as suited for your operating system.

Once installed, you may open anaconda prompt and set up the working environment. The **gep_onsset_env.yml** file - located in the root directory of the repository - contains all necessary packages. You may use it to set up a new virtual environment by using:

```
conda env create --name gep_onsset_env --file gep_onsset_env.yml
```

This might take a while.. When complete, activate the virtual environment using:

```
conda activate gep_onsset_env
```

Now you are set to begin exploring the model!

1.1.3 Python Interfaces & IDEs

Integrated Development Environments are used in order to ease the programming process when multiple or long scripts are required. They allow the user to interact with the code (e.g. call functions, apply changes etc.).

Jupyter notebook (via Anaconda)

Jupyter notebook is a console-based, interactive computing approach providing a web-based application suitable for capturing the whole computation process: developing, documenting, and executing code, as well as communicating the results. Jupyter notebook is for example used for the GEP_generator interface.

PyCharm

For additional experimentation with the code one might prefer to use another IDE. While there are plenty of IDEs developed for Python, we suggest PyCharm Community version as it is a well-known, open access IDE.

Both of the above are described in more detail in the **scenario_run_reference** section.

1.1.4 Additional Info

- Basic navigating commands for DOS (cmd)
- Git Cheat Sheet
- Modules and packages installation documentation from python.org

1.2 GIS data collection

An electrification analysis with **gep_onsset** is based on information collected by a number of GIS layers. These are used to provide all necessary, initial attributes that the model needs to run.

A basic analysis relies on the following "foundamental" GIS layers:

- Distribution of HV lines (current & planned)
- Distribution of MV line
- Location of Substations & Transformers
- Road network
- Global Horizontal Irradiation
- Wind speed
- Location of Small Hydropower potential sites
- Land Cover
- Elevation & Slope
- Administrative boundaries
- Population distribution
- Travel time to nearest town
- Nighttime lights

- Custom Residential Electricity Demand Indicative Target (CREDIT) Layer

Other supplementary layers may be used depending on their availability and support the electrification analysis accordingly.

Below we provide key features of these layers in the form of metadata. The list is not exhaustive but rather focuses on the latest, open access datasets providing global or at least regional coverage. These have informed the scenario analysis available on the GEP Explorer.

Note: It is important to highlight that the selection of these datasets is not set in stone. They are interchangeable and may be replaced by alternative datasets as per case study mandates. Note however, that any update should comply with the suggested data guidelines developed as part of the GEP project.

1.2.1 Infrastructure

HV lines (current & planned)

Dataset	High Voltage (HV) lines
Data Type	Vector
Units	kV
Spatial Resolution	Regional, national
Description	Spatial distribution of (Existing & Planned) the transmission network. HV capacity definition depends on the country but usually refers to lines above 69 kV.
Why we are using this dataset	Identify where HV lines are; identify electrification status in the base year
Author	Open Street Map/The World Bank
Year	2017
Availability	Publickly available
Cleaned/Processed?	not available
Responsible Party	The World Bank
Learn More Link	https://energydata.info/dataset/africa-electricity-transmission-and-distribution-2017
Download from Source	http://africagrid.energydata.info/
Category	Transmission and distribution
Cautions	none
Supplementary Info	This dataset serves as an updated and improved replacement for the Africa Infrastructure Country Diagnostic (AICD) data that was published in 2007.
Geographic Coverage	Africa
CRS of Original File	EPSG:4326 - WGS 84 - Geographic
Date of Publication	March 24, 2017, 7:24 PM (UTC+01:00)
Date of Content	January 10, 2019, 12:06 PM (UTC+01:00)
Frequency of Updates	yearly
Summary of License (Open, Closed, Limited)	Open
License Type (if available)	Creative Commons Attribution 4.0
Link to License	https://creativecommons.org/licenses/by/4.0/
Citation	
Tags	Transmission; energy access; grid

MV lines

Dataset	Medium Voltage (MV) lines
Data Type	Vector
Units	kV
Spatial Resolution	Regional, national
Description	Spatial distribution of the medium voltage transmission network. What is defined as medium voltage depends on the country but usually refers to lines between 11-69 kV.
Why we are using this dataset	Identify where MV lines are; identify electrification status in the base year
Author	Christopher Arderne, Conrad Zorn, Claire Nicolas and Elco Koks
Year	2020
Availability	Publicly available
Cleaned/Processed?	not applicable
Responsible Party	not available
Learn More Link	https://gridfinder.org/
Download from Source	https://zenodo.org/record/3628142#.XxhXF55KhPY
Category	Transmission and distribution
Cautions	none
Supplementary Info	none
Geographic Coverage	Global
CRS of Original File	EPSG:4326 - WGS 84 - Geographic
Date of Publication	January 16, 2020
Date of Content	January 16, 2020
Frequency of Updates	non available
Summary of License (Open, Closed, Limited)	Open
License Type (if available)	Creative Commons Attribution 4.0 International
Link to License	Creative Commons Attribution 4.0 International
Citation	https://www.nature.com/articles/s41597-019-0347-4
Tags	Distribution; energy access; grid

Sub-stations & Transformers

Dataset	Substations & Transformers
Data Type	Vector
Units	kVA
Spatial Resolution	National
Description	The location of currently available substations and transformers.
Why we are using this dataset	Identify where sub-stations are; identify electrification status in the base year
Author	OpenStreetMap
Year	Up-to-date
Availability	Partially available
Cleaned/Processed?	Need to be processed and cross-validates
Responsible Party	OpenStreetMap
Learn More Link	none available
Download from Source	http://download.geofabrik.de/
Category	Grid infrastructure
Cautions	
Supplementary Info	
Geographic Coverage	Global
CRS of Original File	EPSG:4326 - WGS 84 - Geographic
Date of Publication	not available
Date of Content	not available
Frequency of Updates	Frequent
Summary of License (Open, Closed, Limited)	Open
License Type (if available)	not available
Link to License	not available
Citation	not available
Tags	Grid infrastructure; Sub-stations

Road network

Dataset	Road Network
Data Type	Vector
Units	
Spatial Resolution	National
Description	Existing & planned road infrastructure. The road network that is to be used has to include major roads such as highways, primary and secondary roads. There is no need to include smaller desolate roads or trails.
Why we are using this dataset	Calibration of electrification heuristics; fuel cost for diesel; grid penalty costing
Author	OSM (through Mapzen)
Year	2018
Availability	Available
Cleaned/Processed?	Processed
Responsible Party	Mapzen
Learn More Link	https://www.mapzen.com/blog/osmlr-2nd-technical-preview/
Download from Source	
Category	Transport
Cautions	OSMLR provides a stable linear-referencing system atop the ever-changing network of roadways in OpenStreetMap. It's used by the Open Traffic platform to associate statistics like speeds and vehicle counts with roadway segments.
Supplementary Info	OSMLR segments are available as geographic tiles at three levels of roadway hierarchy. The highway level (0) includes drivable road segments with OSM highway tags: motorway, motorway_link, trunk, trunk_link, primary, and primary_link. The arterial level (1) includes drivable road segments with OSM highway tags: secondary, secondary_link, tertiary, and tertiary_link. The local level (2) includes drivable road segments with OSM highway tags: unclassified, unclassified_link, residential, and residential_link.
Geographic Coverage	Global
CRS of Original File	EPSG:4326 - WGS 84 - Geographic
Date of Publication	
Date of Content	
Frequency of Updates	
Summary	Open

1.2.2 Energy (and other) resources

Global Horizontal Irradiation (GHI)

Dataset	Global Horizontal Irradiation (GHI)
Data Type	Raster
Units	kWh/m2/year
Spatial Resolution	0.0083 deg
Description	Provide information about the Global Horizontal Irradiation (kWh/m2/year) over an area.
Why we are using this dataset	
Author	SOLARGIS
Year	2017
Availability	Available
Cleaned/Processed?	Processed
Responsible Party	Energy Sector Management Assistance Program (ESMAP)
Learn More Link	https://globalsolaratlas.info/downloads?c=22.755921,-17.753906,2
Download from Source	
Category	Energy Resources
Cautions	
Supplementary Info	The Atlas covers areas between latitudes 60°N to 45°S. Areas north and south of these coordinates are not covered because the incline of the satellite imagery prohibits an accurate assessment of cloud cover. The primary grid resolution of solar resource data is approximately 3 to 7 km (depending on the latitude), which is enhanced by downscaling to a nominal resolution of approximately 1 km. The spatial resolution of other data parameters has been also harmonized to 1 km.
Geographic Coverage	Global
CRS of Original File	EPSG:4326 - WGS 84 - Geographic
Date of Publication	
Date of Content	
Frequency of Updates	
Summary of License (Open, Closed, Limited)	Open
License Type (if	Creative Commons Attribution license (CC BY 3.0 IGO)

Wind

Dataset	Wind speed or Power Density
Data Type	Raster
Units	m/s or W/m2
Spatial Resolution	0.01 deg
Description	Provide information about the wind velocity (m/sec) over an area. The wind power density map should provide information about the power density (W/m2) at a clearly stated altitude.
Why we are using this dataset	
Author	Technical University of Denmark ("DTU")
Year	2018
Availability	Available
Cleaned/Processed?	
Responsible Party	Energy Sector Management Assistance Program (ESMAP)
Learn More Link	https://globalwindatlas.info/downloads
Download from Source	
Category	Energy Resources
Cautions	
Supplementary Info	Vesrion 2.3. Wind resource mapping at 50, 100 and 200 m a.g.l.
Geographic Coverage	Global
CRS of Original File	EPSG:4326 - WGS 84 - Geographic
Date of Publication	
Date of Content	

Small Scale Hydropower

Dataset	Small scale Hydropower potential
Data Type	Vector
Units	
Spatial Resolution	National
Description	Points showing potential mini/small hydropower potential. Dataset developed by KTH dESA including environmental, social and topological restrictions and provides power availability in each identified point. Other sources can be used but should also provide such information to reassure the proper model function. This information is regarding the location of the plants, their power output, the head and the discharge connected to each point.
Why we are using this dataset	
Author	Alexandros Korkovelos
Year	2017
Availability	Available
Cleaned/Processed?	
Responsible Party	KTH Royal Institute of Technology
Learn More Link	https://energydata.info/dataset/small-and-mini-hydropower-potential-in-sub-saharan-africa
Download from Source	
Category	Energy Resources
Cautions	
Supplementary Info	
Geographic Coverage	Sub-Saharan Africa
CRS of Original File	EPSG:4326 - WGS 84 - Geographic
Date of Publication	June 19, 2017, 8:35 PM (UTC+02:00)
Date of Content	
Frequency of Updates	
Summary of License (Open, Closed, Limited)	Open
License Type (if available)	Creative Commons Attribution 4.0 International License. Full license text available at Creative Commons Attribution 4.0
Link to	https://creativecommons.org/licenses/by/4.0/

Land Cover

Dataset	Land cover
Data Type	Raster
Units	0-16, 254, 255
Spatial Resolution	0.00467 deg
Description	Land cover maps are used in a number of processes in the analysis (Energy potentials, restriction zones, grid extension suitability map etc.). Currently the land cover map used is divided into 17 classes. The classes are described in http://glcf.umd.edu/data/lc/. If this land cover map is replaced the land cover classification in OnSSET has to be altered. It is therefore advantageous if any land cover map that is used is classified similarly to the one described above.
Why we are using this dataset	
Author	GLCF
Year	2010
Availability	Available
Cleaned/Processed?	
Responsible Party	
Learn More Link	http://glcf.umd.edu/data/lc/
Download from Source	
Category	Land cover
Cautions	
Supplementary Info	Global Mosaics of the standard MODIS land cover type data product (MCD12Q1) in the IGBP Land Cover Type Classification are reprojected into geographic coordinates of latitude and longitude on the WGS 1984 coordinate reference system (EPSG: 4326). The data set boundaries are -180.0° <= longitude <= 180.0°; -64.0° <= latitude <= 84.0°. The data are organized as an array of values uniformly spaced across latitude and longitude with the indexed as [0, 0] at 84.0° latitude, -180.0° longitude. Spatially aggregated data for each year in the period 2001–2012 are available at two spatial resolutions: 5' x 5' resolution comprising 1776 rows x 4320 columns at a geographic pixel size of approximately 0.083333°; and 0.5° x 0.5° resolution comprising 296 rows x 720 columns of 0.5° pixels. The global land cover data sets are available as GeoTIFF format files (*.tif) with embedded metadata or as ESRI ASCII Grid format files (*.asc) with limited metadata in header lines. Native resolution data in the GLCF tile framework are available as GeoTIFF format files (*.tif).
Ge-	Global

Elevation

Dataset	Elevation
Data Type	Raster
Units	meters
Spatial Resolution	0.00083 deg
Description	Filled Digital Elevation Model (DEM) maps are used in a number of processes in the analysis (Energy potentials, restriction zones, grid extension suitability map etc.).
Why we are using this dataset	
Author	CGIAR-CSI
Year	2008
Availability	Available
Cleaned/Processed?	
Responsible Party	
Learn More Link	http://www.cgiar-csi.org/data
Download from Source	
Category	Land cover
Cautions	
Supplementary Info	Database v4.1
Geographic Coverage	Global
CRS of Original File	EPSG:4326 - WGS 84 - Geographic
Date of Publication	2008
Date of Content	
Frequency of Updates	
Summary of License (Open, Closed, Limited)	
License Type (if available)	
Link to License	
Citation	Jarvis, A., H.I. Reuter, A. Nelson, E. Guevara, 2008, Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org).
Tags	DEM; elevation map

Slope

Dataset	Slope
Data Type	Raster
Units	degrees
Spatial Resolution	0.00083 deg
Description	A sub product of DEM. The slope map visualizes the terrain slope in degrees. Any slope map that is to be used has to provide the slope in degrees.
Why we are using this dataset	
Author	KTH desa
Year	2017
Availability	Available
Cleaned/Processed?	Processed
Responsible Party	KTH dESA
Learn More Link	
Download from Source	
Category	Land cover
Cautions	
Supplementary Info	
Geographic Coverage	Africa
CRS of Original File	EPSG:4326 - WGS 84 - Geographic
Date of Publication	2017
Date of Content	
Frequency of Updates	
Summary of License (Open, Closed, Limited)	Open
License Type (if available)	Creative Commons Attribution 4.0
Link to License	https://creativecommons.org/licenses/by/4.0/
Citation	
Tags	slope; elevation; Africa

1.2.3 Socio-economic

Administrative units

Dataset	Administrative Boundaries
Data Type	Vector
Units	
Spatial Resolution	National, sub-national
Description	Includes information (e.g. name) of the country(s) to be modelled and delineates the boundaries of the analysis.
Why we are using this dataset	
Author	GADM
Year	2018
Availability	Available
Cleaned/Processed?	
Responsible Party	GADM
Learn More Link	https://gadm.org/download_country_v3.html
Download from Source	
Category	Socio-economic
Cautions	
Supplementary Info	Version 3.6
Geographic Coverage	Global
CRS of Original File	EPSG:4326 - WGS 84 - Geographic
Date of Publication	6 May 2018
Date of Content	
Frequency of Updates	3-6 months
Summary of License (Open, Closed, Limited)	Open
License Type (if available)	The data are freely available for academic use and other non-commercial use. Redistribution, or commercial use, is not allowed without prior permission. Using the data to create maps for academic publishing is allowed.
Link to License	
Citation	
Tags	administrative boundaries

Population

Dataset	Population clusters - distribution & density
Data Type	Vector
Units	
Spatial Resolution	National
Description	Spatial quantification of the population for a selected area of interest (usually country or continent).
Why we are using this dataset	
Author	Babak Khavari, Andreas Sahlberg, Alexandros Korkovelos, Mark Howells
Year	2019
Availability	Available
Cleaned/Processed?	Processed
Responsible Party	KTH dESA
Learn More Link	
Download from Source	https://data.mendeley.com/datasets/z9zfhzk8cr/4
Category	Socio-economic
Cautions	
Supplementary Info	
Geographic Coverage	Malawi
CRS of Original File	EPSG:4326 - WGS 84 - Geographic
Date of Publication	February 1 2019
Date of Content	
Frequency of Updates	yearly
Summary of License (Open, Closed, Limited)	Open
License Type (if available)	Creative Commons Attribution 4.0
Link to License	https://creativecommons.org/licenses/by/4.0/
Citation	http://dx.doi.org/10.17632/z9zfhzk8cr.4
Tags	population; clusters; settlements

Travel time

Dataset	Travel time
Data Type	Raster
Units	minutes
Spatial Resolution	0.0083 deg
Description	Visualizes spatially the travel time required to reach from any individual cell to the closest town with population more than 50,000 people. The unit of these maps should preferably be in minutes but hours is also acceptable.
Why we are using this dataset	
Author	map
Year	2015
Availability	Available
Cleaned/Processed?	
Responsible Party	
Learn More Link	https://map.ox.ac.uk/research-project/accessibility_to_cities/
Download from Source	
Category	Transport; socio-economic
Cautions	
Supplementary Info	In the present study, we quantify and validate global accessibility to high-density urban centres at a resolution of 1×1 kilometre for 2015, as measured by travel time. The last global mapping effort to measure accessibility was for the year 2000, a time that predates both substantial investment and expansion of transportation infrastructure and an extraordinary improvement in the data quantity and quality of accessibility measures. The game-changing improvement underpinning this work is the first-ever, global-scale synthesis of two leading roads datasets – Open Street Map (OSM) data and distance-to-roads data derived from the Google roads database – which resulted in a nearly five-fold increase in the mapped road area relative to that used to produce the circa 2000 map. A major strength of the new roads data is its inclusion of minor roads (e.g., unpaved rural roads), which comprise a large proportion of roads in many low-resource settings and were largely absent or geographically inaccurate in previous roads databases. As such, the improvements in our accessibility map are most prominent in the areas where quality data are most needed for informing sustainable development policies and actions. To illustrate the far-reaching utility of our 2015 global accessibility map, we conduct exploratory analyses that enumerate geographic and wealth-based inequities in

Nighttime Lights

Dataset	Nighttime Lights (NTL)
Data Type	Raster
Units	nW cm^-2 sr^-1
Spatial Resolution	0.00417 deg
Description	Nighttime light maps showing light pollution. The map shows stable light source wiht the unit nW cm^2 sr^1. Available on a yearly basis and monhtly basis. The monthly data is not cleaned of noise and outliers while the yearly one is. Latest yearly dataset is from 2016
Why we are using this dataset	Night-time light maps capture anthropogenic light sources on the surface of the earth using satellite imagery. It is a good proxy for assessing where electrified human settlements are, as these tend to give light pollution. In OnSSET nighttime light maps are used to estimate the location of currently electrified population.
Author	NOAA National Centers for Environmental Information (NCEI)
Year	2016
Availability	Available
Cleaned/Processed?	Cloud free composite
Responsible Party	NOAA National Centers for Environmental Information (NCEI)
Learn More Link	https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html
Download from Source	https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html
Category	Night time lights; Socio-economic
Cautions	Nighttime light maps mostly capture light from outdoor sources; in many cases outdoor light is not a very good indicator of household electricity.
Supplementary Info	
Geographic Coverage	Global
CRS of Original File	EPSG:4326 - WGS 84 - Geographic
Date of Publication	
Date of Content	
Frequency of Updates	yearly
Summary of License (Open, Closed, Limited)	
License Type (if available)	
Link to License	
Citation	
Tags	nighttime lights; NOAA

Residential Electricity Demand target layer

Dataset	Residential demand
Data Type	Raster
Units	kWh/capita/year
Spatial Resolution	0.0083 deg
Description	Layer that indicates electricity demand for residential sector (e.g. WRI's perspective map)
Why we are using this dataset	
Author	KTH dESA
Year	2019
Availability	Potentially available
Cleaned/Processed?	Processed
Responsible Party	KTH dESA
Learn More Link	
Download from Source	
Category	Socio-economic
Cautions	
Supplementary Info	
Geographic Coverage	Malawi
CRS of Original File	EPSG:4326 - WGS 84 - Geographic
Date of Publication	
Date of Content	
Frequency of Updates	yearly
Summary of License (Open, Closed, Limited)	Open
License Type (if available)	Creative Commons Attribution 4.0
Link to License	https://creativecommons.org/licenses/by/4.0/
Citation	
Tags	electricity demand; households: energy access

1.2.4 Supplementary layers

Power Plants (existing & Planned)

Dataset	Power Plants (Existing & Planned)
Data Type	Vector
Units	kW
Spatial Resolution	National
Description	The locations of existing and planned power plants. It is also important that the dataset includes attributes regarding each plant's minimum capacity.
Why we are using this dataset	
Author	
Year	2018
Availability	Available
Cleaned/Processed?	
Responsible Party	World Resources Institute
Learn More Link	http://datasets.wri.org/dataset/globalpowerplantdatabase
Download from Source	
Category	Climate; Energy
Cautions	
Supplementary Info	The Global Power Plant Database is a comprehensive, open source database of power plants around the world. It centralizes power plant data to make it easier to navigate, compare and draw insights for one's own analysis. Each power plant is geolocated and entries contain information on plant capacity, generation, ownership, and fuel type. As of June 2018, the database includes around 28,500 power plants from 164 countries. It will be continuously updated as data becomes available. The most recent release of the Global Power Plant Database 1.1 includes the addition of two countries (China and Fiji), over 3,000 power plants, and nearly 1300 gigawatts of power capacity. We highly recommend using version 1.1, available online as of June 2018.
Geographic Coverage	Global
CRS	
CRS of Original File	
Date	June 11, 2018

Poverty maps

Dataset	Poverty maps
Data Type	Raster or vector
Units	%
Spatial Resolution	0.0083 deg
Description	Poverty maps stating the headcount for the population below the poverty line. These poverty maps should be on the basis of a known administrative areas. The poverty line used should be clearly stated. If the poverty maps are available as raster maps for the studied countries it would preferable.
Why we are using this dataset	
Author	Worldpop
Year	2018
Availability	Available
Cleaned/Processed?	
Responsible Party	
Learn More Link	http://www.worldpop.org.uk/data/get_data/
Download from Source	
Category	Socio-economic
Cautions	
Supplementary Info	DATASET: Alpha version 2008 estimates of proportion of people per grid square living in poverty, as defined by the Multidimensional Poverty Index (http://www.ophi.org.uk/policy/multidimensional-poverty-index/), and associated uncertainty metrics. UNITS: Proportion of residents living in MPI-defined poverty (poverty dataset); 95% credible interval (uncertainty dataset). MAPPING APPROACH: Bayesian model-based geostatistics in combination with high resolution gridded spatial covariates applied to GPS-located household survey data on poverty from the DHS and/or LSMS programs.
Geographic Coverage	Kenya, Malawi, Nigeria, Uganda, Tanzania, Bangladesh, Pakistan
CRS of Original File	EPSG:4326 - WGS 84 - Geographic
Date of Publication	January 2013
Date of Content	
Frequency	

GDP PPP

Dataset	GDP PPP
Data Type	Raster
Units	$
Spatial Resolution	0.0083 deg
Description	GDP map used should be a global raster map and show the purchasing power parity.
Why we are using this dataset	
Author	Kummu Matti, Taka Maija, Guillaume Joseph H.A.
Year	2018
Availability	Available
Cleaned/Processed?	
Responsible Party	
Learn More Link	https://datadryad.org/resource/doi:10.5061/dryad.dk1j0/13
Download from Source	
Category	Socio-economic
Cautions	
Supplementary Info	This global dataset represents the gross domestic production (GDP) of each grid cell. GDP is given in 2011 international US dollars. The data is derived from GDP per capita (PPP) which is multiplied by gridded population data HYDE 3.2 (the years of population data not available (1991-1999) were linearly interpolated at grid scale based on data from years 1990 and 2000). Dataset has global extent at 5 arc-min resolution for the 26-year period of 1990-2015. Detail description is given in a linked article and metadata is provided as an attribute in the NetCDF file itself.
Geographic Coverage	Global
CRS of Original File	EPSG:4326 - WGS 84 - Geographic
Date of Publication	2018-02-01
Date of Content	
Frequency of Updates	

HDI

Dataset	HDI
Data Type	Raster
Units	0-1
Spatial Resolution	0.083 deg
Description	HDI map can be used in combination with GDP maps in order to assess electricity demand goals. These maps should be in raster format as HDI varies considerably within countries.
Why we are using this dataset	
Author	Kummu Matti, Taka Maija, Guillaume Joseph H.A.
Year	2018
Availability	Available
Cleaned/Processed?	
Responsible Party	
Learn More Link	https://datadryad.org/resource/doi:10.5061/dryad.dk1j0/10
Download from Source	
Category	Socio-economic
Cautions	
Supplementary Info	HDI is a composite index of average achievement in key dimensions of human development (dimensionless indicator between 0 and 1). This index is based on method introduced 2010 and updated 2011. The subnational data for HDI were collected from multiple national-level datasets, and national-level HDI was collected from UNDP. Years with missing data were interpolated over time thin plate spines, assuming smooth trend over time. The dataset has a global extent at 5 arc-min resolution, and the annual data is available for each year over 1990-2015. HDI sub-national data covers 39 countries and 66% of global population in 2015.
Geographic Coverage	Global
CRS of Original File	EPSG:4326 - WGS 84 - Geographic
Date of Publication	2018-02-01
Date of Content	
Fre-	

Income level or Energy expenditure

Dataset	Income level or Energy expenditure
Data Type	Vector or Raster
Units	$/year
Spatial Resolution	best available
Description	The income level or energy expenditure in an area could potentially be used for heat-maps identifying higher demand. These maps are preferably available on the basis of known administrative areas
Why we are using this dataset	
Author	
Year	
Availability	Not available
Cleaned/Processed?	
Responsible Party	
Learn More Link	
Download from Source	
Category	
Cautions	
Supplementary Info	
Geographic Coverage	
CRS of Original File	
Date of Publication	
Date of Content	
Frequency of Updates	
Summary of License (Open, Closed, Limited)	
License Type (if available)	
Link to License	
Citation	
Tags	

Biomass

Dataset	Biomass
Data Type	Raster
Units	not available
Spatial Resolution	not available
Description	Current and potentially productive agricultural activity as an indicator of agricultural residues.
Why we are using this dataset	
Author	
Year	
Availability	Potentially available
Cleaned/Processed?	
Responsible Party	
Learn More Link	
Download from Source	
Category	
Cautions	
Supplementary Info	
Geographic Coverage	
CRS of Original File	
Date of Publication	
Date of Content	
Frequency of Updates	
Summary of License (Open, Closed, Limited)	
License Type (if available)	
Link to License	
Citation	
Tags	

Electricity demand for education facilities

Dataset	Productive uses - Electricity demand for education
Data Type	Raster
Units	kWh/year
Spatial Resolution	best available
Description	Locations of schools.If there are additional data on school districts (in order to know to which school the population in a certain cell is going to) or the energy demand in the schools it would be useful.
Why we are using this dataset	
Author	
Year	
Availability	Not available
Cleaned/Processed?	
Responsible Party	
Learn More Link	
Download from Source	
Category	
Cautions	
Supplementary Info	
Geographic Coverage	
CRS of Original File	
Date of Publication	
Date of Content	
Frequency of Updates	
Summary of License (Open, Closed, Limited)	
License Type (if available)	
Link to License	
Citation	
Tags	

Electricity demand for health facilities

Dataset	Productive uses - Electricity demand for health
Data Type	Raster
Units	kWh/year
Spatial Resolution	best available
Description	Locations of health clinics in the study area. If there are estimates of the energy demand in the health clinics this could also potentially be useful for the analysis.
Why we are using this dataset	
Author	
Year	
Availability	Not available
Cleaned/Processed?	
Responsible Party	
Learn More Link	
Download from Source	
Category	
Cautions	
Supplementary Info	
Geographic Coverage	
CRS of Original File	
Date of Publication	
Date of Content	
Frequency of Updates	
Summary of License (Open, Closed, Limited)	
License Type (if available)	
Link to License	
Citation	
Tags	

Electricity demand in commercial facilities

Dataset	Productive uses - Electricity demand for commercial uses
Data Type	Raster
Units	kWh/year
Spatial Resolution	best available
Description	Maps showing electricity demand for commercial activity (mines, stores etc.). This is an important dataset since mines tend to use large quantities of electricity.
Why we are using this dataset	
Author	
Year	
Availability	Not (publicly) available
Cleaned/Processed?	
Responsible Party	
Learn More Link	
Download from Source	
Category	
Cautions	
Supplementary Info	
Geographic Coverage	
CRS of Original File	
Date of Publication	
Date of Content	
Frequency of Updates	
Summary of License (Open, Closed, Limited)	
License Type (if available)	
Link to License	
Citation	
Tags	

Electricity demand for agriculture activities

Dataset	Productive uses - Electricity demand for Agriculture
Data Type	Raster
Units	kWh/year
Spatial Resolution	best available
Description	Maps showing the productive uses of electricity within the agricultural sector or areas that can be expected to have a large amount of agricultural activity are useful when estimating the productive uses.
Why we are using this dataset	
Author	
Year	
Availability	Potentially available
Cleaned/Processed?	
Responsible Party	
Learn More Link	
Download from Source	
Category	
Cautions	
Supplementary Info	
Geographic Coverage	
CRS of Original File	
Date of Publication	
Date of Content	
Frequency of Updates	
Summary of License (Open, Closed, Limited)	
License Type (if available)	
Link to License	
Citation	
Tags	

Mobile coverage

Dataset	Mobile phone coverage
Data Type	Raster
Units	0-1
Spatial Resolution	best available
Description	Indication of where the is mobile phone coverage (service); usually in binary format (1:coverage, 0: no-coverage). It can work as a proxy of locations that are electrified
Why we are using this dataset	
Author	
Year	
Availability	Not (publicly) available
Cleaned/Processed?	
Responsible Party	
Learn More Link	
Download from Source	
Category	
Cautions	
Supplementary Info	
Geographic Coverage	
CRS of Original File	
Date of Publication	
Date of Content	
Frequency of Updates	
Summary of License (Open, Closed, Limited)	
License Type (if available)	
Link to License	
Citation	
Tags	

1.3 Preparation of the primary input file

The basis of an electrification analysis with **gep_onsset** is the geo-location of population or settlements. Distribution of population might be available as a raster layer (e.g. WorldPop) or as vector layer (SEDAC).

1.3.1 Creating population "clusters"

In GEP, we have processed existing population datasets (and HRSL in particular) to create population "clusters". A cluster is a bundle of pixelated areas in close proximity to each other that merge and create a vector polygon. Clusters - in contract to gridded population - have various geometries and sizes and therefore better reflect the geography and behaviour of human settmements.

You may find more information on the underlying methodology in this publication. The methodology is also available for testing and experimentation in the form of an open source Qgis plugin. Population clusters developed for Sub-Saharan African countries are openly accecible via this link.

Note: Although clusters improve the geographic representation of human settlements, their development process comes with its own caveats. For example, densily populated urban areas tend to create big population conglomerates that defy the logic of granular analysis that geo-spatial electrification modelling might bring. You may read more in this publication.

1.3.2 Adding attributes to population settlements

Regardless of the source's format, the population layer is eventually converted into a point layer; each point representing the center of the area it covers. In order to prepare the primary input file to **gep_onsset** the values of the foundamental GIS datasets (see previous section) need to be extracted and attributed to each settlement (or point) respectively.

The extraction process can be executed manually in any GIS environment. This however, involves various commands depending on the type of attribute to be extracted (e.g. Zonal statistics, Raster value to points, nearest neighbor etc.).

Therefore, a Qgis plugin has been developed to automate this process. The plugin comes with installation and how-to-use instructions that allow easy replication of the process for an area of interest.

Note: In case gridded (raster) population is used instead of clusters in the first place, the extraction process may be executed via a modified version of the Qgis Plugin available in this repository.

1.3.3 Example of the primary input file

This process creates the primary input file for the **gep_onsset** model. The Qgis plugin exports the result in the form of a .csv file. An example of how this file should look like, is available here (as Malawi.csv).

Note: The number of columns may differ depending on the GIS layer availability (see previous section). Below follows an example of what one shall expect. In case they are not, you may re-visit the extraction process or add the missing layer.

#	Column	Unit	Description
1	Country	name	Name of the country in focus (e.g. Malawi)
2	NightLights	nW cm^2 sr^1	Average yearly value of stable night lights luminosity. Value is used (toge
3	Pop	people	Number of people living in the settlement, as retrieved from the GIS data
4	id	number	Unique identifier of each settlement
5	GridCellArea	sq.km	Area of the settlement; retrieved from population cluster development and
6	ElecPop	people	Number of people with access to electricity in the base year; Value is retri
7	WindVel	m/s	Yearly average wind speed in the area of the settlement
8	GHI	kWh/m^2/year	Yearly average Global Horizontal Irradiation in the area of the settlement
9	TravelHours	hours	Travel time to nearest town of 50k people; in the case of (polygon) cluster
10	Elevation	m	Above sea level; in the case of cluster this value reflects the average elevat
11	Slope	deg	A product of DEM indicating terrain slope; in the case of cluster this valu
12	ResidentialDemandTIerCustom	kWh/capita/year	Electricity demand target based on a custom-based, bottom up approach (i
13	LandCover	1 to 15	Type of land cover as defined by the source data; refer to documentation
14	SubstationDist	km	Distance to nearest sub-station; based on best available GIS data sources i
15	CurrentHVLineDist	km	Distance to nearest HV line; based on best available GIS data sources in t

Table 1 – continued from

#	Column	Unit	Description
16	CurrentMVLineDist	km	Distance to nearest MV line; based on best available GIS data sources in t
17	RoadDist	km	Distance to nearest (primary/secondary) road; based on best available GIS
18	X_deg	deg	Longitude
19	Y_deg	deg	Latitude
20	TransformerDist	km	Distance to nearest service transformer; based on best available GIS data s
21	PlannedMVLineDist	km	Distance to nearest planned MV line; based on best available GIS data; in
22	PlannedHVLineDist	km	Distance to nearest planned HV line; based on best available GIS data; in
23	HydropowerDist	km	Distance to nearest site with identified small scale hydropower potential
24	Hydropower	kW	Technical potential of the nearest small scale hydropower site
25	HydropowerFID	number	Unique identified of the nearest small scale hydropower site
26	IsUrban	0,1,2	Indicates Urban/Rural status of the settlement; 2: urban, 1 and 0: rural; 0 r
27	PerCapitaDemand	kWh/capita/year	Electricity demand target based on urban/rural status and targets set for ea
28	HealthDemand	kWh/year	Electricity demand target to support health related activities in the settlem
29	EducationDemand	kWh/year	Electricity demand target to support education related activities in the sett
30	AgriDemand	kWh/year	Electricity demand target to support agriculture related activities in the set
31	ElectrificationOrder	number	Indicates the loop in which the settlement gets electrified; serves only dev
32	Conflict	0-4	Indicates conflict level, which in turn affects costing of electrification; 0: r
33	CommercialDemand	kWh/capita/year	Electricity demand target to support commercial activities in the settlemer
34	ResidentialDemandTier1	kWh/capita/year	Electricity demand target for Tier 1
35	ResidentialDemandTier2	kWh/capita/year	Electricity demand target for Tier 2
36	ResidentialDemandTier3	kWh/capita/year	Electricity demand target for Tier 3
37	ResidentialDemandTier4	kWh/capita/year	Electricity demand target for Tier 4
38	ResidentialDemandTier5	kWh/capita/year	Electricity demand target for Tier 5

1.4 Input file calibration and update

The primary input file (see previous section) includes rough data as extracted from the GIS layers. Before proceeding with the electrification analysis, these values need to be conditioned and/or calibrated.

- **Conditioning** makes sure that physical values (e.g. GHI, Wind speed, land cover, elevation etc.) are within acceptable limits.
- **Calibration** makes sure some of the parameters (e.g. total population, urban/rural ration, electrification rate etc.) are in line with official statistics.
- **New columns** are also created and support later stages of the analysis (e.g. wind capacity factor, grid penalty ratio, electrification status etc.).

1.4.1 Calibration with GEP Generator

The GEP Generator is an interactive interface, developed as a jupyter notebook (.ipynb) in order to support calling functions in the **gep_onsset** model. The GEP Generator is located in the root directory of the repository. You may access it by simply navigating there via anaconda prompt using:

```
> cd ..\my_designated_local_directory

..\my_designated_local_directory> jupyter notebook
```

Jupyter notebook will open on your default browser; simply select to open the `GEP Generator.ipynb` and you are set to go!

The GEP Generator runs in 9 steps (or blocks). Steps 1-5 are used to define calibration parameters and conduct the conditioning/calibration process. In particular,

- **Step 1** requires that the user provides the primary input file (see previous section)
- **Steps 2 & 4** allow the user to interactively provide the calibrating parameters
- **Step 5** conducts the conditioning and calibration process

---

Note:

- **Step 3** is related to the definition of scenario parameters and is discussed in more detail in the following section.
- The GEP generator **does not** store the calibrated results in a separate file but rather continues right away to the scenario runs. This makes the process faster on the one hand, but it means that the conditioning and calibration process runs anew everytime a scenario is executed (only one at a time).

1.4.2 Calibration with gep_runner

The **gep_runner.py** is an alternative way to call functions from the **gep_onsset.py**. You may execute **gep_runner** in any IDE of preference, we suggest PyCharm. Interaction with the code using **gep_runner** takes place in the python console of your IDE. Upon initiation, the code will prompt you to select one of the three following options:

- 1: To split countries in case of multiple country runs (used rarely)
- **2: To prepare/calibrate the GIS input file**
- 3: To run scenario(s)

For calibration you may select option 2. Execution requires two files:

- **The primary input file** (see *Malawi.csv* from previous section)
- **The specs file** (see example of specs_mw_one_scenario.csv).

The specs file contains the parameters and their values against which the GIS data are conditioned or calibrated. The user shall fill in all necessary values in the `SpecsData` sheet. A description of the parameters is presented below.

Parameter	Description
Country	Name of the country
CountryCode	ALPHA-2 country code as per ISO 3166 international standard
StartYear	Base year of the analysis; usually selected based on data availability
EndYEar	End year of the analysis
PopStartYear	Official population at the base year
UrbanRatioStartYear	Official urban population ratio in the base year
UrbanCutOff	Population threshold above which a settlement can be considered urban (optional)
UrbanRatioModelled	This value is provided by the model after calibration
PopEndYearHigh	Expected population at the end year based on high growth rate
PopEndYearLow	Expected population at the end year based on high growth rate
UrbanRatioEndYear	Expected urban population ration in the end year
NumPeoplePerHHRural	Number of people per household - rural settlements
NumPeoplePerHHUrban	Number of people per household - urban settlements
GridCapacityInvestmentCost	Expected investment cost per kW of additional capacity in the central grid system
GridLosses	Expected transmission ans distribution losses in the grid network
BaseToPeak	Average to peak load ratio for the grid; used for sizing additional capacity due grid e
ExistingGridCostRatio	Persentage of capital cost increase in each grid extension iteration; used to accomm

Table 2 – continued from previous page

Parameter	Description
MaxGridExtensionDist	Maximum distance (in km) that MV lines can reach in each iteration loop
NewGridGenerationCapacityAnnualLimitMW	Capacity that can be added to the central grid per year of analysis
ElecActual	Official national electrification rate in the base year
Rural_elec_ratio	Official national electrification rate in rural areas in the base year
Urban_elec_ratio	Official national electrification rate in urban areas in the base year
ElecModelled	This value is provided by the model after calibration
urban_elec_ratio_modelled	This value is provided by the model after calibration
rural_elec_ratio_modelled	This value is provided by the model after calibration
MinNightLights	Nighttime light value above which a settlement can be consedered electrified; used t
DistToTrans	Distance to transformers (in km) above which a settlement can be consedered electri
MaxGridDist	Distance to T&D network (in km) above which a settlement can be consedered elect
MaxRoadDist	Distance to road network (in km) above which a settlement can be consedered electr
PopCutOffRoundOne	This value is provided by the model after calibration
PopCutOffRoundTwo	This value is provided by the model after calibration

With **gep_runner** the calibration process is separated from the scenario runs. That is, the code stops once the conditioning and calibration process is complete. The result is exported in a "calibrated" input file.

The result file and the updated parameters in the specs file should be reviewed to decide whether the result is satisfactory or the process requires further calibration. Key outputs to cross-check include:

- Population projection
- Modelled urban/rural classification
- Modelled electrification rate (national, urban, rural)

Note: The conditioning & calibration process is driven by relevant functions located in **gep_onsset.py**. One can access and modify these functions in case their existing form does not serve the intended purpose. This requires some experience with the model; in case you are a new user you may experiment with the GEP Generator first before engaging in modification of the core code.

1.4.3 Example of the calibrated input file

The calibration process will add the following columns to the input file.

#	Column	Unit	Description
39	PopStartYear	people	Calibrated population to match with official statistics in the base year
40	Pop<year>High	people	Projected population in the specified <year> based on high growth indicators; for intermediate and end years
41	Pop<year>Low	people	Projected population in the specified <year> based on low growth indicators; for intermediate and end years
42	Pop<base year>	people	Calibrated population to match with official statistics in the base year
43	RoadDistClassified	1 to 5	Classified value of distance to road used to calculate grid penalty factor
44	SubstationDistClassified	1 to 5	Classified value of distance to sub-station used to calculate grid penalty factor
45	LandCoverClassified	1 to 5	Classified value of land cover type used to calculate grid penalty factor
46	ElevationClassified	1 to 5	Classified value of elevation used to calculate grid penalty factor
47	SlopeClassified	1 to 5	Classified value of slope used to calculate grid penalty factor
48	GridClassification	1 to 5	Grid extension suitability index; Higher value indicates higher suitability; based on an Analytic Hierarchy Process (AHP) over the above parameters
49	GridPenalty	number	Grid extension cost multiplier based on above classification; default value 1 induces no additional costs
50	WindCF	from ~0 to 1	Wind capacity factor estimated based on available wind speed and power rating of Vestas V-44 600kW turbine
51	ElecPopCalib	people	Number of people with access to (grid) electricity, calibrated to match official statistics in the base year
52	ElecStart	0,1	Electrification status in the base year; 0: non-electrified 1: electrified (by the grid)
53	GridDistCalibElec	km	Distance to nearest power infrastructure element (transformer, MV, HV); based on their availability. In case transformers are not available it will lookup the next available element (e.g. MV)
54	Elec_Initial_Status_<base year>	0,1	Grid-side electricity status in the specified base year; 0: non-electrified 1: electrified by the grid
55	Elec_Init_Status_Offgrid_<base year>	0,1	Off-grid electricity status in the specified base year; 0: non-electrified 1: electrified by an off-grid technology
56	Actual_Elec_Status_<base year>	0,1	Overall electrification status in the specified year; 0: non-electrified 1: electrified by any technology
57	FinalElecCode<base year>	1 to 8 or 99	Code of electrifying technology in the specified year (1: grid, 2: stand-alone diesel, 3: stand-alone PV, 4: Mini-grid diesel, 5: Mini-grid PV, 6: Mini-grid Wind, 7: Mini-grid Hydro, 8: Hybrid Mini-grid, 99: not-electrified)
58	GridReachYear	year	Estimated year that the grid might be able to reach this settlement; currently deactivated and not used in the GEP

When the calibration process is complete you may proceed with running an electrification scenario (see next section)!

1.5 Scenario run

If the previous steps have been successful, running an electrification scenario with **gep_onsset** is a fairly straightforward process.

1.5.1 Running with GEP Generator

The simplest way to run a scenario is via `GEP Generator.ipynb`. You may refer to the previous section on how to get is up running. In order to run a scenario you will need:

- **The primary input file** (see *Malawi.csv* from previous section)
- **Fill in the calibration parameters** (same as *Steps 2,4 and 5* from previous section)
- **Provide the scenario parameters** (*Steps 3 & 6*)

Note: The GEP "levers"

The GEP levers refer to 7 key decision parameters, the selection of which can drastically change the output of the electrification analysis. They cover the following:

- Population growth rate
- Level of electricity consumption in to-be-electrified settlements
- Targeted national electrification rate in the intermediate year
- Expected electricity generating cost for the central grid
- Capital cost of photovoltaic systems
- Diesel price
- Electricity demand targets for productive uses (agriculture, health, education)
- Rollout plan - prioritization (e.g who gets electricity first and how?)

Each lever has 1-3 possible options the combination of which can generate 216 scenario as presented in the GEP Explorer. The GEP generator guides the selection process with embedded documentation and link to the relevant sources.

Step 7 runs the electrification analysis for the specified scenario. **Note** that in this step, you may also provide the parameter *cost_choice* if you want to include (1) or exclude (2) break down of investment costs. The first option will add computational time in the analysis.

Step 8 prepares a summary table, four graphs and a map over the key results of the analysis for a quick, on-the-fly review.

Step 9 exports the results into 3 csv files:

- The **_Variables**.csv file provides a summary of input variables
- The **_Summaries**.csv file provides a summary of key results of the electrification analysis
- The **_Results**.csv file provides the electrification results in full granularity. An overview of the content is available in the next section

1.5.2 Running with gep_runner

The **gep_runner** is usually used to run multiple scenarios at once. As shown in the previous section, interaction with the code takes place in the python console. Upon initiation, you may select option 3 for scenario run(s).

- 1: To split countries in case of multiple country runs (used rarely)
- 2: To prepare/calibrate the GIS input file
- **3: To run scenario(s)**

Note:

1. In the latest update, you will also be prompted to provide the parameter *cost_choice* if you want to include (1) or exclude (2) break down of investment costs. The first option will add computational time in the analysis.
2. It is also highly recommended that you use the existing patterns `# TODO`, `# RUN_PARAM` and `# REVIEW` to navigate through the **gep_onsset.py** and **gep_runner.py** code. You can find more info on how to activate those in PyCharm here.

Execution requires two files:

- **The specs file** (see previous section)
- **The calibrated input file** (see from previous section)

Scenario definition is possible in the `ScenarioInfo` sheet of the specs file. There one can parameterize the "levers" accordingly and create a bundle of scenarios. Each row represents one potential scenario. The **gep_runner** will run as many scenarios as defined in this sheet.

The following table gives on overview of the potential scenario combinations.

Lever	Option	Description
Population_Growth	0, 1	Expected population in the country by the end year of the analysis; 0: low population growth, 1: high population growth
Target_electricity_consumption_level	0, 1, 2	0: low electricity demand target (e.g. U4R1), 1: high electricity demand target (e.g. U5R3), 2: use the custom residential demand target layer (from GIS)
Electrification_target_5_years	0, 1	0: low electrification target in the intermediate year (e.g. 35%), 1: high electrification target in the intermediate year (e.g. 60%)
Grid_electricity_generation_cost	0, 1	0: low generating cost for the grid (e.g. 0.03), 1: high generating cost for the grid (e.g. 0.08)
PV_cost_adjust	0, 1, 2	0: PV capacity cost as defined by the user, 1: PV capacity cost reduced by 25%, 2: PV capacity cost increased by 25%
Diesel_price	0, 1	0: low diesel price , 1: high diesel price
Productive_uses_demand	0, 1	0: not including productive uses of electricity, 1: including productive uses of electricity
Prioritization_algorithm	0, 1, 2	0: least cost prioritization, 1: forced grid within 1km, 2: forced grid within 2km

Therefore, the scenario `0_0_0_0_0_0` will respectively represent:

- low population growth
- low electricity demand target (e.g. U4R1)
- low electrification target in the intermediate year (e.g. 35%)
- low generating cost for the grid (e.g. 0.03)

- PV capacity cost as defined by the user
- low diesel price
- not including productive uses of electricity
- least cost prioritization

Note that in the `ScenarioParameters` sheet one can customize how the aforementioned codes are translated to tangible input variables in the **gep_onsset** code.

The **gep_runner** yields two csv files for each scenario.

- The **_Summaries**.csv file that provides a summary of key results of the electrification analysis
- The **_Results**.csv file that provides the electrification results in full granularity

---

Note: The scenario coding convention is applied in the naming process of the output result files as well. For example the same scenario for Malawi would yield the result file names `mw-1-0_0_0_0_0_0.csv`. You may refer to GEP Data Ingest documantation for additional info.

An overview of the content is available in the next section.

1.6 Result file overview

1.6.1 Full-result csv file descriptor

The following table provides a brief explanation of the parameters included in the full result csv file.

Note that columsn 1-38 are the same with the `primary input file` presented in a previous section. Similarly, columns 39-58 include the results of the calibration process that remain the same for all scenarios. These are excluded here.

The remaining columns are informed by the electrification analysis per se. They repeat as many time as the intermediate years of the analysis. Therefore, these columns are presented only once below and are (usually) accompanied by the identifier <year>.

#	Column	Unit	Description
59	Pop<year>	people	Population in the specified year taking in to account the scenario parameteri
60	NewConnections<year>	people	Number of people to get electrified in the specified year; excludes populatio
61	NumPeoplePerHH	people	Number of people per type of household (different for urban/rural)
62	Tier	1 to 5	Reflective targeted “tier” of electrification for the settlement; based on para
63	EnergyPerSettlement<year>	kWh/year	Total electricity demand (to be covered) in the settlement in the specified ye
64	TotalEnergyPerCell	kWh/year	Total electricity demand in the settlement (already covered + to be covered)
65	MG_Hydro<year>	$/kWh	Levelized cost of electricity for mini-grid hydro power in the settlement in t
66	MG_PV<year>	$/kWh	Levelized cost of electricity for mini-grid PV in the settlement in the specifi
67	MG_Wind<year>	$/kWh	Levelized cost of electricity for mini-grid wind in the settlement in the spec
68	MG_Diesel<year>	$/kWh	Levelized cost of electricity for mini-grid diesel in the settlement in the spe
69	SA_Diesel<year>	$/kWh	Levelized cost of electricity for stand alone diesel in the settlement in the sp
70	SA_PV<year>	$/kWh	Levelized cost of electricity for stand alone PV in the settlement in the spec
71	Minimum_Tech_Off_grid<year>	name	Name & type of off-grid technology that provides the lowest lcoe in the spe
72	Minimum_LCOE_Off_grid<year>	$/kWh	Levelized cost of electricity of the off-grid technology that provides the low
73	Off_Grid_Code<year>	2 to 8	Code of electrifying off-grid technology in the specified year that provide th
74	Grid<year>	$/kWh	Levelized cost of electricity for grid in the settlement in the specified year; i

#	Column	Unit	Description
75	new_connections_household	kWh/year	Targeted electricity demand per household; new connections divided with n
76	MinimumOverall<year>	name	Name & type of technology that provides the lowest lcoe in the specified ye
77	MinimumOverallLCOE<year>	$/kWh	Levelized cost of electricity of the technology that provides the lowest lcoe
78	MinimumOverallCode<year>	1 to 8	Code of electrifying technology in the specified year that provide the lowest
79	InvestmentCost<year>	million $	Investment cost for electrification of the settlement in the specified year, by
80	InvestmentCostLV<year>	million $	Investment cost for LV line development in the settlement in the specified y
81	InvestmentCostMV<year>	million $	Investment cost for MV line development in the settlement in the specified y
82	InvestmentCostHV<year>	million $	Investment cost for HV line development in the settlement in the specified y
83	InvestmentCostTransformer<year>	million $	Investment cost for transformer development in the settlement in the specifi
84	InvestmentCostConnection<year>	million $	Investment associated with fee connection costs in the settlement in the spec
85	CapitalCapacityInvestment<year>	million $	Investment cost directly associated with capacity additions in the settlement
86	RecurringCosts<year>	million $	Aggregated recurring costs (O&M, fuel) in the settlement in the specified ye
87	ElecStatusIn<year>	0,1	Electrification status of the settlement in the specified year, before the appli
88	InvestmentCapita<year>	million $	Investment cost per capita (new connections) for electrification of the settle
89	GridElecIn<year>	0,1	Grid electricity status in the specified year; 0: non-electrified 1: electrified b
90	OffGridElecIn<year>	0,1	Off-grid electricity status in the specified year; 0: non-electrified 1: electrifi
91	FinalElecCode<base year>	1 to 8 or 99	Code of electrifying technology in the specified year (1: grid, 2: stand-alone
92	NewCapacity<year>	kW	Additional capacity required to electrify the settlement in the specified year

1.6.2 Summary csv file descriptor

Summary files provide aggregated results (at national level) of key findings from the electrification analysis. The following table provides a brief explanation of those.

Note: All investment costs are discounted to the base year of the analysis.

Variable	Data Type	Units	Description
Population<technology><year>	Integer	people	Total Population electrified by the specified <technology> in the specified <year>
NewConnections<technology><year>	Integer	people	Population that gains access through the specified <technology> in the specified <year> (not electrified before)
Capacity<technology><year>	Float	kW	Added Capacity to related to the specified <technology> in the specified <year>
Investment<technology><year>	Float	USD	Total investment required to deploy the specified <technology> in the specified <year>
InvestmentLV<technology><year>	Float	USD	Investment related to the development of LV lines required by specified <technology> in the specified <year> (for grid extension and mini-grids)
InvestmentMV<technology><year>	Float	USD	Investment related to the development of MV lines required by specified <technology> in the specified <year> (for grid extension and mini-grids)
InvestmentHV<technology><year>	Float	USD	Investment related to the development of HV lines required by specified <technology> in the specified <year> (for grid extension)
InvestmentTrans<technology><year>	Float	USD	Investment related to the development of sub-stations and/or transformers required by specified <technology> in the specified <year> (for grid extension and mini-grids)
InvestmentCon<technology><year>	Float	USD	Investment related to connection fees required by specified <technology> in the specified <year> (for grid extension and mini-grids)
InvestmentCap<technology><year>	Float	USD	Investment related (only) to the capacity additions required by specified <technology> in the specified <year>
RecurringCosts<technology><year>	Float	USD	Sum of recurring costs related to O&M and fuel for the specified <technology> in the specified <year>

1.7 Post-analysis & Visualization

1.7.1 Post-analysis module(s)

This section is under development. Once completed it aims to provide access and guiding instructions to scripts and/or modules that perform common post analysis on the **gep_onsset** electrification results.

Stay tuned for the updates!

1.7.2 Visualization

One of the benefits of geospatial analysis is that the results can be vizualized on maps thus convey information in a more direct and "digestible" way. In addition, one might leverage the spatial attributes of the results and in combination with other geo-coded information provide location based analytics.

Using GIS software

The **gep_onsset** full result file can be visualized in any GIS medium (QGIS, GeoPandas etc.) with the use of the coordinate `X_deg` and `Y_deg` columns. These columns provide the longitude and latitute of each population settlement in `WGS84 CRS - EPSG:4326`.

Using the GEP Explorer

The GEP Explorer is open source. That is, it can be configured and used locally and independently of the online server. You can find more information of how to configure it on your local machine at the explorer repository on GitHub.

1.8 Parameter Dictionary

The **gep_onsset** can get overwhelming for a new user - or even for experienced ones :) - due to the number of parameters, variable and naming conventions used. The following table aims to provide a brief explanation of common parameters one might encounter in the process.

Dataset	Data Type	Units	Description
Country	String		The official name of
Country code	String		Official country code
Start year	Integer	year	The year to be used a
End year	Integer	year	The final year of the
Discount Rate	Float	Percentage (8% used)	Rate used for discou
Settlement Area	Float	square kilometers (km2)	Size of the settlemen
Plant Capacity	Float	Kilowatts (kW)	Size of plant
Investment Cost	Float	USD/Kilowatts ($/kW)	Cost per unit of insta
Operation and Maintenance Cost	Float	Percentage	Percentage of Invest
Capacity Factor	Float	Ratio (0-1)	Ratio of actual electr
Expected Lifetime of Technology	Float	Years	Average predicted li
Base to Peak ratio	Float	Ratio (0-1)	Ratio used to estima
Fuel Price	Float	USD($) per liter	Average price of fue
Grid Electricity Generation Cost	Float	USD/kWh	Expected average lev
Grid Generation Capacity Cost	Float	USD/kW	Relative cost per cap
Expected Lifetime of Technology	Float	Years	Expected lifespan, w
Grid Extension Limit	Float	Kilometers	Maximum distance c
Capacity Factor	Float	Ratio (0-1)	Ratio of actual electr
Grid Penalty Ratio	Float	Ratio (0-1)	Obtained for each po
Grid Extension Cost Ratio	Float	Percentage	Percentage increase
Grid capacity ramp up limit	Float	MW	Limit of additional g
Base to Peak ratio	Float	Ratio (0-1)	Ratio used to estima
Expected Lifetime	Float	Years	Lifespan of transmis
HV line cost (108 kv)	Float	USD/Kilometer	Cost of 108 kv per k
HV line cost (69 kv)	Float	USD/Kilometer	Cost of 69 kv per kil
MV line cost (33 kv)	Float	USD/Kilometer	Cost of 33 kv per kil
MV line cost (11 kv)	Float	USD/Kilometer	Cost of 33 kv per kil
LV line cost (.4 kv)	Float	USD/Kilometer	Cost of .2 kv per kilo
LV line cost (.2 kv)	Float	USD/Kilometer	Cost of .2 kv per kilo
HV to MV substation (1000 kVA)	Float	USD/unit	Cost of substation pe
MV to MV substation (400 kVA)	Float	USD/unit	Cost of substation pe
Service transformer (50 kVA)	Float	USD/unit	Cost of substation pe
Max LV line length	Float	km	Maximum span of a
Load moment	Integer	kW m	Used in LV line dim
Maximun demand nodes served per transformer	Integer	nodes	Maximum number o
Additional Connection cost per household connected to the grid	Float	USD/Household	Additional Connecti
Additional Connection cost per household connected to the minigrid	Float	USD/Household	Additional Connecti
Transmission and Distribution Losses	Float	Percentage	Percentage of Total t

Table 4 – continued from pre

Dataset	Data Type	Units	Description
Operation and Maintenences costs of distribution	Float	Percentage	Percentage of total c
MV line maximum reach (extension limit)	Float	Kilometer	Maximum reach of M
Power factor	Float		The ratio of the activ
Population, start year	Integer	People	Population at the beg
Population, end year	Integer	People	Predicted population
Urban share, start year	Float	Percentage	Percentage of total p
Urban share, end year	Float	Percentage	Percentage of total p
Number of people/house (urban)	Float	People/household	Average household s
Number of people/house (rural)	Float	People/household	Average household s
Current National Electrification Rate	Float	Percentage	Percentage of electri
Current National Electrification Rate Rural	Float	Percentage	Percentage of rural e
Current National Electrification Rate Urban	Float	Percentage	Percentage of urban
Urban cut off value	Integer	people	Number of people ab
Urban ratio modelled	Float	Percentage	Urban ration identifi
Minimum night light value	Integer	(0-64)	Value that implies el
Minimum distance to service transformers	Float	km	Value that implies el
Maximum distance to roads	Float	km	Value that implies el
Maximum distance to the grid network	Float	km	Value that implies el
Population cut off value	Integer	people	Value that implies el
Electrification rate modelled	Float	Percentage	National electrificati
Urban electrification rate modelled	Float	Percentage	Urban electrification
Rural electrification rate modelled	Float	Percentage	Rural electrification

1.9 Contact

For any questions, feedback or general inquiries please to not hesitate to contact the development team.

The World Bank

- Benjamin Stewart <bstewart@worldbankgroup.org>
- Alexandros Korkovelos <akorkovelos@worldbank.org>

KTH dES

- Andreas Sahlberg <asahl@kth.se>
- Babak Khavari <khavari@kth.se>