LEAP-RE

Research and Innovation Action (RIA)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 963530

Start date : 2020-10-01 Duration : 63 Months
http://www.leap-re.eu/

Report of characterization of the electricity needs and resource assessment methodologies

Authors : Dr. Riccardo MEREU (POLIMI), António Couto, João Cardoso, Paula Costa, Teresa Simões (LNEG); Nicolò Stevanato, Ivan Sangiorgio, Riccardo Mereu (POLIMI); Davide Fioriti (UniPi); Sandra Banda (SU)

LEAP-RE - Contract Number: 963530

Project officer: Bernardo Luis ABELLO GARCIA

Document title	Report of characterization of the electricity needs and resource assessment methodologies
Author(s)	Dr. Riccardo MEREU, António Couto, João Cardoso, Paula Costa, Teresa Simões (LNEG); Nicolò Stevanato, Ivan Sangiorgio, Riccardo Mereu (POLIMI); Davide Fioriti (UniPi); Sandra Banda (SU)
Number of pages	82
Document type	Deliverable
Work Package	WP13
Document number	D13.2
Issued by	POLIMI
Date of completion	2022-12-01 07:48:45
Dissemination level	Error (13) !

Summary

Characterization of the electricity needs and resource assessment methodologies

Approval

Date	By
2022-12-01 07:49:07	Dr. Riccardo MEREU (POLIMI)
2022-12-01 11:00:07	Mr. Léonard LéVêQUE (LGI)

LEAP-RE

Long-Term Joint EU-AU Research and Innovation Partnership on Renewable Energy

Research & Innovation Action

Characterization of the electricity needs and resource assessment methodologies

November 2022

D13.2

Version N°1

Authors:

LNEG:
António Couto
João Cardoso
Paula Costa
Teresa Simões

PoliMI:
Nicolò Stevanato
Ivan Sangiorgio
Ricardo Mereu

UniPi:
Davide Fioriti

Strathmore:
Sandra Banda

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Disclaimer

The content of this report reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Document information

Grant Agreement	963530
Project Title	Long-Term Joint EU-AU Research and Innovation Partnership on Renewable Energy
Project Acronym	LEAP-RE
Project Coordinator	Vincent Chauvet (Vincent.chauvet@lgi-consulting.com) – LGI
Project Duration	1st October 2020 – 31st December 2025 (63 Months)
Related Work Package	WP13 – SETaDiSMA
Related Task(s)	Task 13.1
Lead Organisation	LNEG
Contributing Partner(s)	PoliMI, UniPi, Strathmore
Due Date	30/09/2022
Submission Date	29/11/2022
Dissemination level	Public

History

Date	Version	Submitted by	Reviewed by	Comments
29/11/2022	1	LNEG	PoliMI	First version of the deliverable.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Summary

The present report was developed as part of the research activities of the Sustainable Energy Transition and Digitalization of Smart Mini-Grids for Africa (SETaDiSMA) work package of the LEAP-RE project. SETaDiSMA aims to tackle the African mini-grid sector as a whole, addressing technological and energy planning challenges, digitalization research and development and related capacity building. The work to be developed focuses on case studies in Algeria, Kenya and Rwanda.

This report, part of SETaDiSMA's Task 13.1, System design and planning from social needs to technical design, constitutes deliverable *D13.2 - Characterization of the electricity needs and resource assessment methodologies,* identifies databases available to support SETaDiSMA's activities, as well as methodologies for demand estimation in the different contexts. A description of the mini-grids in the three countries of interest is also provided.

A reliable assessment of the primary resource for wind and solar power is essential to quantify the energy available to be converted at a particular site/region over a period of interest. In this work, a review of existing databases publicly available for performing a wind and/or solar resource assessment was performed. The databases identified can be split into three main groups according to the main source of information: i) ground-based stations; ii) numerical-based, and iii) satellite-based. In addition, databases with a combination of the different methods were also identified. For each database, the main features are provided.

For the countries of interest, the number of available ground stations with data suitable to the identify and/or validate the most suitable database for renewable energy (RE) resource estimation was limited: 7 with wind speed and direction - covering three stations in Algeria, two in Kenya and two in Mozambique, and 3 with global horizontal irradiance (GHI) and direct (or beam) normal irradiance – all in Kenya.

The results of the validation show that the wind database quality depends on the terrain and roughness types. This conclusion was expected since the analysed databases present a coarse spatial resolution being unable to represent quite well the wind behaviour over complex terrain or locals with medium/high roughness. In the case of solar databases, it was possible to benefit from another LEAP-RE work package (WP 10 - PURAMS) to identify the most suitable database for solar power estimation and, therefore, results are briefly presented in this report. The performance of solar databases, namely using satellite-based information, is high, being capable to provide hourly data with correlation values above 0.96 in the three stations analysed. While some static yearly or climatological resource assessment information is already publicly available (e.g., Global Wind Atlas), databases with hourly (or sub-hourly) data with high accuracy, as needed for dynamic studies of wind power integration, are still scarce for the African countries.

On the demand estimation side, an extensive literature review is conducted to assess the adopted techniques in rural electricity demand estimation in scientific literature. Thanks to this review the reader can have an understanding of the main socio-economic and cultural drivers that influence the electricity demand in rural areas.

Subsequently to the review, two different methodologies for demand estimation are proposed as possible novel approaches compared to the reviewed literature. First is hypothesized the creation of a public database for correlating already connected communities' appliance adoption patterns with the socio-economic and technical characteristics of the community and in this way estimate the possible appliance adoption pattern of the community under study, taking advantage of similarities in the socio-economic structure. Secondly a set of energy-modelling-ready archetypes of rural areas

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

users are developed, to serve as first-round approximated inputs in case of scarcity of data to formulate realistic load profiles for the study area.

The proposed archetypes are then validated against data collected from the studied mini-grids adopting a Mean Bias Error technique on the number of appliances predicted versus the observed data.

Keywords

Renewable energy databases; Wind resource assessment; Solar resource assessment; Validation procedure, Mini-grids, load assessment, load profile estimation.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Table of contents

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

List of figures

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

List of tables

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Abbreviations and Acronyms

Acronym	Description
ADEV	Absolute deviation
ADP	Automated Data Processing
ADPSFC	Global Surface Observational Weather Data
AI	Artificial intelligence
ANN	Artificial Neural Network
ANOVA	Analysis of variance
AS	April to September months
ASCII	American Standard Code for Information Interchange
AY	All Year Cooling
BSRN	Baseline Surface Radiation Network
CAMS	CAMS developed by the Copernicus Atmosphere Monitoring Service
CERES	Clouds and the Earth's Radiant Energy System
CFS	Climate Forecast System
CM SAF	Satellite Application Facility on Climate Monitoring
CSV	Comma separated values
DCs	Developed countries
DER-CAM	Distributed Energy Resources Customer Adoption
DHI	Difuse Horizontal Irradiance
DNI	Direct Normal Irradiance
ECMWF	European Centre for Meteorological Weather Forecasts
EICV	Integrated Household Living Conditions
EPMs	Energy planning models
ERA	European reanalysis
EU	European Union
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
GEBA	Global Energy Balance Archive
GFS	Global forecast system
GHI	Global Horizontal Irradiance
GIS	Geographical interface system
GMAO	Global Modeling and Assimilation Office
GOES	Geostationary Operational Environmental Satellite
GPS	Global Positioning System
GTI	Global Tilted Irradiance
GTS	Global Telecommunications System
HadISST	Hadley Centre Sea Ice and Sea Surface Temperature
HOMER	HOMER - Hybrid Renewable and Distributed Generation System Design Software
IHOGA	Improved Hybrid Optimization by Genetic Algorithm
IPPs	Independent Power Producers
JRA	Japanese reanalysis
JRC	Joint Research Centre
JSON	JavaScript Object Notation
KPLC	Kenya Power and Lighting Company
LASSO	Least Absolute Shrinkage and Selection Operator
MACC-RAD	Monitoring Atmospheric Composition and Climate
MAGICSOL	Mesoscale Atmospheric Global Irradiance Code Solar
MBE	Mean bias error
MERRA	Modern-Era Retrospective analysis for Research and Applications

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

METAR	Meteorological Terminal Air Report
M-LED	Multi-sectoral Latent Electricity Demand
MLP	Multi-Layer Perceptron
MSG	Meteosat Second Generation
MVIRI	Meteosat Visible and Infrared Imager
NASA	National Aeronautics and Space Administration
NC	No Cooling Days
NCEP	National Center for Environmental Prediction
NOAA	National Oceanic and Atmospheric Administration
NREL	National Renewable Energy Laboratory
NSRDB	National Solar Radiation Database
NWP	Numerical Weather Prediction
OLS	Ordinary least squares
OM	Cooling Days from October to March
OSeMOSYS	Open Source Energy Modeling System
POWER	Prediction Of WorldwidE Resources
PURAMS	Rural African Markets using Standalone Solar
PV	Photovoltaic
PVGIS	PhotoVoltaic Geographical Information System
RAD	Radiation
RAMP	Remote-Areas Multi-energy systems load profiles model
RE	Renewable energy
REG	Rwanda Energy Group
REREC	Rural Electrification and Renewable Energy Corporation
RES	Renewable energy sources
RMSE	Root mean square error
RTTOV11	Radiative Transfer for TOVS
SARAH	SurfAce solar RAdiation data set Heliosat
SASSCAL	Climate Change and Adaptive Land Management
SAURAN	Southern African universities radiometric network
SDG	Sustainable development goal
SEVIRI	Spinning Enhanced Visible and InfraRed Imager
SHS	Solar home system
SOM	Self-Organizing Map
SPECMAGIC	Spectrally Resolved Mesoscale Atmospheric Global Irradiance Code
SREP	Scaling-Up Renewable Energy Programme
SSA	Sub-Saharan Africa
TAHMO	Trans-African Hydro-Meteorological Observatory
TMY	Typical Meteorological Year
TOVS	Operational Vertical Sounder
UTC	Universal Time
WMO	World meteorological organization
WRDC	World Radiation Data Centre
WRMC	World radiation monitoring centre
WT	Wealth tier

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

1. Introduction

The LEAP-RE project aims to establish a long-term partnership between African and European stakeholders in the field of renewable energy. Among its planned actions are a set of eight internal research and innovation projects. One of those projects is the Sustainable Energy Transition and Digitalization of Smart Mini-Grids for Africa (SETaDiSMA) project (LEAP-RE WP13). SETaDiSMA addresses the use of renewable energy sources for mini-grid applications in Africa, considering both the repowering of existing conventional mini-grids (brown-fields) and the electrification of new communities (green-fields), focusing on three countries: Algeria, Kenya and Rwanda.

The present report was developed as part of the research activities of the WP 13 SETaDiSMA project, Task 13.1 – System design and planning from social needs to technical design, corresponding to the project deliverable D13.2. It presents methodologies and tools to perform the characterization of the electricity needs and renewable energy resources assessment to support the design and deployment of renewable-powered (including hybrid) mini-grids in Africa.

The seventh sustainable development goal (SDG) defined by the United Nation aims to "*ensure access to affordable, reliable, sustainable and modern energy for all*"[1] (Jayachandran et al., 2022). Guaranteeing access to modern clean energy options will impact the life of billions of persons due to the new economic opportunities and jobs, citizens empowerment, better education, better health and financial services, and more sustainable and inclusive communities, and inclusive, while contributing to climate change mitigation actions (ESMAP, 2022; USAID, African Union, & Power Africa, 2021). The "Agenda 2063" released by the African Union (Africa Union Comission, 2015) established that an adequate electricity infrastructure at different spatial levels from continental to local levels, including rural areas is crucial for the socioeconomic development of African countries. Despite the improvements observed in the latest years, population without access to electricity is still high in Africa, especially in Sub-Saharan countries, Figure 1.

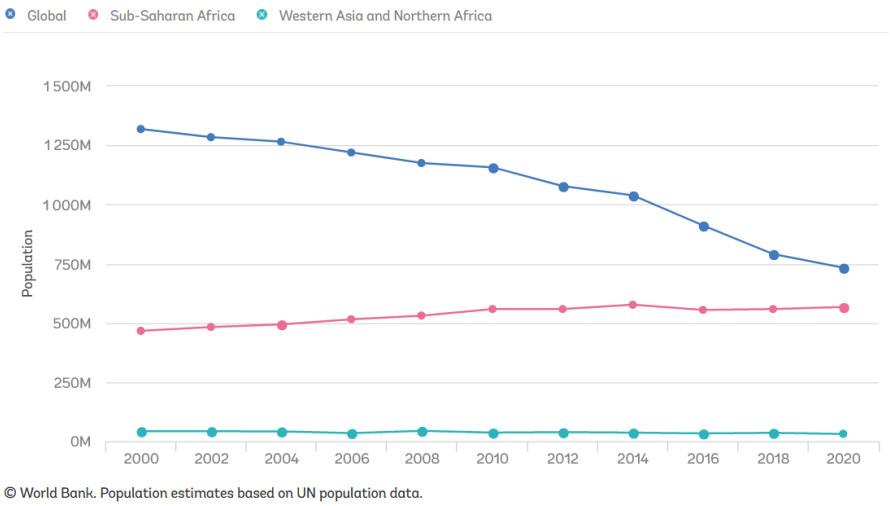


Figure 1 - Population without access to electricity, millions of people (Total). Source: https://trackingsdg7.esmap.org

[1] More details available at: https://www.un.org/en/chronicle/article/goal-7-ensure-access-affordable-reliable-sustainable-and-modern-energy-all

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

According to United States Agency for International Development (USAID et al., 2021), sixty percent of Africans live in rural areas. Of these, only around 5% have access to modern electricity options[2]. These numbers are mainly explained by the high dispersion of end-users, low economic activity limiting the ability of populations to support the costs, distance from the national network and low population density (USAID et al., 2021; Zebra, van der Windt, Nhumaio, & Faaij, 2021). With the rapid technological development, mini-grids have become a practical solution to the challenge of electrifying rural areas, being a sustainable alternative to the expansion of the national grid. Thus, in many cases, mini-grids are the most efficient and economical way of providing access to energy, since they can i) be "easily" installed, ii) be flexible and modular, being adjusted according to the electricity demand, and iii) be connected to the main national grid if it expands (Zebra et al., 2021). Although mini-grids are not a completely new phenomenon, the technologies that they use improved significantly since 2012 becoming a very interesting solution for the electrification of rural zones (SEforALL & BloombergNEF, 2019).

By definition, mini-grids include generation systems (using conventional, renewable energy and storage technologies) and distribution of electrical energy suitable to supply electricity to only a few spatially dispersed/remote customers or bring energy to a large number of people in a predetermined area. Typically, mini-grids involve only small-scale electricity generation that ranges from 10 kW to 10MW. A mini-grid can be composed of different generation technologies and are capable of operating independently from the national main grid (SEforALL & BloombergNEF, 2019; Zebra et al., 2021). Initially, most mini-grids depended on conventional energy sources such as fossil fuels (diesel and kerosene) or small hydroelectric power plants. With the strong cost reduction observed in recent years (in opposition with fossil fuels whose prices are on a continuous rise) mini-grids are adopting time-variable renewable energy sources (vRES) technologies in their generation mix (Mbinkar, A. Asoh, Tchuidjan, & Baldeh, 2021). Solar power, in specific, is the leading technology installed in the last years in mini-grid projects (Figure 2) due to its modular and relatively easy to install in remote areas features, solar power.

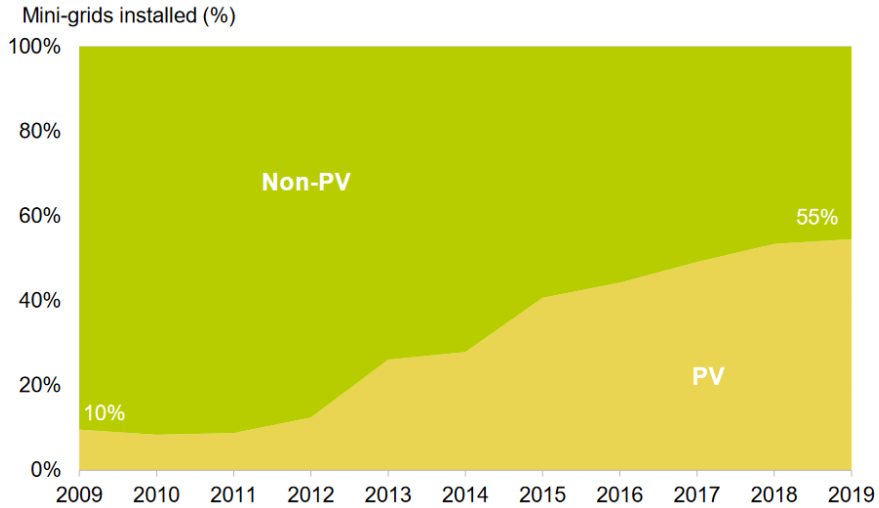


Figure 2 – Evolution of weight of solar PV power in the installed capacity of the mini girds (image extracted from (SEforALL & BloombergNEF, 2019)).

Unlike conventional power plants, vRES (mainly wind and solar power) are weather dependent and, therefore, have high temporal variability, bringing additional challenges to a safe and reliable power system. To partially accommodate this variability and always

[2] Modern electricity options refer to the energy-based from sources obtained through commercialized market channels rather than the traditional biomass with low energy value.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

satisfy the electricity supply/demand equilibrium, battery systems and/or fossil fuel generators are used as a backup. Therefore, within a mini-grid, a reliable assessment of the vRES resource, as well as the expected demand, is crucial for sizing the different components of the system, allowing to reduce initial and operational costs while contributing to reach SDG #7, i.e., ensuring access to affordable, reliable and sustainable energy for all.

- **Overview of mini-grids in Kenya**

Traditionally, mini-grids in Kenya were diesel-powered and run by the national utility, the Kenya Power and Lighting Company (KPLC). Since 2011, several diesel powered mini-grids have been transformed into hybrid systems that have an additional solar or wind power component, while several new mini-grids, powered exclusively by renewable energy technologies, have been deployed. Further hybridisation of existing mini-grids is planned for the near future. The development of hybrid mini-grids is one of Kenya's important projects within the Scaling-Up Renewable Energy Programme (SREP)[3]. Hybridisation with renewables has had positive impacts in reducing generation costs. Research indicates that northern Kenya is attractive for mini-grids with large penetration of solar energy (>85%). The use of pre-paid meters, smart meters and remote monitoring, has positively impacted the performance of mini-grids in Kenya. There are about 400 mini-grids in Kenya with 282 of them run by KPLC.

Around two-thirds of the Kenyan population live in the fertile southern part of the country and can be reached by national grid extension. The remaining one-third is spread across the arid and semi-arid Northern and North-Eastern areas of the country, which are sparsely populated and therefore expensive to connect to the national grid. One of the solutions for electricity access is decentralized electricity generation and distribution networks, including off-grid solutions and mini-grids.

For the SETaDiSMA activities, the following Kenyan mini-grids are studied:

Eliye (Illiye) Spring mini-grid station located in Turkana Central, north west of Kenya was commissioned in 2019; constructed by Rural Electrification and Renewable Energy Corporation (REREC) and operated by KPLC. The mini-grid is of hybrid technology with an installed capacity of 120 kW using both diesel and solar PV. The solar PV is the predominantly used technology with an installed capacity of 80 kW. The diesel generator has a capacity of 5 kVA operating on a power factor of 0.8.

Figure 3 – Picture of the Eliye (Illiye) Spring mini grid station.

[3] More details available at: https://www.worldbank.org/en/country/armenia/brief/srep

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Kakuma mini-grid station in Turkana, Kakuma town, is a purely diesel operated mini-grid owned by REREC and operated by KPLC. It has an installed capacity of 1.2 MW with 3 generators, each with an installed capacity of 500 kVA.

Figure 4 – Pictures of the Kakuma mini grid station.

Lokichogio mini-grid station in Turkana County is a diesel operated mini-grid that was constructed by REREC and is currently being run by KPLC. Two diesel generators have been installed with a combined capacity of 640 kW which serve around 614 users; both commercial and residential.

Figure 5 – Pictures of the Lokichogio mini grid station.

Longech mini-grid in Turkana county is a privately owned hybrid mini-grid that runs on solar PV and diesel generators with a combined capacity of 67 kW. The installed capacity for the solar mini-grid is 50 kW while the diesel generator is of 22 kVA.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Figure 6 – Pictures of the Longech mini grid.

Powerhive mini-grid sites in Kisii and Nyamira counties use solar photovoltaic technology in their operation. Each mini-grid can be connected to 120 households within a 1 km radius.

Faza station in Lamu county, is a diesel operated mini-grid owned by REREC and operated by KPLC with a capacity of 2 030 kW with 6 000 users dependent on it.

Gakira Micro Hydropower in Nyeri county, Mathira town, is a private hydro-powered mini-grid with a capacity of 600 kW.

- **Overview of mini-grids in Rwanda**

According to Rwanda Energy Group (REG) the number of households accessing electricity has increased from 10% in 2010 to 73% as of May 2022 with the target of being 100% electrified by 2024. Some areas especially those located in Kigali city have currently the highest rate with 99% while the lowest rate is above 50%. Rwanda topography/mountainous nature is a hindrance to electricity access with respect to national grid extensions. To overcome this geographic limitation there is an increasing need for mini-grids development. According to the National electrification plan revision of 2021, the national grid is expected to contribute with 89.9% and with the remaining 10.1% coming from off-grid and mini-grid systems (Table 1).

Table 1 - Electricity plan by 2024 (National Electricity plan 2021).

Province	Electrification Technology per villages						Grand Total
	GE & Fill-In	%	Microgrid	%	SAS	%	
North	2,439	88,9%	24	0.9%	281	10.2%	2,744
South	2,993	85.5%	62	1.8%	446	12.7%	3,501
East	3,410	89.9%	80	2.1%	301	8%	3,791
West	3,309	91.5%	16	0.4%	292	8.1%	3,617
Kigali City	1,163	100.0%		0.0%		0.0%	1,163
Grand Total	**13,293**	**89.7%**	**203**	**1.4%**	**1,320**	**8.9%**	**14,816**

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

By the end of 2019, mini-grids connected 3 236 households across Rwanda, with 84 mini-grids installed with a total capacity of around 250 kW but managing to produce only 182 kW. Currently, off-grid systems are operated by private companies, however, considering the renewable electricity potential in Rwanda this sector is underutilized.

Rwanda is generally characterized by the Savannah climate and its geographical location endows it with sufficient solar radiation intensity for the deployment solar systems, with approximately 5 kWh/m^2/day and peak sun hours of approximately 5 hours per day. Solar energy is a promising solution to meet the demand for rural households' electricity services in remote locations. As of May 2021, 16% of Rwandan households are accessing electricity through off-grid systems, mainly solar photovoltaic systems.

Rwanda's hydropower sector showed tremendous progress. Overall installed capacity of power is about 221.1 MW, with hydropower contributing approximately 46% of it. This was achieved by involving private investors in the energy sector as Independent Power Producers (IPPs).

Briefly, Rwanda possesses a conductive legal and regulatory environment for private investors to invest in renewable energy. Currently, the elaboration and development of 20 green mini-grid feasibility studies and roll out plans are being carried out and will be handed over to the private sector to increase the private sector contribution in energy generation. Government of Rwanda (GoR) will lease out these sites to private investors to better operate, maintain and connect them to the off-grid.

For the SETaDiSMA study, the 8 mini-grids will be studied in Rwanda, being the first selection: "Rutenderi AC Solar Mini-grid"; Rushonga AC Solar Mini-grid; Mudasomwa PHPP; Banda AC Solar Mini-grid; Rutobotobo and Rwamacumu AC solar Mini-grid; Nyankorogoma PHPP Ltd; Nyakiramba PHPP; Nyakiramba PHPP.

Aiming to support the SETaDiSMA activities, this report is structured as follows: Section 2 presents a review of existing databases for estimating the wind and solar resource assessment and provides the results obtained to identify the most adequate database for the different renewable technologies; Section 3 provides also a literature review to identify the most common techniques for estimate the rural electricity demand. Two different methodologies for demand estimation are proposed and validated against data collected from the studied mini-grids; and, finally, in section 4, some final remarks are provided.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

2. Resource assessment

Resource assessment is a key step in the design process of hybrid mini-grids in rural areas. It consists in the evaluation of the availability (both time-wise and energy-wise) of energy resources at a given location. It can be performed by means of direct ground measurements, however, this approach may not always be suitable and/or feasible worldwide. Due to seasonal and multi-annual variability of renewable energy sources such as solar and wind, resource assessment requires long-term data. Hence the need of exploiting meteorological datasets, from which data related to solar irradiance, temperature and wind speed may be derived and employed to evaluate the energy production potential of a technology for a given location.

The resource assessment for renewable energy (RE) technologies entails the characterization of the primary resource of these technologies (wind speed in the case of wind power and solar irradiance in the case of solar power), which is available for energy conversion at a specific location or region over a period of interest. In the case of the SETaDiSMA project, the resource assessment will support the activities of renewable energy sources integration into mini-grids to increase the share of solar and wind power into this type of system, as well as to allow its decarbonization.

For solar energy applications, the focus will be on the different components of the solar irradiance relevant to the solar sector (Sengupta, Habte, Wilbert, Gueymard, & Remund, 2021):

- Global horizontal irradiance (GHI): the power density of the solar radiation received on a horizontal surface. This component is the sum of the direct and diffuse components of solar radiation.
- Direct (or beam) normal irradiance (DNI): the power density of solar radiation incident on a surface perpendicular to the sun's rays emanating from the solar disk and the circumsolar region.
- Diffuse horizontal irradiance (DHI): the power density of solar radiation scattered and reflected by the air, clouds and other particles in the atmosphere, incident on a horizontal surface.
- Global tilted irradiance (GTI): the power density of solar radiation, including radiation emanating from the solar disk, the circumsolar region, the scattered radiation and the reflected radiation incident on a tilted surface.

In Figure 7 a schematic representation of the relevant components of solar global irradiance is provided.

In the case of wind power, the resource assessment focuses on the wind speed and direction due to the cubic dependency of the generated power from this parameter (equation 1). The wind direction is also included in the description of the existing databases since this information is crucial for optimizing the wind turbine layouts in wind parks.

$$P = \frac{1}{2}\rho A C_p v^3 \qquad (1)$$

In equation (1), P is wind power, ρ is the air density, A is the rotor area of a wind turbine, C_p is the power coefficient and *v* is the wind speed. In addition, some databases already provide the resource data as power (or energy) without being necessary to apply any conversion from the primary resource data. These databases were also included in this report.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

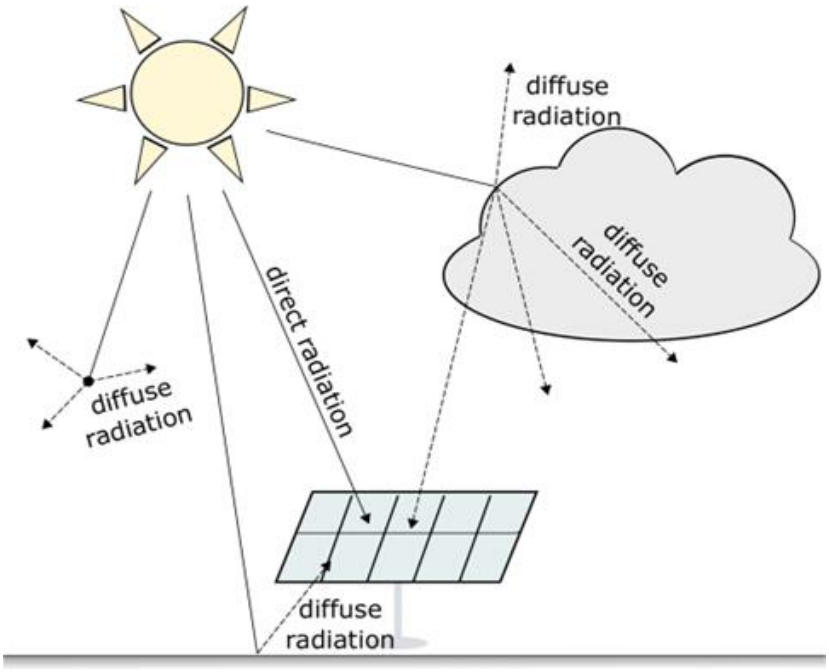


Figure 7 - Global solar radiation components: direct (or beam), diffuse, and reflected radiation incident on a tilted solar photovoltaic (PV) panel (image extracted from (Doddy Clarke & Sweeney, 2022)).

2.1 Existing databases for RE (wind and solar) resource estimation

This section presents and briefly characterizes the main open-access databases available for wind and solar resource estimation. Other relevant databases, with access under data purchase or paid subscription, can be found in existing literature such as (Sengupta et al., 2021). Some well-known open-access data platforms, such as the Global Wind Atlas (available at https://globalwindatlas.info) or the Global Solar Atlas (available at https://globalsolaratlas.info), are not included in this document since the temporal resolution of the data is not suitable for mini-grid applications beyond simplified pre-feasibility studies. Other tasks, e.g. the mini-grid design, require dynamical studies which demand data with short time intervals (one hour or less).

2.1.1 Ground measured datasets

The number of ground stations available with measured data is very important for the resource assessment since they provide data with high degree of accuracy and high temporal resolution. However, it has high maintenance requirements and its spatial resolution is very limited, since the data is only representative for the station location. Ground measured data is also required for validation of reanalysis or satellite derived datasets (even if the available data is only daily or monthly averages).

This type of dataset consists of data collected for a specific location using ground stations. Pyrheliometers and pyranometers are the two most common types of radiometers used to measure solar irradiance. In the case of wind power, anemometers and wind vanes are the types of sensors used to collect wind speed and direction data, respectively. For some specific cases/applications, a LiDAR (Light Detection And Ranging) can also be used. This type of system enables to infer the vertical wind speed and direction profiles from 10 meters up to 500 meters above ground level. The installation and maintenance of these types of equipment are quite expensive and therefore, typically, the time series for characterization of the resource are of short duration (except for climatological stations of meteorological institutes/airports).

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

A special care should be taken when using measured data since not all databases perform data quality assessments of their records. Several methodologies for quality control of solar irradiance and wind measurement data are available in the literature (see for example (Brower, Bailey, Doane, & Eberhard, 2012; Marques, Páscoa, Carvalho, & Cardoso, 2020; Sengupta et al., 2021).

Figure 8 - Sensors used for solar irradiance measurements left top - pyranometer for global irradiance measurements; left bottom - two-axis tracked pyrheliometer for direct normal irradiance measurements; and right - pyranometer with shading ball for diffuse irradiance measurements (Saad, M. Amen, Refaat, & Morad, 2012).

- ***World Radiation Monitoring Center - Baseline Surface Radiation Network***

The Baseline Surface Radiation Network (BSRN) provides high quality observations for short and long-wave surface radiation fluxes. BSRN observations have a global spatial coverage, coming from a set of ground stations (49 by the end of 2021) located across the world in different climatic zones, Figure 9. Three of these stations are located in the African continent: Tamanrasset in Algeria; Gobabeb in Namibia; De Aar in South Africa. Date coverage varies from station to station, starting from 1992 till the present date. The number and type of measurements vary by station, however, in all cases, the basic irradiance parameters are given, including GHI, DNI and GHI with temporal resolution of 1 minute (3 minutes resolution for stations before 2009). The data acquired through the BSRN is managed by the World Radiation Monitoring Center (WRMC) and can be accessed online at https://bsrn.awi.de.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

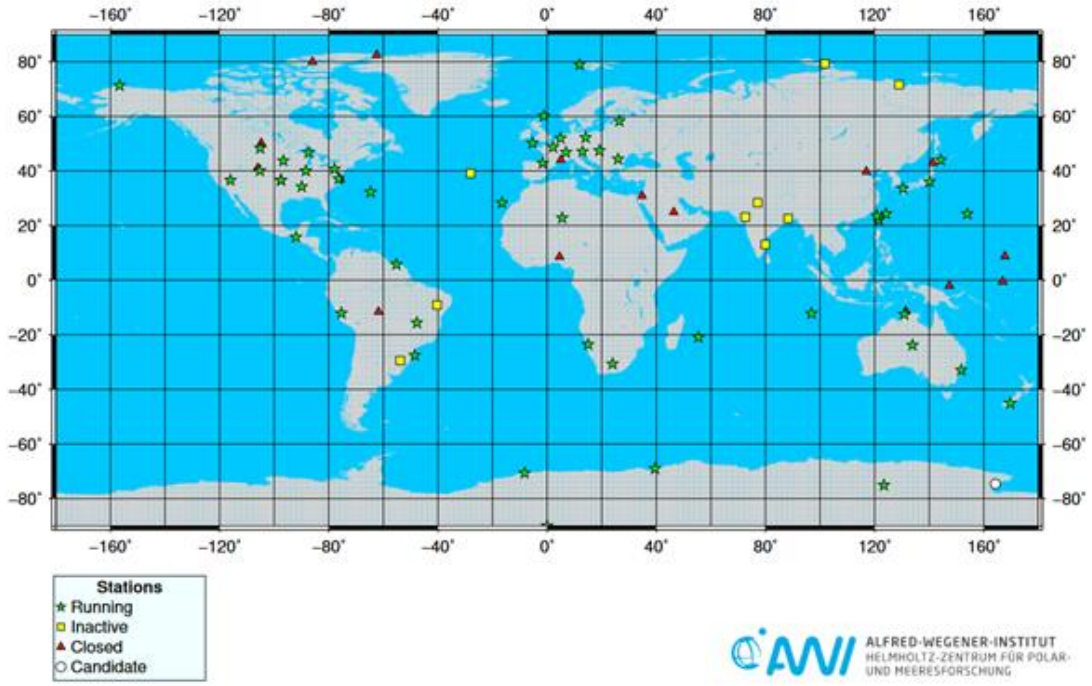


Figure 9 - Location of the BSRN stations (December 2021). Source: https://bsrn.awi.de

- ***World Meteorological Organization – World Radiation Data Centre***

Since its inception in 1964, the World Radiation Data Centre (WRDC), an initiative under the scope of the World Meteorological Organization, collects, archives and publishes radiometric data for several meteorological stations, having a global spatial coverage, Figure 10. The WRDC currently processes and publishes solar radiation data collected by more than 500 stations across 56 countries and has a historical archive covering more than 1200 stations[4]. Some of these stations are located in African countries[5] and from those, very few are currently feeding data to this archive – in the latest reports available covering a full year (2020) only an Algerian station (Tamanrasset) can be found. Considering the last five years, only in 2017 is possible to find other active stations located in an African country (in Egypt and Morocco). Data coverage varies from year to year and from station to station, being available from 1964 onwards (notice there is an approximately one year gap between the current date and the latest date available in the WRDC archive). The

[4] Further details available at: https://community.wmo.int/world-radiation-data-centre

[5] Covering less than 60% of the African countries: Algeria, Angola, Burkina Faso, Cape Verde, Central African Republic, Djibouti, Egypt, Ethiopia, Gambia, Ghana, Guinea-Bissau, Kenya, Madagascar, Mali, Morocco, Mozambique, Namibia, Niger, Nigeria, Republic of South Africa, S. Tomé and Príncipe, Senegal, Sudan, Tanzania, Tunisia, Uganda, Zaire, Zambia and Zimbabwe.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

number and type of measurements vary by station, presenting global irradiation data for all stations and diffuse irradiation for some of them[6]. The data is presented as daily and monthly totals. Monthly means of hourly totals are also available for some of the stations. WRDC data archives can be accessed online at http://wrdc.mgo.rssi.ru.

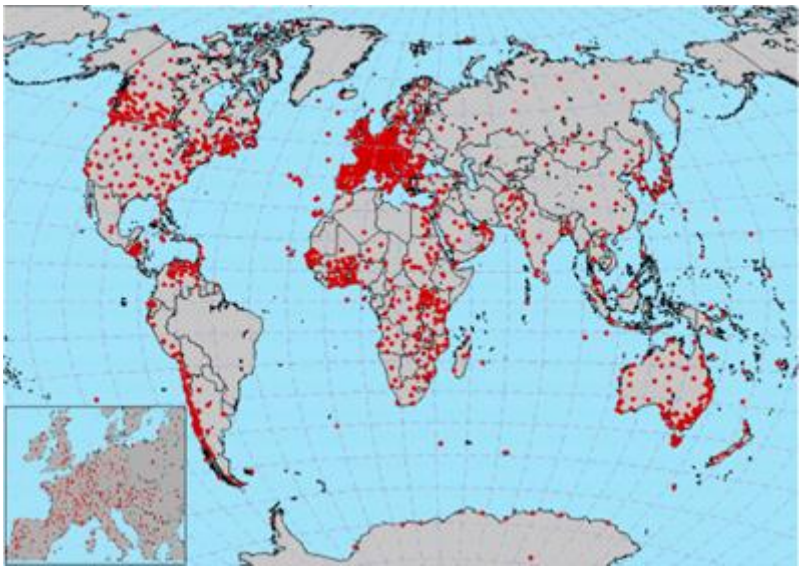


Figure 10 - Location of the stations with data present in the WRDC archives. Source: http://wrdc.mgo.rssi.ru/wrdc_en_new.htm

- ***Global Energy Balance Archive***

The Global Energy Balance Archive (GEBA) (Wild et al., 2017) is a database curated by ETH Zurich that contains surface energy flux measurement data (including global, direct and diffuse irradiance) for climatological applications. The 2017 version of the GEBA database includes data from 2500 different locations, with 1155 locations having more than 3 years of data, Figure 11. Some ground stations included in the GEBA overlap with other datasets (e.g., with the BSRN or WRDC). Data from several African locations (particularly from sub-Saharan Africa) are included in this database. However, data availability varies significantly from year to year and from station to station. Global irradiance data is available for 2249 locations, being the oldest measurement from 1919. Data for the diffuse and direct irradiance is scarcer, comprising 787 and 109 sites, respectively. The data are presented as monthly average values and are available online at https://geba.ethz.ch.

[6] Other parameters available, although less relevant for the resource assessment, include the sunshine duration and radiation balance.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

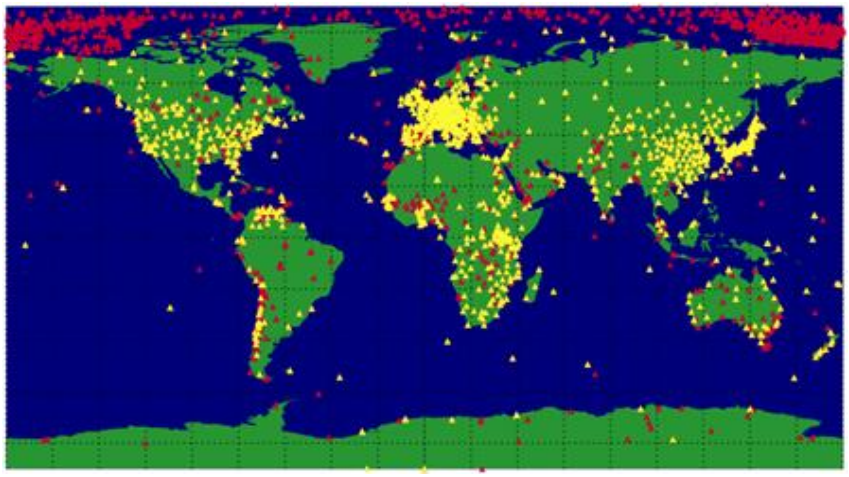


Figure 11 – Location of the GEBA measurement stations: red symbols represent locations with at least one monthly entry; yellow symbols indicate locations with at least 3 years of data (image extracted from (Wild et al., 2017)).

- ***Southern African Universities Radiometric Network***

The Southern African Universities Radiometric Network (SAURAN) is an initiative of the Stellenbosch University and the University of KwaZulu-Natal, established in 2014, that gathers 16 African, European and North American partners. SAURAN's dataset includes data for a total of 24 stations located in Botswana, Namibia and South Africa (including offline stations), Figure 12. Currently, operational stations are only located in Namibia and South Africa. Data availability varies from station to station both in terms of timespan and measurands. For example, the earliest data start is from 2010 (Stellenbosch University station) while the latest is from 2022 (University of KwaZulu-Natal, Pietermaritzburg station). All stations present measured data for direct normal irradiance and global horizontal irradiance. Moreover, two stations present data for diffuse horizontal irradiance and some stations also present data for other meteorological parameters such as wind speed and direction, ambient temperature, relative humidity, etc. The data are presented as daily, hourly and minute averages and are available online at https://sauran.ac.za/.

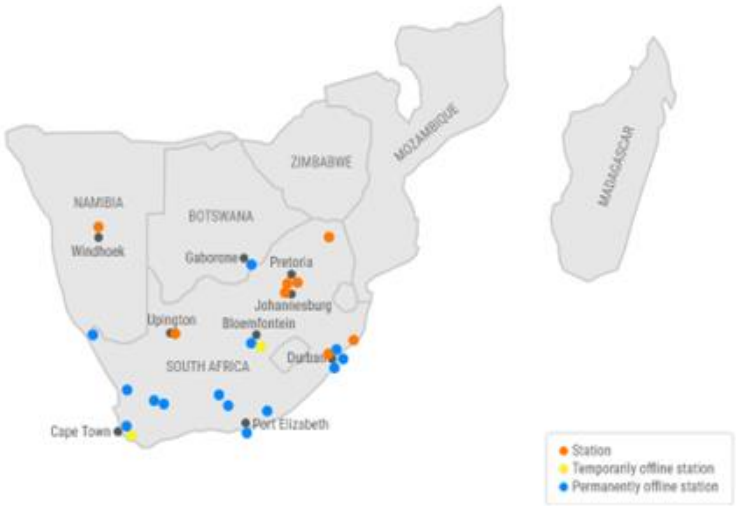


Figure 12 - Location of the SAURAN stations. Source: https://sauran.ac.za/

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

- ***Trans-African Hydro-Meteorological Observatory***

The Trans-African Hydro-Meteorological Observatory (TAHMO) initiative aims to develop a large scale network of low-cost meteorological stations (up to 20 000) in sub-Saharan African countries, gathering European, North American and African partners. Currently, the network gathers meteorological data, including solar global irradiance, wind speed and wind direction, in 23 African countries, Figure 13. The data timespan varies from station to station, and can have high temporal resolution of up to one minute. TAHMO data can be accessed at https://tahmo.org/, being available for free for governmental and scientific use. However, a fee is incurred for commercial applications. Despite the general interest of such initiative, currently, this dataset is of limited use for renewable energy resource assessment, since both solar irradiance and wind measurements have not been suitably validated against reference measurements in African settings (Schunke et al., 2021).

Figure 13 - Location of the TAHMO stations. Source: https://tahmo.org

- ***ENERGYDATA.INFO - World Bank Group***

The ENERGYDATA.INFO is an online open data platform (https://energydata.info) maintained by the World Bank Group that gathers energy related datasets and analytics, including data from solar irradiance and wind measurements at African ground stations, Figure 14. This repository has data pertaining to solar irradiance measurements (direct normal, global horizontal and diffuse horizontal), wind measurements and other meteorological parameters (e.g., ambient temperature, relative humidity, etc.) measurements for weather stations located in Benin, Kenya, Liberia, Malawi, Mali, Senegal, Tanzania and Zambia. The temporal resolution of this data is 1 minute and its timespan is variable, depending on both country and station location, being the earliest data from 2015. The wind data also varies from station to station in terms of height of measurement, with data acquired at 3 m for Kenya, Senegal and Tanzania and 10 m for the others. This data repository also has a wind measurement dataset for 17 Ethiopian sites with a 10 minute resolution for the years 2018 till 2020.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

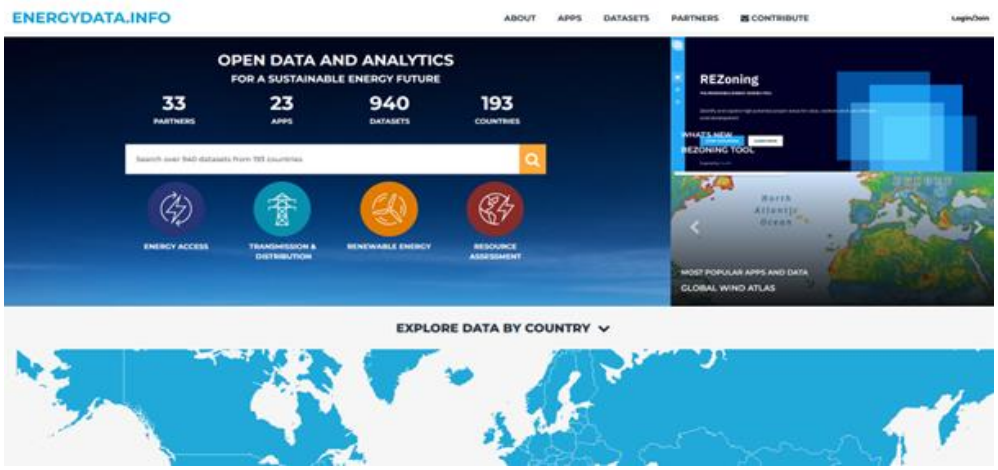


Figure 14 – Entrance portal for the ENERGYDATA.INFO platform. Source: https://energydata.info

- ***Southern African Science Service Centre for Climate Change and Adaptive Land Management WeatherNET***

The Southern African Science Service Centre for Climate Change and Adaptive Land Management (SASSCAL) WeatherNET is a network of meteorological stations resulting from a joint initiative of Angola, Botswana, Namibia, South Africa, Zambia and Germany. The corresponding database covers 164 locations across the 5 African countries, being 96 of the stations offline for more than one month or unavailable on the website, Figure 15. These stations collect parameters as global horizontal irradiance and wind speed and direction measurements. Other meteorological variables are also provided in this dataset, such as ambient temperature, relative humidity, etc. The timespan of the available data varies from station to station, being the oldest data from 2009. The data are presented as hourly, daily and monthly average values and are available online at http://www.sasscalweathernet.org and http://data.sasscal.org.

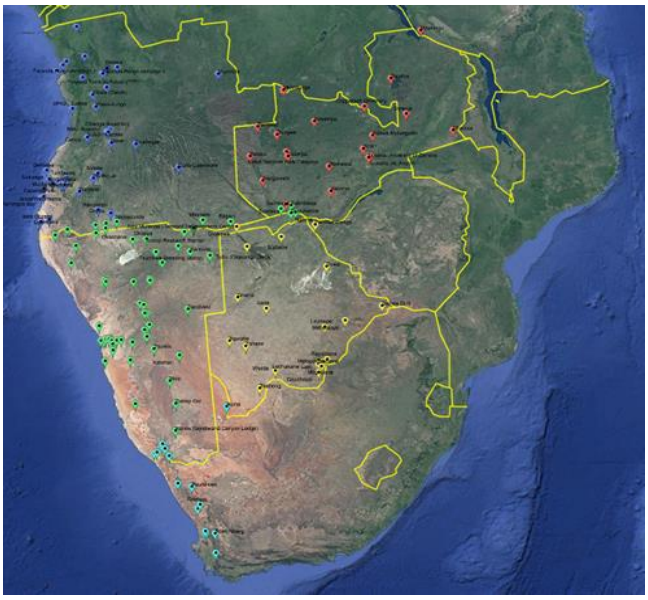


Figure 15 - Location of the SASSCAL WeatherNet stations. Source: http://www.sasscalweathernet.org

- ***NCEP ADP Global Surface Observational Weather Data, October 1999 – continuing (RDA database)***

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Automated Data Processing (ADP) Global Surface Observational Weather Data (ADPSFC) database is based on weather surface reports and it is hosted by the National Centers for Environmental Prediction (NCEP). This database covers the entire globe (Figure 16) and includes data from i) land stations, e.g., land synoptic stations (fixed and mobile), aviation (METAR), land reports; and ii) marine surface reports from Ships, Drifting and moored buoys transmitted through the global telecommunications system (GTS). These stations collect parameters such as air temperature, surface pressure, wind direction and speed. The timespan of the available data varies from station to station, being the oldest data from 1999. The temporal resolution also varies among the stations ranging from 30 minutes to 3 hours. Although the data are based on stations from meteorological institutes and used to support air traffic users should perform rigorous quality control checks before using these data since anomalous behaviours (e.g., frozen records) and erroneous values.

The data from this database are publicly available at: https://rda.ucar.edu/datasets/ds461.0 (the website requires registration). On this website, data can be extracted for a specific location[7] or for a region of interest[8].

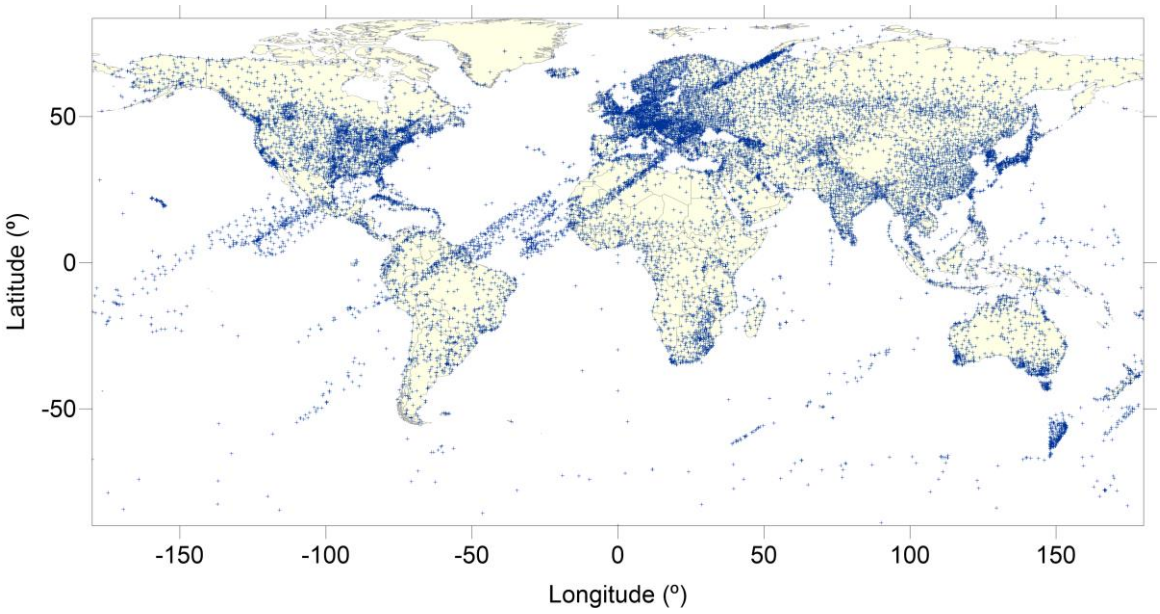


Figure 16 - Location with measured data in ADPSFC database from 1999 to 2022 (blue points).

2.1.2 Meteorological Reanalyses (Numerical model database)

[7] See database website for further details and use the following link to identify the ground-based stations: https://oscar.wmo.int/surface

[8] Within the scope of LEAP-RE project some scripts were developed to obtain the wind speed and direction data, apply some basic control and quality check procedures and sort by station. These scripts have been made publicly available to allow access to project partners as well as other interested parties (see further details at: https://github.com/AntCouto/LEAP-RE).

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Reanalyses combine meteorological observations and numerical weather prediction models to obtain a long-term retrospective estimation of many atmospheric and climatic variables such as humidity, clouds composition, temperature, wind and radiation properties in a uniform spatial grid with a certain resolution. Regarding wind energy applications, reanalyses are the only tool which can provide past wind speed estimates for a long time span and at different heights.

This type of dataset consists of a blend of numerical weather prediction (NWP) models with data (gathered from ground stations, buoys, weather balloons, airplanes, etc.) assimilation schemes. NWP models parameterize and simulate the atmosphere and its circulation mechanisms, ensuring the atmospheric dynamic consistency, while the assimilation schemes keep the model close to the real conditions, compensating the deviations associated with the model physics.

Reanalysis data have a relatively low spatial resolution, which is a drawback for accurately representing some local effects. Currently, ERA5 from the European Centre for Medium-Range Weather Forecasts (ECMWF) provides the highest resolution with a spatial grid of 0.28º x 0.28º (latitude x longitude). The temporal resolution varies according to the reanalysis database ranging from one to six hours. The main advantages of using reanalysis to perform renewable energy (RE) resource assessment are: uniformed grid data for the entire world, temporally consistent without missing data, and long-term period (in some cases with more than 30 years). The accuracy of this type of dataset strongly relies on the amount and quality of the observations used as initial conditions as well as in the development of the physical formulations and parameterization.

Currently, several reanalysis databases are publicly available providing meteorological data (wind speed and direction, cloud cover, etc.). Below a brief description of these databases is provided. Most of the data is available in the Netcdf format, which can be read using the Panoply software from NASA/GISS[9]. It should be mentioned that other examples of databases with reanalysis are also available, such as: the NCEP-R1 produced and released by National Centre for Environmental Prediction (NCEP), later replaced by NCEP-R2 (NCEP/DOE, 2000), the ERA-40 or ERA-Interim (Berrisford et al., 2011) produced and released by European Centre for Medium-Range Weather Forecasts (ECMWF), the JRA-25 produced and released by Japanese Meteorological Agency, and CAMS developed by the Copernicus Atmosphere Monitoring Service (CAMS). Analysis datasets[10] as the NCEP Global Forecast System (NCEP-GFS) (National Centers for Environmental Prediction NOAA, U.S. Department of Commerce, 2015) and the NCEP Final Analysis (NCEP-FNL) (National Centers for Environmental Prediction NOAA, 2000) are provided with spatial and temporal characteristics similar to reanalysis and are also publicly available. However, all the previous datasets have coarser temporal resolutions (which vary from three to six hours) and are not suitable for the evaluation of the RE resource in the context of mini-grids applications. Typically, these reanalyses and analysis datasets are used as initial and boundary conditions in limited area/mesoscale models, allowing these models to perform simulations with high temporal and spatial resolutions.

[9] Software available at: https://www.giss.nasa.gov/tools/panoply

[10] Analysis datasets consist of observations on an irregular grid to produce a representation of the atmospheric state over a regular grid. See further details at: https://rda.ucar.edu/datasets/ds083.2/docs/Analysis.pdf

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

- ***ECMWF-ERA5***

ERA5 is the fifth generation of global climate atmospheric reanalysis produced by the ECMWF replacing the ERA-15, ERA-40 and more recently the ERA-Interim product widely used in several studies. Compared to ERA-Interim, improvements in ERA-5 include, e.g., the use of data from global sea ice and sea surface temperature analyzes (HadISST.2), reprocessed climate data from ECMWF and implementation of the Radiative Transfer for TOVS 11 (RTTOV11) radiative transfer model.

ERA5 covers the period from January 1950 providing information until near real-time. This database provided information with a spatial resolution of 0.25º x 0.25º (approximate resolution of 31 kilometers). Atmospheric data are simulated for 37 pressure levels (same levels as in ERA-Interim). Surface or single level above ground data is also available, such as meridional and longitudinal wind speed at various levels (10, 100 meters), shortwave radiation, etc. The data from this database are publicly available at: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form or https://rda.ucar.edu/datasets/ds633.0/#!description (both websites require registration).

- ***NASA-MERRA-2***

Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) is developed by the NASA's Global Modeling and Assimilation Office (GMAO) (Bosilovich et al., 2015). MERRA-2 represents the GMAO's commitment to making climate data available continuously and in near real-time. This database incorporates observation types not available in its predecessor, MERRA, as well as advances made in the assimilation system that allow the use of modern observations of hyperspectral and microwave radiation, and other data types.

MERRA-2 provides global atmospheric data from 1980 until near real-time with hourly resolution and a spatial resolution of 0.50º x 0.63º (latitude x longitude). Atmospheric data are simulated for 72 pressure levels. Surface or single level above ground data is also available, such as meridional and longitudinal wind speed at different levels (10, 50 meters), cloud cover, incident shortwave radiation, etc. The data are publicly available at: https://disc.gsfc.nasa.gov/ (website requires registration).

- ***NCEP-CFSv2***

The Climate Forecast System, version 2 (CFSv2) has been developed at NCEP. It is an integrally coupled model capable of representing the atmosphere's interactions as well as the interactions of the atmosphere with the land surface, ocean, and sea ice. This version became operational at NCEP in March 2011, replacing the previous version - Climate Forecast System Reanalysis (CFSRv1) implemented in 2004. At the base of this model are the improved versions of the NCEP models, namely, the well-known Global project Reanalysis 2 and Global Forecast System (GFS). Thus, compared to previous NCEP reanalyses, CFSv2 provides, among others, finer spatial and temporal resolution, an improved model, advanced assimilation schemes capable to use satellite data and atmosphere-ocean-land-ice coupling (Saha et al., 2010, 2011).

CFS v2 provides global atmospheric data from 2011 and is initialized four times a day (00, 06, 12 and 18 UTC). The original spatial resolution is 0.5º x 0.5º. Surface or single level above ground data is also available, such as meridional and longitudinal wind speed at different levels (10 meters), downward shortwave radiation flux, total cloud cover, etc.

The data from this database are publicly available at https://rda.ucar.edu/datasets/ds094.1. Data from this database can be obtained in Netcdf or comma-separated values (CSV) files. In this website, data are available with hourly

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

resolution and with horizontal resolutions of 0.2º, 0.5º, 1.0º and 2.5º. This hourly resolution data can be obtained by combining 1) the analysis time step and the remaining 5 hours with the forecast results, or 2) all forecast hours from each initialization.

2.1.3 Satellite-based dataset

This type of dataset can be obtained using meteorological satellites that:

a) use sensing imagery useful to derive the influence of clouds on solar radiation on the Earth's surface or for estimating the wind velocity and direction at cloud level by analyzing the positions of a distinguishable cloud/group of clouds between two sequential images. In general, for this purpose, geostationary satellites are used. They have a high time resolution of 15 to 30 minutes, reasonable spatial resolution (few km) and their orbit is nearly 36 000 km away over the equator. They circle the earth at the same speed as the earth's rotation, such that they are always above the same location.

b) use dedicated equipment/sensors such as imager radiometer and vertical sensors. The data acquisition process involves interactions between the incident radiation and the target of interest (e.g., ocean waves). The energy used in remote sensing is electromagnetic radiation, whose most important characteristics are the frequency and wavelength of the radiation. Active sensors emit energy and measure the properties of the signal that returns to these instruments after being absorbed, reflected or scattered across the surface of the imaged target. This intrinsic feature only allows to perform the resource assessment for offshore regions using these sensors, which are regions outside the scope of the SETaDiSMA activities. For this purpose, polar- orbiting satellites are used. They have low time resolution (1-2 images per day) and a global coverage. This type of satellite has higher spatial resolution compared to geostationary satellites, orbiting at an altitude of near 800 km.

Currently, several satellite-based databases are publicly available providing mainly data for the solar energy sector. Below a brief description of these databases is provided.

- ***CAMS Radiation service***

Copernicus Atmosphere Monitoring Service radiation service (CAMS-RAD), is part of the Copernicus Programme, an Earth observation programme coordinated and managed by the European Commission in partnership with the European Space Agency. CAMS combines satellite-based irradiance data with a NWP model to provide optical variables as well as solar radiation parameters. It is based in the Heliosat-4 method, integrated into the MACC—RAD service (Hoyer-Klick C., Lefèvre, Schroedter-Homscheidt, & Wald, 2015). This method is capable to treat separately the clear sky radiation (CAMS McClear method) and the effects of clouds (CAMS McCloud method) using images acquired by the Meteosat Second Generation (MSG) satellite. In Figure 17, a schematic diagram of the CAMS radiation service methodology is depicted.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

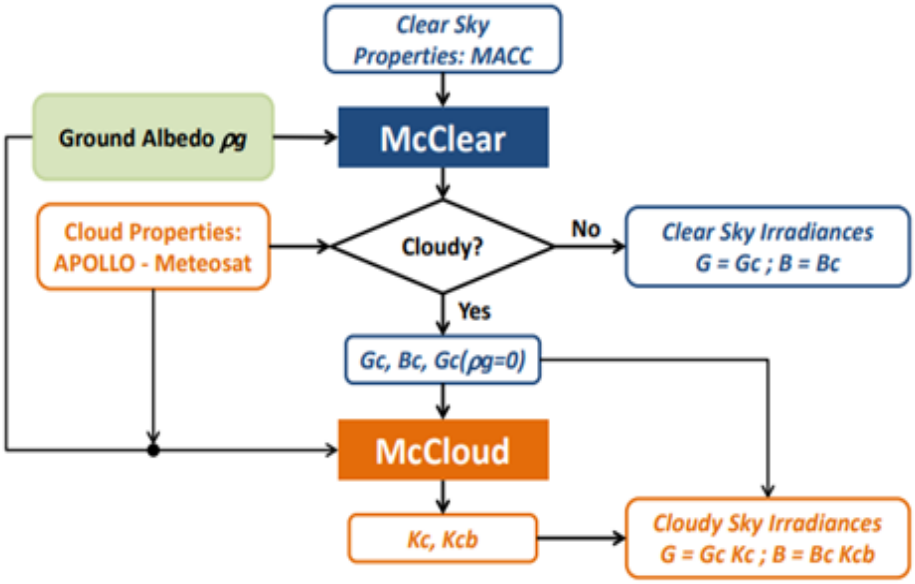


Figure 17 – Schematic diagram of the CAMS radiation service methodology. Adapted from (Blanc & Wald, 2015).

CAMS-RAD provides GHI, DHI, and DNI data from 2005 up to 2 days ago, with resolution ranging from 1 min to 1 month. The spatial coverage of this dataset is -66º to 66º (both in latitude and longitude) and data can be extracted for a specific location of interest, Figure 18. Registration is needed to download the data and each user can perform up to 100 requests per day. Data are available in CSV and Netcdf format at: https://www.soda-pro.com/web-services/radiation/cams-radiation-service

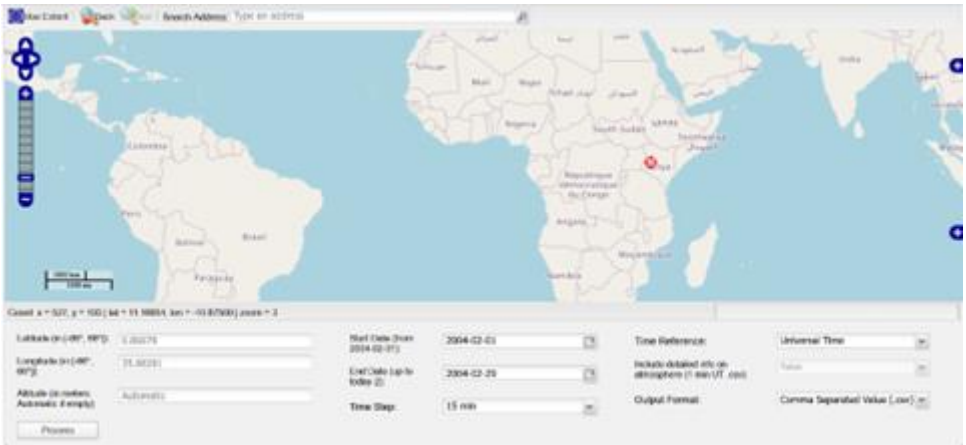


Figure 18 - Entrance portal for the CAMS radiation database. Source: https://www.soda-pro.com/web-services/radiation/cams-radiation-service

- ***SARAH-2***

The Surface solar RAdiation data set-Heliosat, Edition 2 (SARAH-2) has been developed by the Satellite Application Facility on Climate Monitoring (CM SAF) - a center established by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) in 2000, Figure 19. SARAH-2 data are derived from the observations of the visible channels of the Meteosat Visible and Infrared Imager (MVIRI) and the Spinning Enhanced Visible and InfraRed Imager (SEVIRI) instruments onboard the MSG satellites. SARAH-2 is

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

produced using the Mesoscale Atmospheric Global Irradiance Code Solar (MAGICSOL) method, which is a combination of the well-established Heliosat method (Hammer et al., 2003) with the SPECMAGIC clear-sky model (Mueller, Behrendt, Hammer, & Kemper, 2012). The aerosol information used in SARAH-2 is climatological, lacking inter-annual variation.

It contains data from global and direct radiation, sunshine duration and cloud albedo. The spatial coverage of this dataset is -66º to 66º (both in latitude and longitude) with a spatial resolution of 0.05º ×0.05º. Data are available between 1983 and 2017 as monthly and daily means and as 30 minutes instantaneous time resolution. Registration is needed to download the data, which are available in Netcdf format at: https://wui.cmsaf.eu/safira/action/viewProduktDetails?eid=21832_22008&fid=28

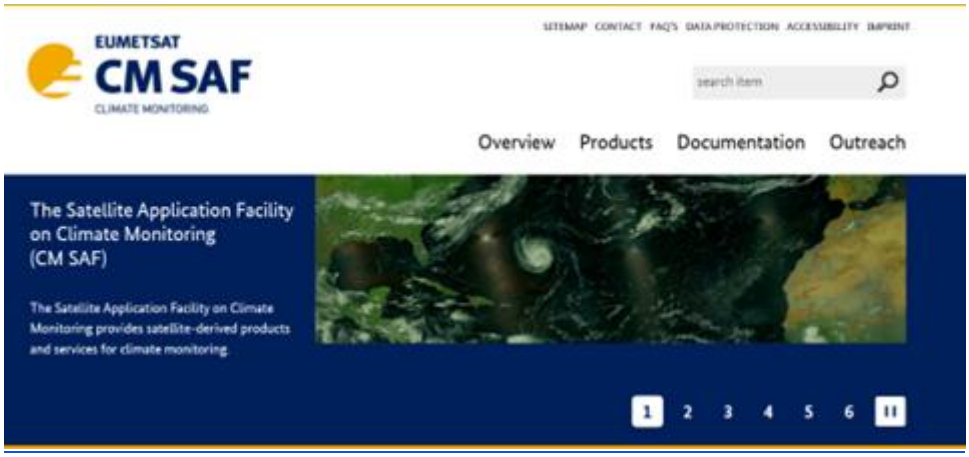


Figure 19 - Entrance portal for the CMSAF database. Source: https://www.cmsaf.eu

- ***CERES***

The Synoptic Fluxes and Clouds (SYN1deg) is a product based on Clouds and the Earth's Radiant Energy System (CERES) satellite instruments that is being developed by NASA's Earth Observing System program. CERES's instruments are capable to measure directly the emission of thermal infrared radiation to space and the reflected solar radiation across all wavelengths between the ultraviolet and infrared (Doelling et al., 2013). SYN1deg is a level-3 product, which comprises average data of fluxes and clouds. This product comprises information from several sources, namely, the Geostationary Operational Environmental Satellite (GOES) and HIMAWARI-8 satellite (Figure 20).

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

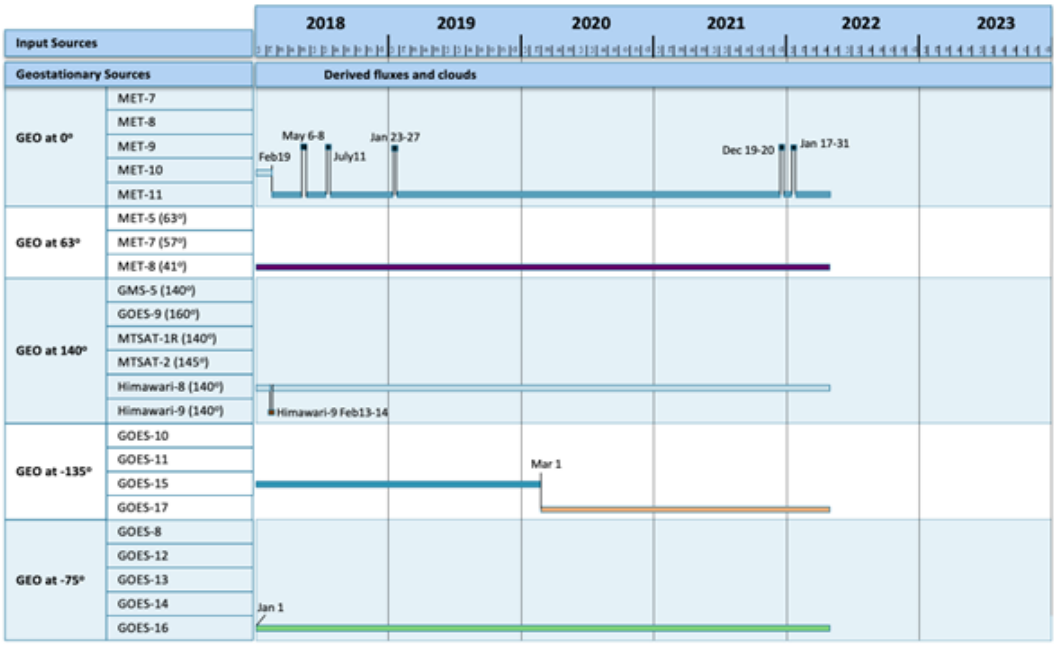


Figure 20 - Satellites used in the CERES-SYN1deg database according the longitude. Source: https://ceres.larc.nasa.gov/data/general-product-info

CERES-SYN1deg benefits from the use of different satellites to provide a product that covers the entire globe with a spatial resolution of 1º ×1º. This database provides data from several parameters such as the top-of-atmosphere and surface radiative fluxes as well as fluxes at four atmospheric pressure levels (70, 200, 500, and 850 hPa) with the following temporal resolution: monthly hourly, daily, 3-hourly and hourly. Data are available since 2000 to date (with a latency of 5 months). A registration is needed to download the data, which are available in Netcdf format at: https://ceres.larc.nasa.gov/data/#syn1deg-level-3

2.1.4 Web platforms and tools

Some web platforms and tools provide data considering hybrid solutions (e.g., combining satellite and reanalysis data, combining satellite data with interpolation of data from ground stations) or even additional processing in relation to the original data source. Below a list of these platforms and tools is presented, considering exclusively the ones that provide free data for the African continent suitable for mini-grid applications.

- ***NASA POWER***

Prediction Of Worldwide Energy Resources (POWER), developed by NASA, provides solar and meteorological data at global scale. Solar data, such as solar irradiance and cloud properties, are based on the satellite based products NASA GEWEX (for data from 1984 to 2001), NASA CERES SYN1deg (2001-few months within real-time) and CERES FLASHFlux Version 4A, which guarantees low latency data (about 4 days of near real time). Meteorological data like wind speed, wind direction and temperature are instead based upon MERRA-2 (from 1979 to few months within real time) and Geostationary Operational Environmental Satellite (GOES) 5.12.4, a weather prediction model that allows to obtain

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

data with a latency of 2 days up to near real-time. POWER time-series are available at a minimum temporal resolution of 1 hour for the time period 2001- real time, while daily, monthly and annual averages are available for the entire temporal coverage. The spatial resolution for meteorological and solar parameters is 0.50º x 0.63º (MERRA-2) and 1° lat × 1° lon (GEWEX, CERES). Wind speed data can be retrieved at any desired height between 10 and 300 meters, since MERRA-2 native data at 10 and 50 meters are used to interpolate wind speed with a power law for several types of terrains. Data are publicly available at https://power.larc.nasa.gov/ in several formats (NetCDF, ASCII, JSON, CSV).

- ***National Solar Radiation Database***

The National Solar Radiation Database (NSRDB) is a collaborative effort of several North American institutions led by NREL that provides solar irradiance (global horizontal, diffuse horizontal and direct normal components, including for clear sky) and meteorological data (such as wind speed and direction, ambient temperature, relative humidity, etc.). NSRDB data is generated through modelling of multi-channel measurements performed by geostationary satellites (see Figure 21).

The present version of NSRDB provides data for the African continent for the years 2017 to 2019 with a spatial resolution of 4 km. The user can select up to three different temporal resolutions: 15, 30 and 60 minutes data. Data can be retrieved directly from the web platform at https:// nsrdb.nrel.gov/ or through an application (see for example https://developer.nrel.gov/ docs/solar/nsrdb/nsrdb_data_query).

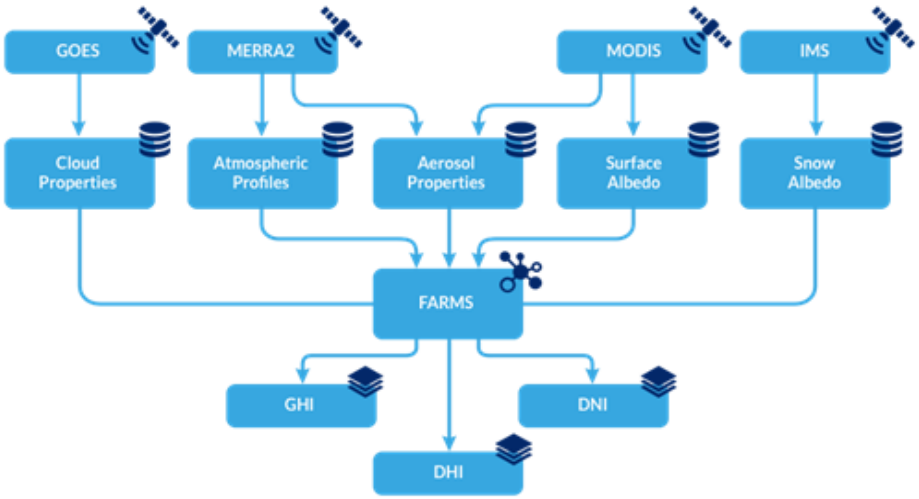


Figure 21 – Data flow used by NSRDB for solar irradiance calculations. Source: https://nsrdb.nrel.gov

- ***Renewables.ninja***

The Renewables.ninja web application results from a collaborative effort of Stefan Pfenninger and Iain Staffell. It uses weather data from the NASA MERRA reanalysis (entire globe) and CM-SAF's SARAH satellite (only cover Europe and North Africa) data and converts it to power output data through the Global Solar Energy Estimator model (Pfenninger & Staffell, 2016) and the Virtual Wind Farm model (Staffell & Pfenninger, 2016). This tool has global coverage and generates data with a temporal resolution of one hour, providing information regarding solar irradiance (direct and diffuse horizontal

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

components), wind speed, ambient temperature and electric power production[11]. The dataset generated with this tool covers the years 2000 till 2018. Data can be retrieved through the web platform at https://www.renewables.ninja (Figure 22).

Figure 22 - Entrance portal for the Renewables.ninja web platform. Source: https://www.renewables.ninja

- ***Photovoltaic Geographical information System***

The Photovoltaic Geographical information System (PVGIS) is a free web application developed by the European Commission Joint Research Centre (JRC) that uses satellite and reanalysis databases to provide data on solar radiation and photovoltaic energy production. The spatial resolution for African locations is approximately 5 km (considering the use of the default database for Africa: the PVGIS-SARAH2), Figure 23. Different types of datasets can be obtained through PVGIS, including monthly average solar irradiation, average daily irradiance (global horizontal, diffuse horizontal and direct normal) and ambient temperature profiles for specific months, hourly irradiance (global, diffuse and reflected) data. The available data timespan covers the years 2005 till 2020. PVGIS also allows the generation of Typical Meteorological Year (TMY) datasets, which include solar irradiance (global horizontal, diffuse horizontal, and direct normal), wind speed and direction and ambient temperature, relative humidity, air pressure and downwards infra-red radiation. Data can be retrieved through the web platform at https://re.jrc.ec.europa.eu/pvg_tools/en.

[11] To estimate the electric power production the user needs to input the capacity of the PV system or the wind turbine.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

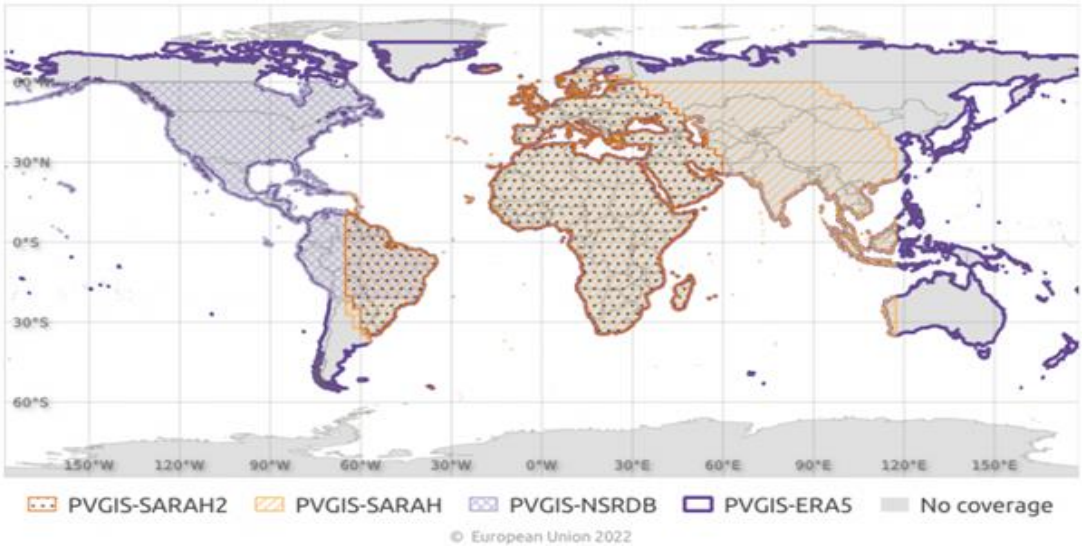


Figure 23 – Coverage of PVGIS solar radiation databases. Source: https://joint-research-centre.ec.europa.eu/pvgis-photovoltaic-geographical-information-system/getting-started-pvgis/pvgis-user-manual_en

2.2 Observed data

Collecting observed data is always a crucial step for validating a RE database. Regarding observed data within the SETaDiSMA work package, a form was established and distributed among the partners to identify data that can be used. Three partners indicated data that could be used, but, until the submission of this deliverable, it was not possible to work with such data. Nevertheless, if the data identified become available within the SETaDiSMA activities timeframe, upcoming deliverables will include a similar analysis as the one presented below.

Although Mozambique does not belong to SETaDiSMA group countries, LNEG had authorization to use, exclusively for validation purposes, data from two anemometric masts operating in Mozambique territory during the years 2012 to 2013. This observational data covers one complete year of data with a 10 minute time resolution, which is very adequate for validation of the wind RE databases. Without access to other observational data from the partner countries, several public databases described in sections 2.1.1 to 2.1.3 were consulted to identify available ground stations with solar or wind data suitable for validation and identification of the most adequate RE database for use in SETaDiSMA studies. In terms of solar data, only three ground-base stations located in Kenya were identified and the validation study of these stations is extensively presented in the Deliverable 10.2 (Couto et al., 2022) of the PURAMS project, which was also developed within the scope of LEAP-RE. Therefore, in section 2.3.2 only the main findings of the solar validation procedure are presented.

In terms of wind studies, the database ADPSFC presents several meteorological stations with data for the countries under analyzed in SETaDiSMA, Figure 24. In this figure, the local of all identified meteorological stations with wind data is plotted and the wind data availability (in %) is highlighted using a color scale.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

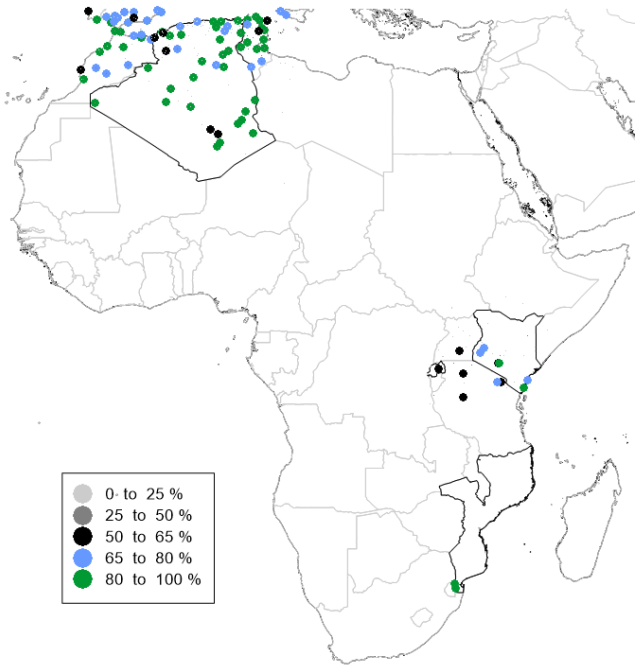


Figure 24 – Public ground-based stations in LEAP-RE countries with wind data and their hourly data availability (in %) (except for Mozambique whose data is not public).

Figure 24 shows a great number of observational wind data stations for Algeria. Most of these stations have good data availability for a timespan of three years. Data measurements were performed at 10 meters height and are provided with a one hour time step. For this country, the three stations with data availability above 90% are considered to validate the RE databases.

For Kenya, only a few meteorological stations are available and only two of the identified stations have data availability above 80% of the time for a timespan of three years of data and time steps of 1 hour.

For Rwanda only one station is identified but their data availability and data quality are very poor to be used in validation studies. Therefore, this station was rejected.

Apart from the previous countries, wind data from Mozambique was used by LNEG. These data were available in the LNEG database and come from two stations with one complete year of data with data intervals of 10 min. Both stations operated in the South part of Mozambique and have 100% data availability.

In addition to the data availability, stations were also chosen according to their localization in the territories aiming to cover different aspects of the orography and roughness such as the mountainous areas, coastal areas or mainland areas. This will elucidate how RE databases are adequate to provide wind data under different conditions. The behaviour of the wind flow depends strongly on the complexity of the orography and surface roughness elements as well as the wind phenomena propagated by atmospheric turbulence and stratification of the atmosphere inside the boundary layer. A well description of wind flow behaviour is still a challenge in the most up-to-date atmospheric or general circulation atmospheric models. Therefore, the outcome of the public RE databases validation can infer how the RE databases generated by general atmospheric models are or not effective in describing the wind phenomena at the mountainous or coastal or mainland areas which are evaluated by this deliverable.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

2.3 Validation of RE databases with locally measured data

In this section, the validation results of the RE wind databases according to observed data identified in section 2.2 are presented. The validation covers three stations in Algeria, two in Kenya and two in Mozambique[12]. Two main RE databases (section 2.1.2) were used for validation: ECMWF-ERA5 and NASA-MERRA2. For the case of the NCEP-CFSv2, due to a long maintenance since summer for updating/substitution the main servers of this database, extraction was not definitely possible to be done or when was possible, data came always with several "*NaN*" errors or data faults or even data with erroneous values. For this reason and at the present time, this database is unsuitable to be used for validation under the development of this report.

The data extraction from the ECMWF-ERA5 was performed for two level heights, 10 m and 100m while from NASA-MERRA2 data was extracted for 10 m and 50m heights. For the case of NASA-MERRA2 the extraction can be bi-linear interpolated spatially to the exact location of the station while the ECMWF-ERA5 the extraction is only possible for the nearest grid-point close to the location of the observational station.

The two stations in Mozambique operated at 40 m and 60 m above ground, which is totally different from the height of meteorological synoptic stations (10 m). The RE renewable databases do not provide data for the 40 m or 60 m heights. Therefore, and for this particular case of Mozambique, the validation on the two stations will be performed for the height of 40 m above ground (the lowest level available). The wind speed is estimated for each record using the following power law equation:

$$\frac{v_2}{v_1} = \left(\frac{h_2}{h_1}\right)^{\alpha} \tag{2}$$

where v_2 is the wind speed from RE database level most closely above 40m height being $h_2$that level and $v_1$the velocity most closely below 40m height and h_1 that height. The wind shear exponent, α is given by:

$$\alpha = \frac{\ln\left(\frac{v_2}{v_1}\right)}{\ln\left(\frac{h_2}{h_1}\right)} \tag{3}$$

For the ECMWF-MERRA5, α is given by:

$$\alpha = \frac{\ln\left(\frac{v_{100m}}{v_{10m}}\right)}{\ln\left(\frac{100}{10}\right)} \tag{4}$$

while for NASA-MERRA2. α is given by:

[12] Mozambique does not belong to SETaDiSMA project although is a partner in other LEAP-RE projects and is geographically located near to countries under analysed in this project.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

$$\alpha = \frac{\ln\left(\frac{v_{50m}}{v_{10m}}\right)}{\ln\left(\frac{50}{10}\right)} \tag{5}$$

After obtaining the wind shear exponent of each location, the wind speed data at 10 m of each RE database can be used to estimate the value of wind speed at 40 m through the power law:

$$v_{40m} = v_{10m}\left(\frac{40}{10}\right)^{\alpha} \tag{6}$$

In terms of wind direction, no estimation was done since the wind direction typically does not vary significantly with the height. Therefore, it is considered that: in the case of NASA-MERRA2 wind direction at 40 m height is the same as at the 50 m height level; in the case of ECMWF-ERA5 the wind direction at 40 m has the same values as at the level of 10 m.

The validation process is performed using hourly resolution data for both wind speed and wind direction. For the wind speed case, the most common statistical parameters used by the wind industry are calculated: the Pearson-Correlation coefficient, the BIAS, the absolute deviation and the root mean square error (RMSE). For the wind direction case, the absolute deviation and RMSE are computed.

Several graphs are also presented. They are the most commonly used graphs produced for validation purposes by the wind industry, which are the daily average wind speed profile, the annual wind speed profile, the wind speed distribution, the scatterplot of the wind speeds and the wind rose.

- **Procedure for validation of RE databases with local measured data**

The following equations describe the validation parameters used for validation which are, the BIAS, the absolute deviation (ADEV) and the RMSE:

$$\mathrm{BIAS} = \frac{1}{n}\sum_{i=1}^{n}\left(x_m(i) - x_o(i)\right) \tag{7}$$

$$\mathrm{ADEV} = \frac{1}{n}\sum_{i=1}^{n}\left|x_m(i) - x_o(i)\right| \tag{8}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_m(i) - x_o(i)\right)^2} \tag{9}$$

where n is the number of data pairs, x_m is the modelled value from the RE database and x_o is the observed/measured data. The Pearson Correlation coefficient is defined according to:

$$\mathrm{CC} = \frac{\sum_{i=1}^{n}(x_m(i) - \overline{x_m}) \cdot (x_o(i) - \overline{x_o})}{\sqrt{\sum_{i=1}^{n}(x_m(i) - \overline{x_m})^2 \cdot (x_o(i) - \overline{x_o})^2}} \tag{10}$$

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

with

$$\overline{x_o} = \frac{1}{n}\sum_{i=1}^{n} x_o(i) \tag{11}$$

and

$$\overline{x_m} = \frac{1}{n}\sum_{i=1}^{n} x_m(i) \tag{12}$$

The Pearson correlation coefficient gives real values between -1 and +1. Values near 0 mean no correlation/values totally in disagreement. Values near +1 means observation and RE modelling database are in good agreement over time while near −1 means they still are in agreement but in the opposite phase. In the following sub-chapters, the results obtained for Algeria, Kenya and Mozambique are presented.

2.3.1 Validation of databases for wind resource assessment

- **Results for Algeria**

Three stations for validation in Algeria are considered. Figure 25 depicts the orography map of Algeria and the location of the stations used and their nicknames DAAE; DABS and DAUA. All stations have wind data measured at 10 m height with data recorded at 1 hour interval and cover the years 2019 to 2021 with data availability greater than 90%.

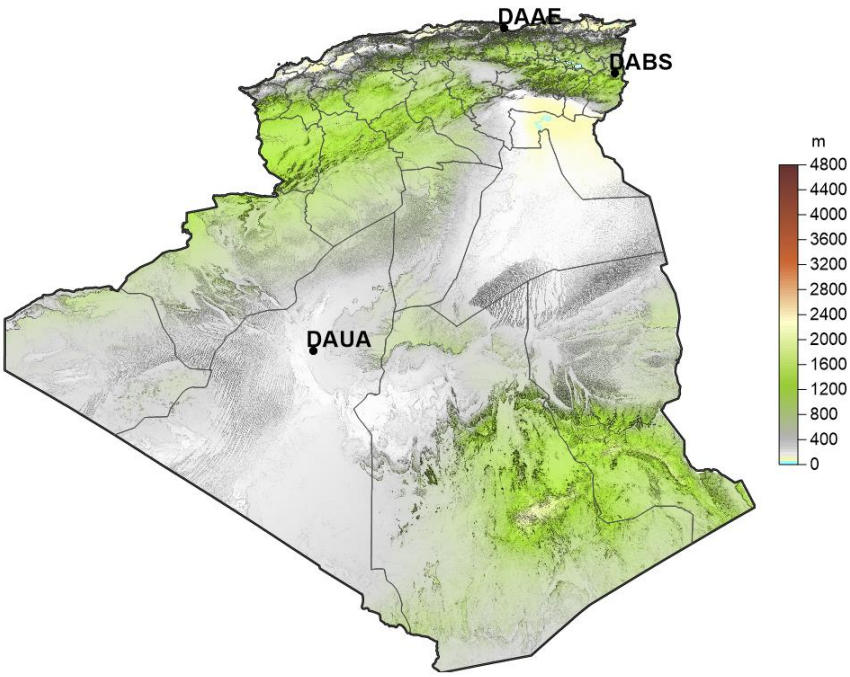


Figure 25 – Orography map of Algeria. Location of the three observational stations accepted for validation. Data availability above 90% covering the years 2019 to 2021. All stations measured wind at 10m height.

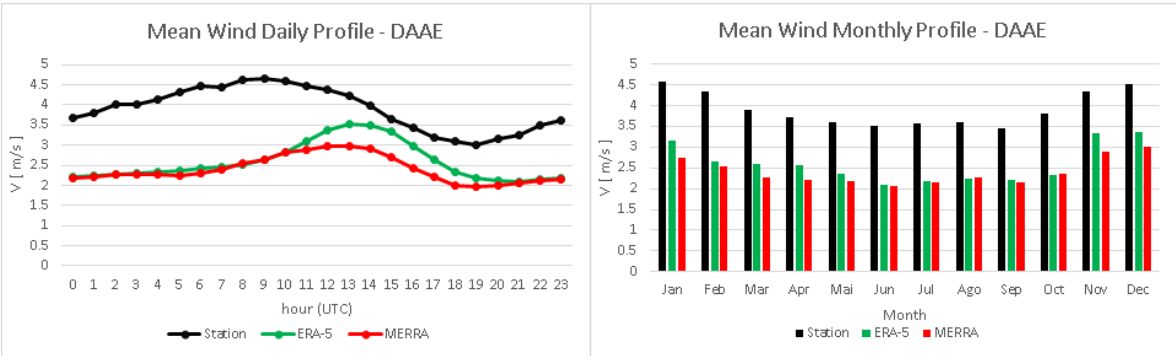
Station DAAE is located at the northern coast of Algeria within a region of complex orography with a long valley (towards southwest to east) surrounded by scarped cliffs. Station DABS is located in a less complex terrain but surrounded by high-cliff dunes (almost the beginning of the Sahara Desert) and very near to the border of Tunisia. Station DAUA is a station located in the middle of Sahara Desert in the mainland country, with no complex orography and low roughness.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Table 2 to Table 4 show the validation results obtained for each station and RE dataset. Figure 26 to Figure 34 depict the mean wind speed daily and monthly profiles, the scatterplots of the wind speed and the wind rose and wind speed frequency distribution for the stations.

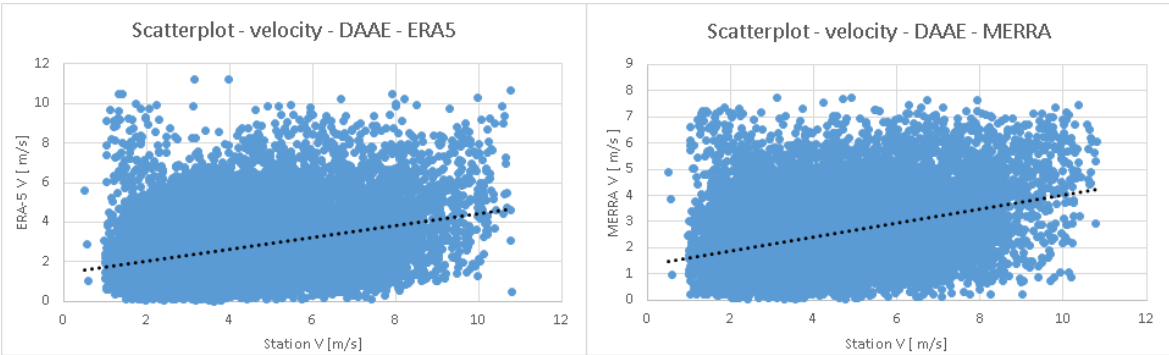
Table 2 – Validation result of the statistical parameters for station DAAE.

Station DAAE	Correlation	BIAS [m/s]	Absolute Deviation [m/s]	RMSE [m/s]	Absolute Deviation [°]	RMSE [°]
ECMWF-ERA5	0.35	-1.31	1.76	2.25	62.2	78.01
NASA-MERRA2	0.37	-1.50	1.76	2.24	70.6	85.35



(a) **(b)**

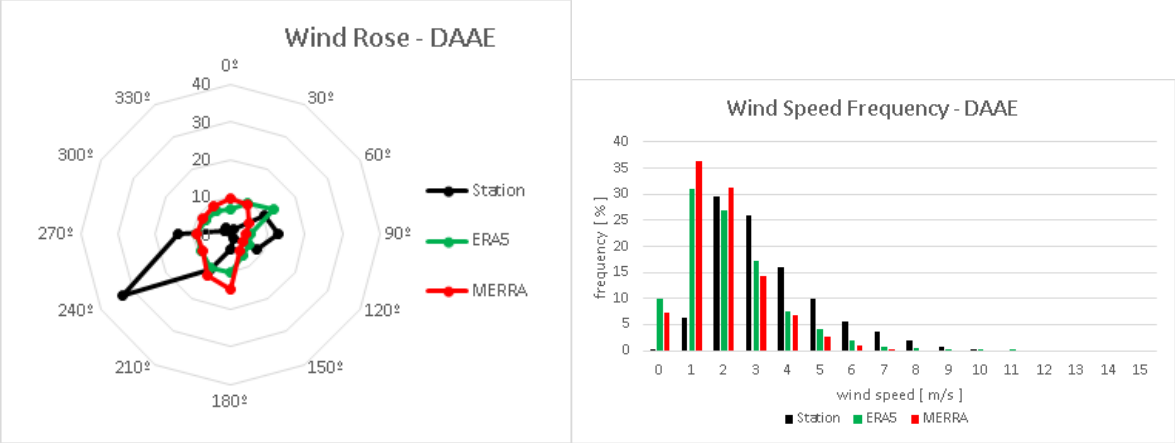
Figure 26 – Wind speed: a) daily and b) monthly profiles for station DAAE and respective RE databases. Validation at 10 m height.



(a) **(b)**

Figure 27 – Wind speed scatterplots observed vs; a) records ECMWF-ERA5 and b) NASA-MERRA2 for DAAE station. Validation at 10 m height.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

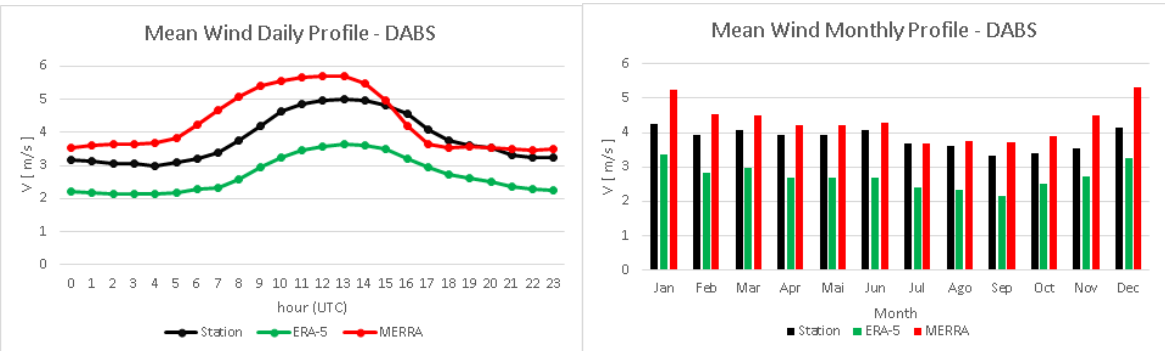


(a) (b)

Figure 28 – a) Wind rose and b) wind speed frequency distribution plot for DAAE station and the RE Databases. Validation at 10 m height.

Table 3 – Validation result of the statistical parameters for station DABS.

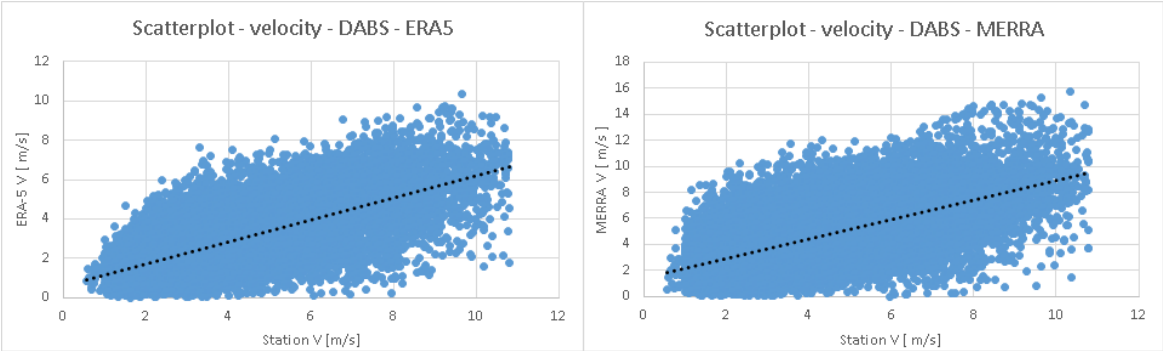
Station DABS	Correlation	BIAS [m/s]	Absolute Deviation [m/s]	RMSE [m/s]	Absolute Deviation [°]	RMSE [°]
ECMWF-ERA5	0.71	-1.11	1.37	1.76	57.19	73.09
NASA-MERRA2	0.59	0.48	1.48	1.92	63.93	81.17



(a) (b)

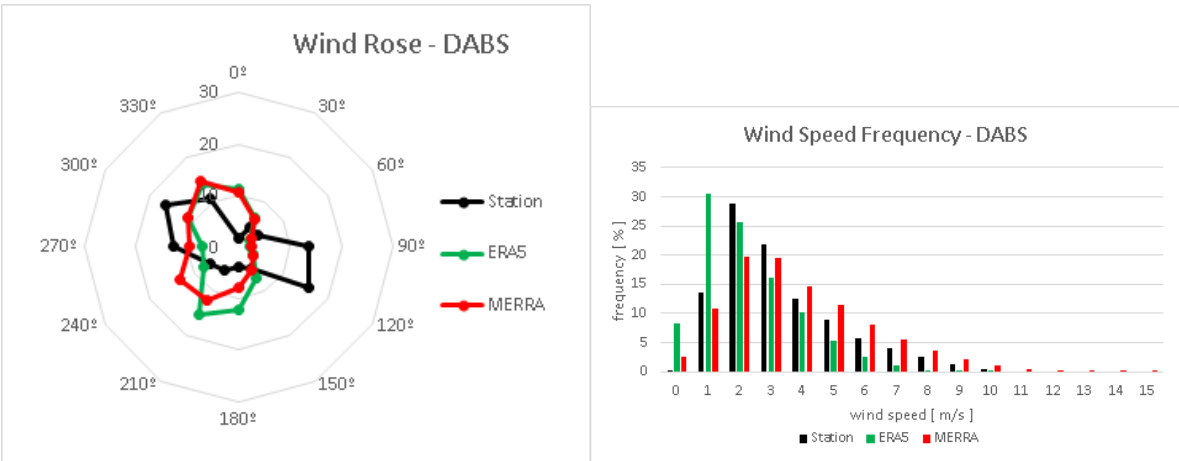
Figure 29 – Mean wind Daily (a) and Monthly (b) profile obtained for station DABS and respective RE databases. Validation at 10 m height.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.



(a) **(b)**

Figure 30 – Scatterplots between wind speed records over time between DABS station and ECMWF-ERA5 (a) and NASA-MERRA2 (b). Validation at 10 m height.



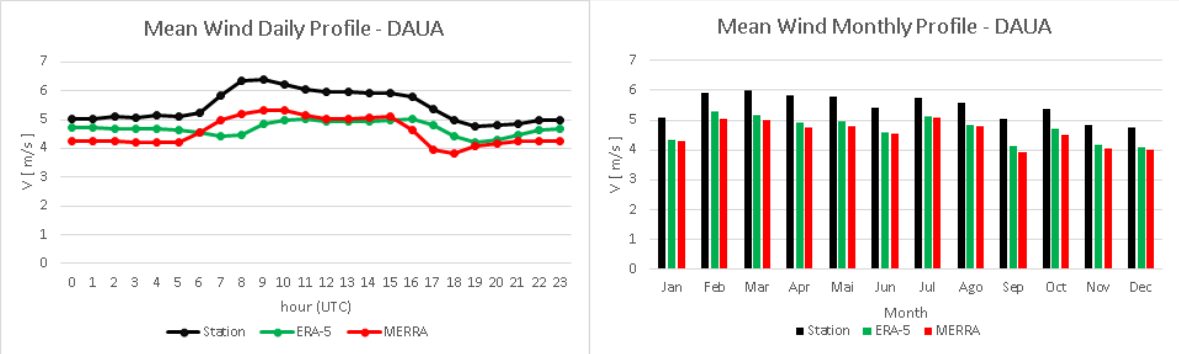
(a) **(b)**

Figure 31 – Wind Rose (a) and Wind speed frequency distribution plot (b) from DABS station and RE Databases. Validation at 10 m height.

Table 4 – Validation Result of the statistical parameters for station DAUA.

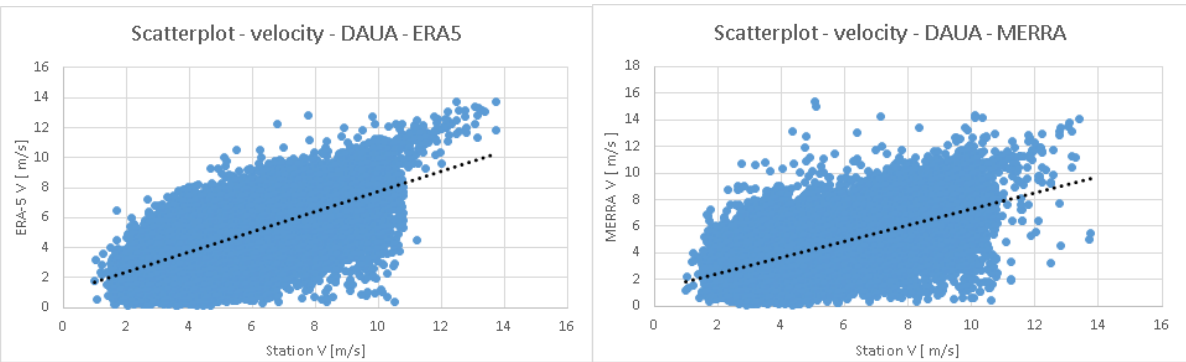
Station DAUA	Correlation	BIAS [m/s]	Absolute Deviation [m/s]	RMSE [m/s]	Absolute Deviation [°]	RMSE [°]
ECMWF-ERA5	0.72	-0.76	1.31	1.75	26.84	40.02
NASA-MERRA2	0.61	-0.88	1.49	1.95	35.28	48.80

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.



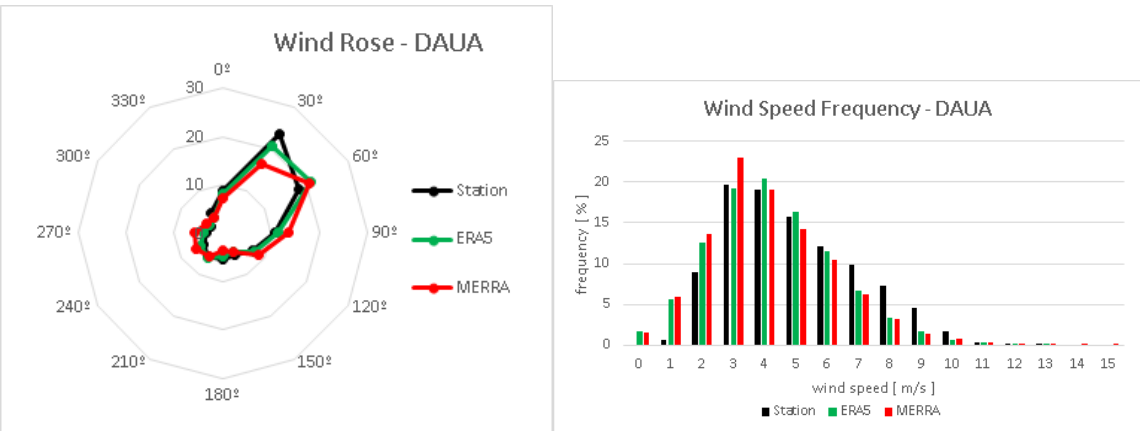
(a) **(b)**

Figure 32 – Mean wind Daily (a) and Monthly (b) profile obtained for station DAUA and respective RE databases. Validation at 10 m height.



(a) **(b)**

Figure 33 – Scatterplots between wind speed records over time between DAUA station and ECMWF-ERA5 (a) and NASA-MERRA2 (b). Validation at 10 m height.



(a) **(b)**

Figure 34 – Wind Rose (a) and Wind speed frequency distribution plot (b) from DAUA station and RE Databases. Validation at 10 m height.

Previous results show that both RE databases present poor performance for representing the wind flow for DAAE station - a coastal area surrounded by scarp cliffs. This means both databases may have an inappropriate spatial resolution to represent an orography of this

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

kind. Therefore, results suggest that both RE databases are poorly recommended to provide wind estimates for this kind of locals. A high-resolution RE database capable to represent more efficiently this type of terrain is desirable to provide more realistic wind data.

In the case of station DABS, the ECMWF-ERA5 shows a good correlation for wind speed but underestimates it (bias equal to -1.11 m/s). On the contrary, NASA-MERRA2 has less correlation but gives little overestimation of the wind speed as revealed by the mean wind speed monthly profile and wind speed frequency distribution. In terms of wind direction, both RE databases show poor performance. Both RE databases were capable to describe the mean daily profile. Although this station is located in a less complex terrain but surrounded by high-cliff dunes – the beginning of the Sahara Desert – probably with moving dunes making the wind direction less accurate, both databases may be equally used although the ECMWF-ERA5 gave better correlation and less wind speed deviation values despite of underestimate the wind speed.

For the case of DAUA station, in the middle of the desert with low complex terrain and low complex roughness both databases behave well in terms of deviations and wind rose.

- **Results for Kenya**

For Kenya two stations are used. Figure 35 depicts the orography map of Kenya and the location of the stations with their nicknames HKJK and HKMO. These two stations recorded wind data at 10 m height with 1 hour interval covering the years 2019 to 2021 with data availability above 85%.

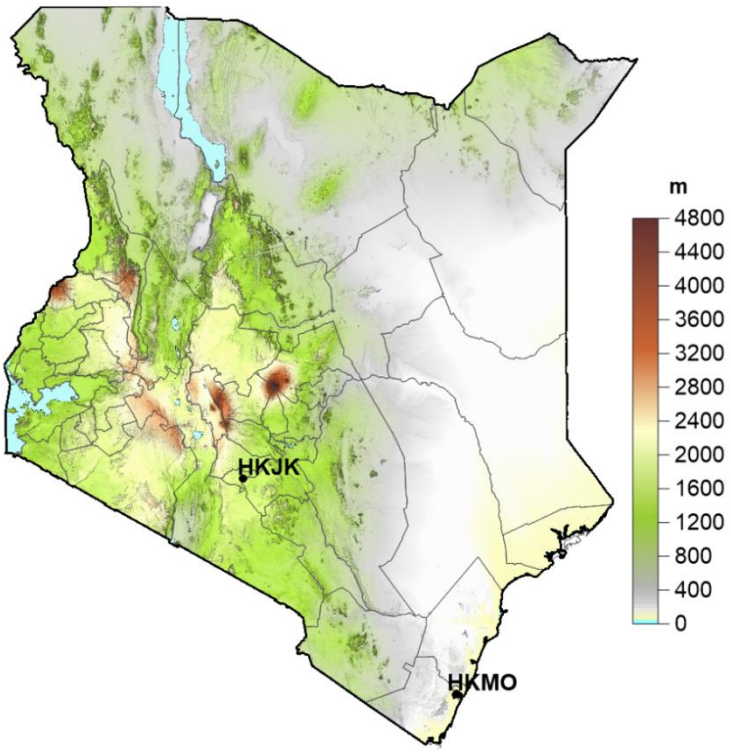


Figure 35 – Orography map of Kenya. Location of the three observational stations used for validation with data availability above 90% covering the years 2019 to 2021. All stations measured wind at 10 m height.

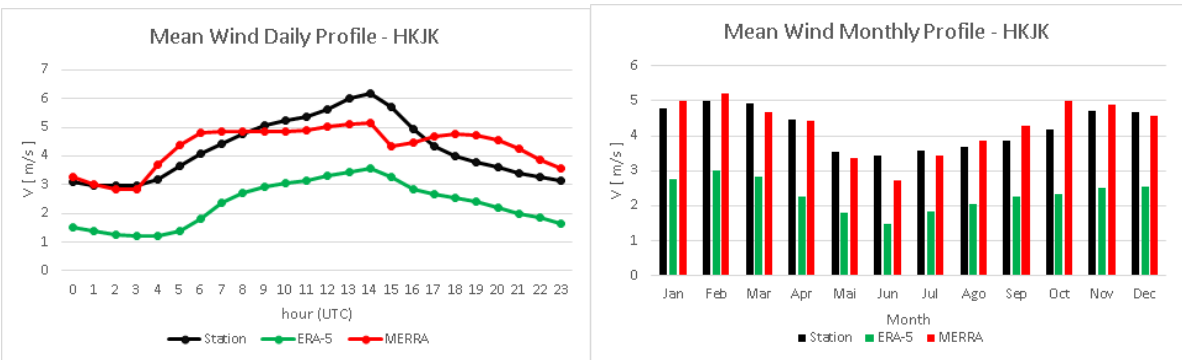
This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Station HKJK is located in a planar zone with low complex orography and low roughness but with high altimetry, around 1700m above sea level, while station HKMO is a coastal station with low orography and roughness in the surrounding of the location.

Table 5 and Table 6 show the validation results obtained in each station for the different RE datasets. Figure 36 to Figure 41 depict the mean wind daily and monthly speed profiles, the scatterplots of the wind speed and the wind rose, and wind speed frequency distribution graph through stations.

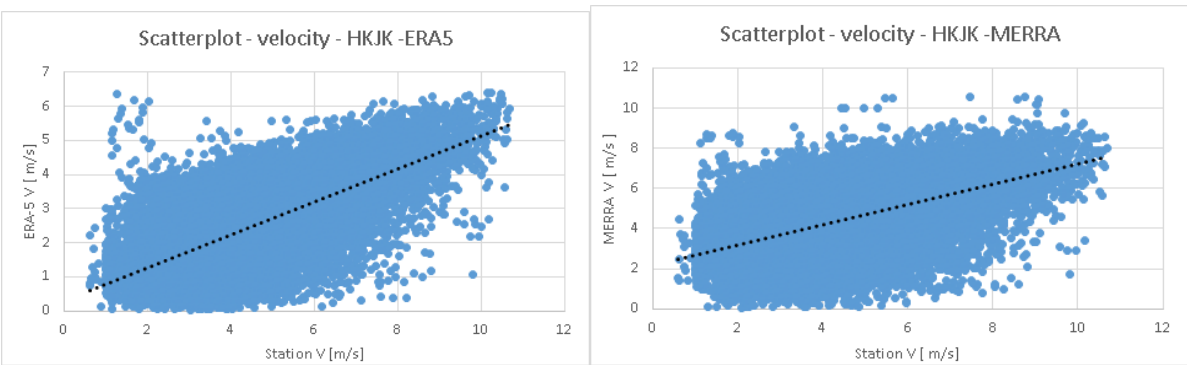
Table 5 – Validation Result of the statistical parameters for station HKJK.

Station HKJK	Correlation	BIAS [m/s]	Absolute Deviation [m/s]	RMSE [m/s]	Absolute Deviation [°]	RMSE [°]
ECMWF-ERA5	0.70	-1.92	1.98	2.30	39.48	57.75
NASA-MERRA2	0.55	0.06	1.30	1.63	43.15	59.12



(a) **(b)**

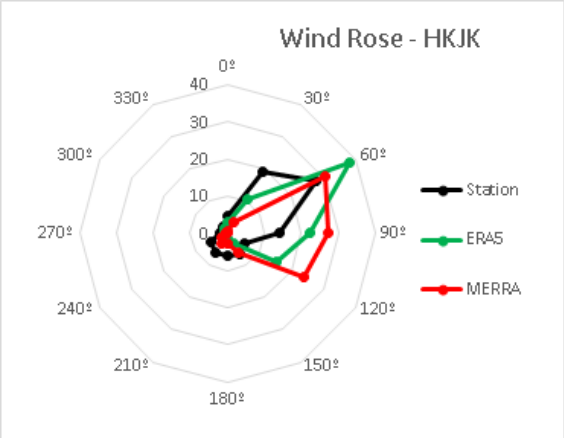
Figure 36 – Mean wind Daily (a) and Monthly (b) profile obtained for station HKJK and respective RE databases. Validation at 10 m height.



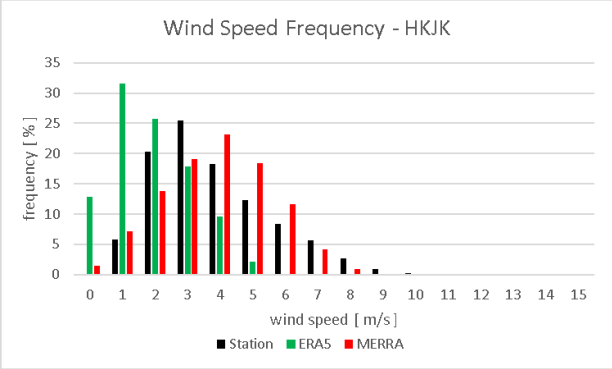
(a) **(b)**

Figure 37 – Scatterplots between wind speed records over time between HKJK station and ECMWF-ERA5 (a) and NASA-MERRA2 (b). Validation at 10 m height.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.



(a)

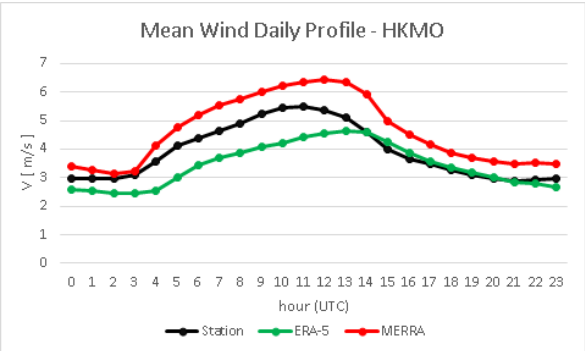


(b)

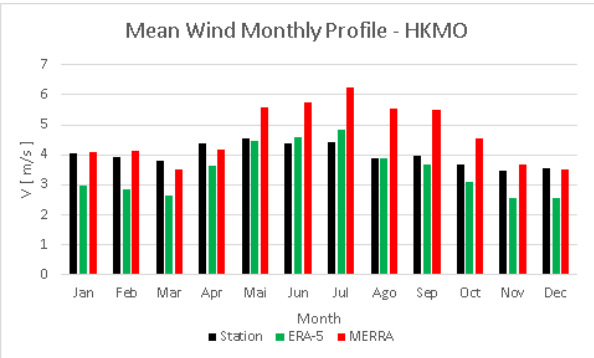
Figure 38 – Wind Rose (a) and Wind speed frequency distribution plot (b) from HKJK station and RE Databases. Validation at 10 m height

Table 6 – Validation result of the statistical parameters for station HKMO.

Station HKMO	Correlation	BIAS [m/s]	Absolute Deviation [m/s]	RMSE [m/s]	Absolute Deviation [°]	RMSE [°]
ECMWF-ERA5	0.64	-0.49	1.03	1.34	21.92	31.64
NASA-MERRA2	0.66	0.72	1.28	1.61	29.57	42.28



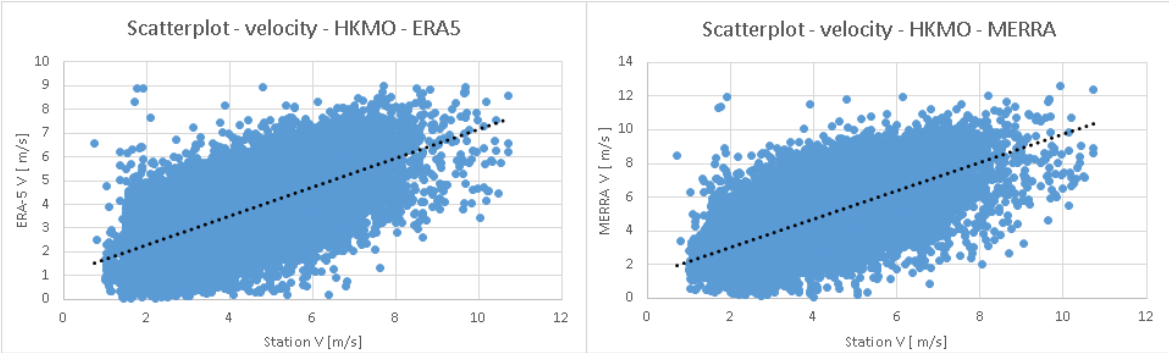
(a)



(b)

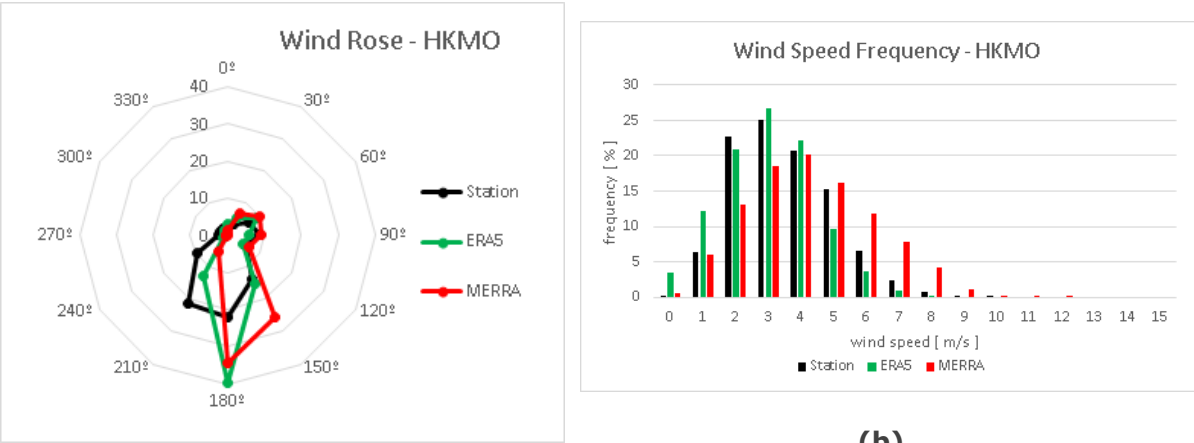
Figure 39 – Mean wind Daily (a) and Monthly (b) profile obtained for station HKMO and respective RE databases. Validation at 10 m height.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.



(a) **(b)**

Figure 40 – Scatterplots between wind speed records over time between HKMO station and ECMWF-ERA5 (a) and NASA-MERRA2 (b). Validation at 10 m height.



(a) **(b)**

Figure 41 – Wind Rose (a) and Wind speed frequency distribution plot (b) from HKMO station and RE databases. Validation at 10 m height.

Results for HKJK station show a good agreement in the correlation for ECMWF-ERA5 but with wind speeds underestimated (almost 2 m/s). On the other hand, NASA-MERRA2 revealed a weaker correlation when compared with ECMWF-ERA5, but the mean wind speed is better represented as seen by the mean wind speed monthly profile and frequency distribution..

ECMWF-ERA5 present a better estimation of the wind rose with less deviation in direction values when compared with NASA-MERRA2. For this particular case, and as an overview result, both databases may be suitable as a first approach to represent the wind. Nevertheless, the NASA-MERRA2 may be better suitable to represent the wind in this type of terrain and roughness.

Validation results using data from the HKMO station show that both databases have similar correlation, although in terms of wind and direction deviations the ECMWF-ERA5 has better performance than NASA-MERRA2. In fact, NASA-MERRA2 overestimates the wind speed as seen in the mean daily and monthly wind profiles as well as in the wind speed distribution values. Also, the wind rose is better represented by the ECMWF-ERA5 rather than NASA-MERRA2. For this particular case, a coastal area with low complex orography and low roughness the ECMWF-ERA5 has better performance than NASA-MERRA2.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

- **Results for Mozambique**

For the Mozambique case, two non-public ground-based stations are used. These stations can be used for validation purposes by LNEG. These two stations recorded data at two levels, 40 m and 60 m height for the years 2012 to 2013 with one complete year of data at 10 minutes interval. Figure 42 depicts the orography map of Mozambique and the location of the stations with their name corresponding to the closer town. Data availability is 100% for both stations. The validation results are performed for 40 m height selecting all records multiple of 1 hour to be compatible with the RE databases.

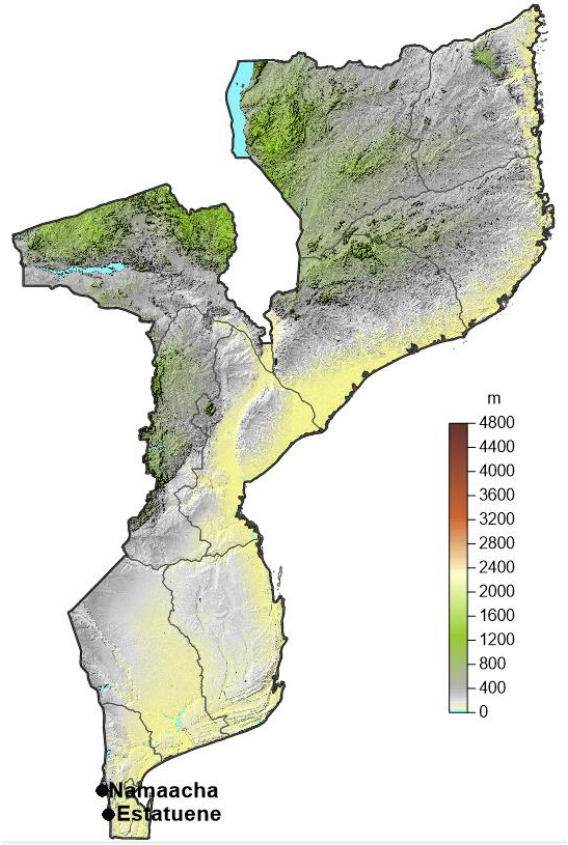


Figure 42 – Orography map of Mozambique. Location of the two observational stations used for validation. Data availability is 100% covering years 2012 to 2013. All stations measured wind speed and direction at 40 m height.

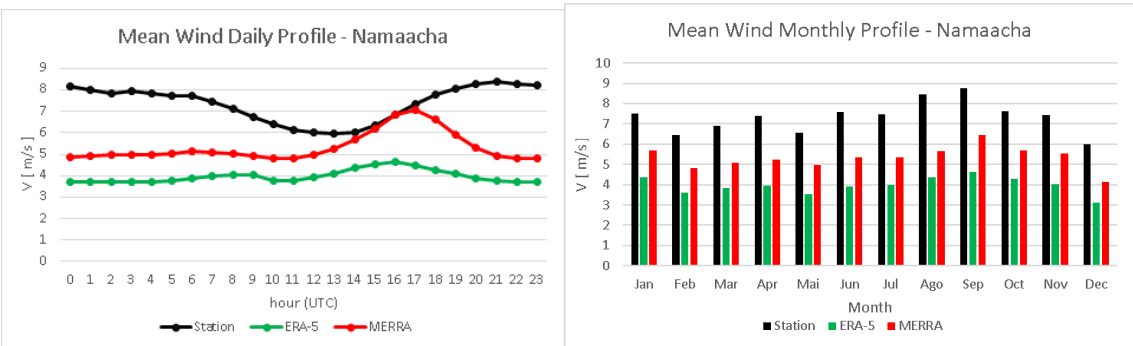
Both stations are located in the most Southern region of Mozambique. The station in Namaacha area is located in a zone with irregular middle-complex orography with medium roughness while the station in Estatuene region is placed in a flat area with low complex orography but with medium roughness.

Table 7 and Table 8 show the validation results obtained for each station using the two RE datasets under analysis. Figure 43 to Figure 48 depict the mean wind daily and monthly profiles, the scatterplots of the wind speed and the wind rose and wind speed frequency distribution graph through stations.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

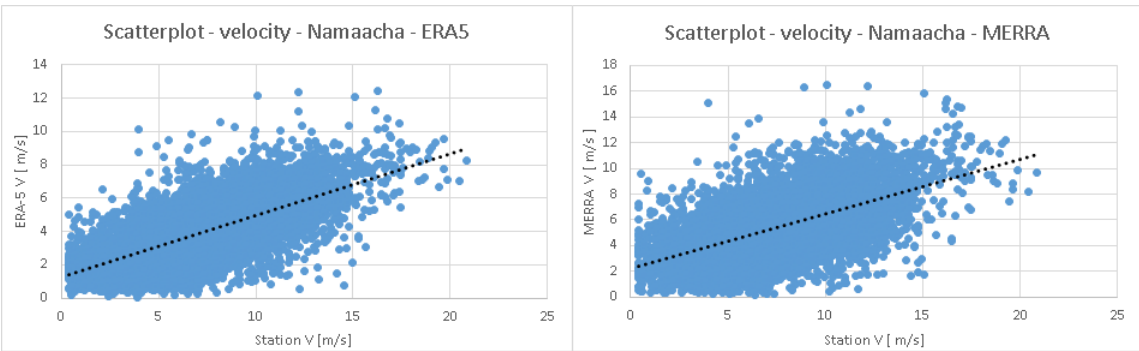
Table 7 – Validation Result of the statistical parameters for station in Namaacha.

Station Namaacha	Correlation	BIAS [m/s]	Absolute Deviation [m/s]	RMSE [m/s]	Absolute Deviation [°]	RMSE [°]
ECMWF-ERA5	0.66	-3.38	3.52	4.15	37.38	51.15
NASA-MERRA2	0.58	-2.01	2.66	3.32	33.66	46.97



(a) **(b)**

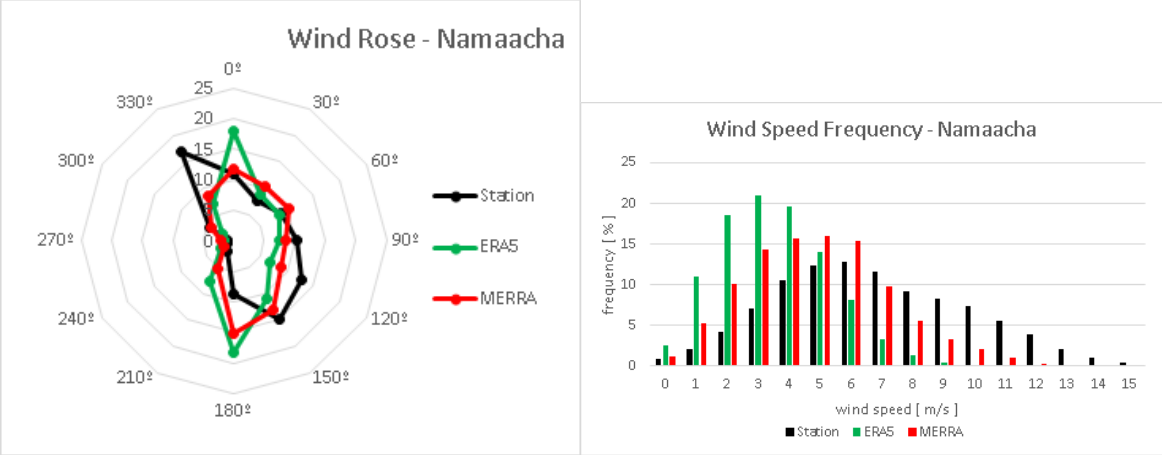
Figure 43 – Mean wind Daily (a) and Monthly (b) profile obtained for station in Namaacha and respective RE databases. Validation at 40 m height.



(a) **(b)**

Figure 44 – Scatterplots between wind speed records over time between station in Namaacha and ECMWF-ERA5 (a) and NASA-MERRA2 (b). Validation at 40 m height.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

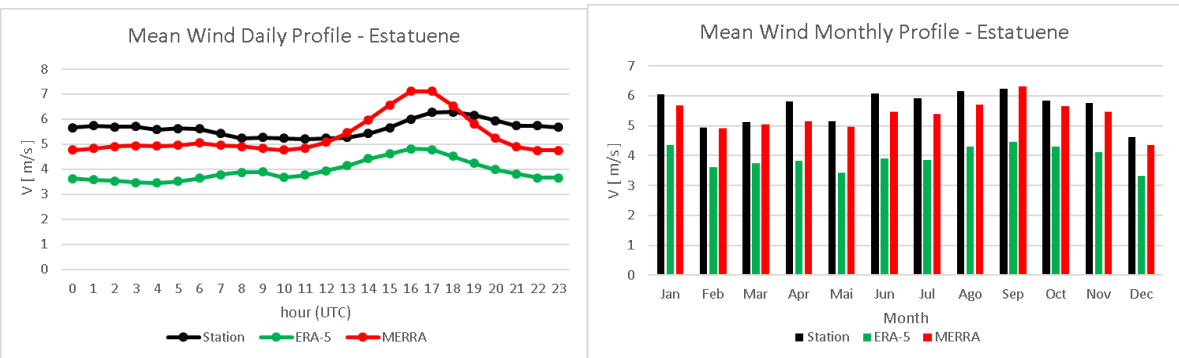


(a) **(b)**

Figure 45 – Wind Rose (a) and Wind speed frequency distribution plot (b) from station in Namaacha and RE Databases. Validation at 40 m height.

Table 8 – Validation Result of the statistical parameters for station in Estatuene.

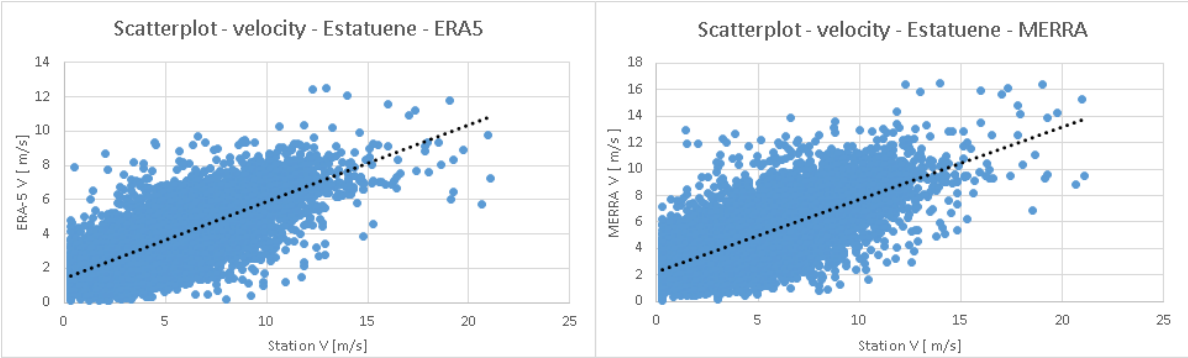
Station Estatuene	Correlation	BIAS [m/s]	Absolute Deviation [m/s]	RMSE [m/s]	Absolute Deviation [°]	RMSE [°]
ECMWF-ERA5	0.72	-1.71	2.05	2.59	34.95	46.49
NASA-MERRA2	0.68	-0.30	1.60	2.11	30.74	43.66



(a) **(b)**

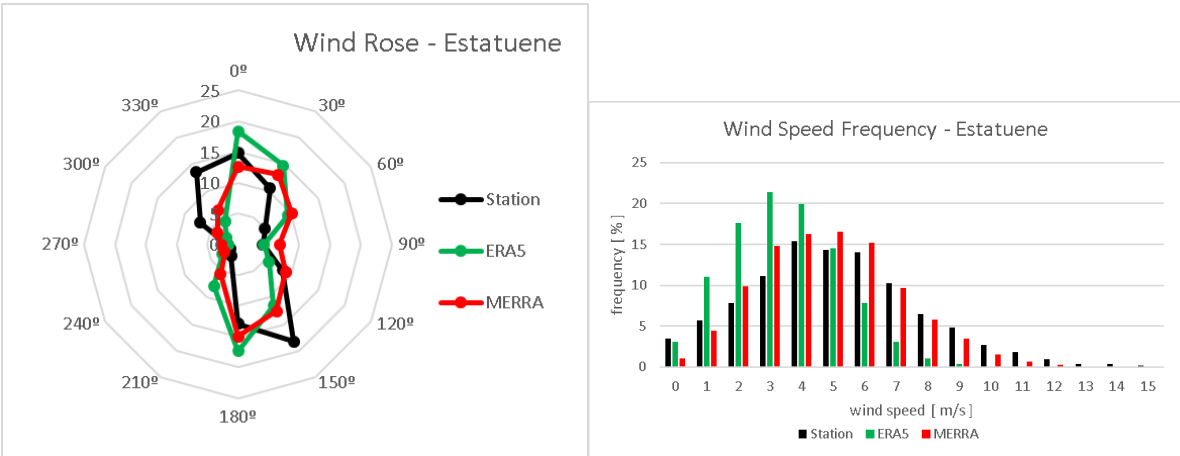
Figure 46 – Mean wind Daily (a) and Monthly (b) profile obtained for station in Estatuene and respective RE databases. Validation at 40 m height.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.



(a) **(b)**

Figure 47– Scatterplots between wind speed records over time between station in Estatuene and ECMWF-ERA5 (a) and NASA-MERRA2 (b). Validation at 40 m height.



(a) **(b)**

Figure 48– Wind Rose (a) and Wind speed frequency distribution plot (b) from station in Estatuene and RE Databases. Validation at 40 m height.

The analysis for Namaacha station data shows that both RE databases may be less effective in representing the wind for this kind of middle/complex orography terrains with medium roughness. Both RE databases show different patterns. Although NASA-MERRA2 show less correlation in wind speed than ECMWF-ERA5, it gave better results in deviation values both for wind speed and wind direction. It is visible that NASA-MERRA2 wind speed monthly profile and frequency distribution are quite similar to observation rather than the ECMWF-ERA5 which clearly differs from the station. Other evidence is ECMWF-ERA5 underestimate clearly the wind values. As a general overview of this local, taking into account the orography and roughness conditions as well the wind rose, it is desirable a high-resolution database capable of representing better the orography to potentially provide better results. However, the results achieved for this station revealed that the NASA-MERRA2 has better performance than ECMWF-ERA5.

For Estatuene station both RE databases are capable, in a first approach, to represent the wind speed for a low complex terrain with medium roughness. It is clearly visible that ECMWF-ERA5 has better correlation that NASA-MERRA2 but as similar to Namaacha station. The statistical parameters revealed that NASA-MERRA2 has much less deviations

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

in both wind speed and direction rather than ECMWF-ERA5. In fact, NASA-MERRA2 represented quite satisfactory the mean monthly profile and the wind speed frequency distribution. Again, ECMWF-ERA5 underestimated quite high wind speed values. As a general overview for this local with low complex orography but with medium roughness, NASA-MERRA2 had better performance than ECMWF-ERA5.

- **Overall validation results**

As a summary of the validation results for the evaluated countries – Algeria, Kenya and Mozambique, it was clear that the RE database quality depends on the terrain and roughness types. This conclusion was expected since both RE databases have limited resolution to represent quite well the wind behaviour over complex terrain or locals with medium roughness. Therefore, ideally, locals dominated by complex orography or medium/high roughness should be better represented by high-resolution spatial RE databases instead of using the RE databases analysed – ECMWF-ERA5 and NASA-MERRA2 that showed poor results over this type of terrain and roughness.

For low complex terrains or low roughness both RE databases have good quality in wind speed and direction. However, if there is a mixture of low complex terrain with medium roughness, both RE databases can still be used as a first approach for inquiring the wind speed and direction on each territory, but for this kind of mixing terrain/roughness, NASA-MERRA2 evidenced slightly better performance in comparison with ECMWF-ERA5.

In terms of observations, no data was provided from SETaDiSMA partners. Therefore, the majority of the stations used were merely publicly synoptical meteorological stations measuring the wind at 10 m height. Although data seemed to be with good quality there is no information about the presence of obstacles in the nearby stations that might affect the wind data used for the validations presented above.

2.3.2 Validation of databases for solar resource assessment

The identification of databases publicly available for solar resource assessment as well as their validation for use in solar resource assessment was performed within the scope of the LEAP-RE work package Rural African Markets using Standalone Solar (WP10 - PURAMS) and is presented in the report D10.2 - Standalone solar cooking appliance design metrics (Couto et al., 2022). Since this deliverable no further ground-based data was retrieved for the reference countries. In this sense, to avoid overlap information already provided, only a brief outline of the main results identified in D10.2 is provided below.

Three ground-based stations with adequate data for validation purposes were identified for the countries under study in SETaDiSMA and PURAMS work packages. These ground-based stations are in Kenya, located at: Homa-Bay, Laisamis and Narok (Figure 49). All data from these stations are provided by the World Bank Group database and GeoSun Africa Company. Therefore, the validation assessment was restricted to this country. The ground-based stations are equipped with two thermopile pyranometers located at 2 m height above ground level with sensors from Delta-T and Hukseflux manufacturers. Measured solar data comprises DHI and GHI in W/m² at a time interval of 1 minute, covering the complete period between 08 December 2019 to 31 December 2021. The measured data was compared with satellite-based databases' GHI and DHI parameters, namely, the CAMS radiation service, PVGIS and POWER.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

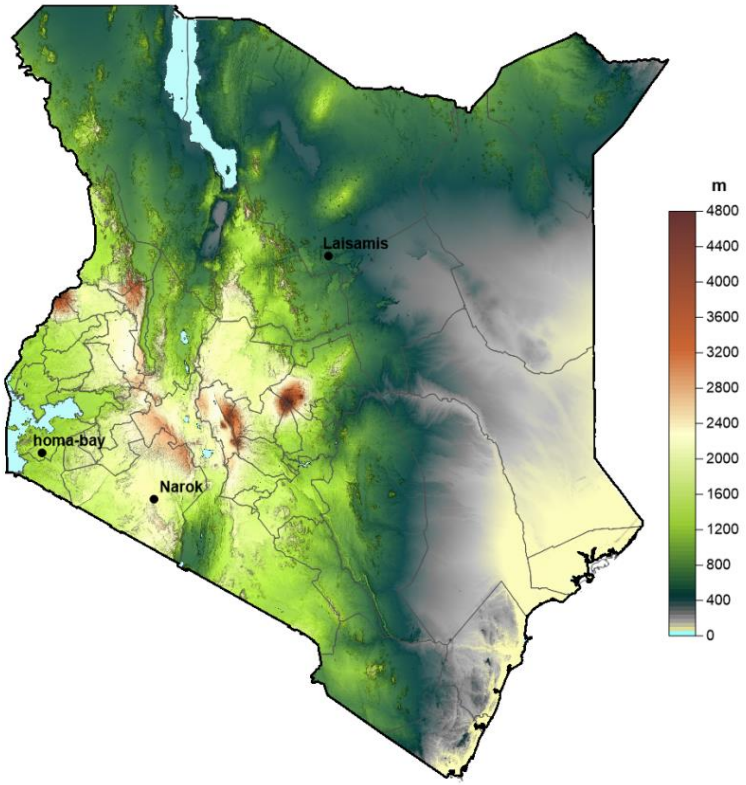


Figure 49 - Location of the ground-based stations in Kenya with solar data used for validation purposes overlayed with the terrain map of Kenya.

The metrics computed are the same applied in the case of wind power. In Table 9 the main basic statistics derived from the validation study between each ground-based station and satellite-based databases – CAMS, PVGIS and POWER are provided. It should be noted that PVGIS database results were only processed for the ground-based station "Laisamis". For the other stations, PVGIS database returned a time-series of "NaN" values meaning no available data or a temporary problem in retrieving data for these locals. In due date of this report, PVGIS database kept returning the same behaviour of "NaN" values for GHI for the local stations "Homa-Bay" and "Narok".

Broadly speaking, results from Table 9 confirm that satellite-based databases have capacity to provide solar data for the locals under study with correlation values above 0.96 for all cases. This high value of correlation means that solar variability is well estimated. The maximum bias error, in absolute value, was 5% and the maximum RMSE value was 22%. These values are in line with the global validation studies of SolarGIS and CAMS radiation service for African Countries. The mean daily profiles and monthly daily profiles show bias values varying between 1% and 9% except for the PVGIS case that show values around 14% - also in accordance with the global validation studies of SolarGIS and CAMS radiation service for African Countries. Nevertheless, the validation results for Kenya show that all satellite-based databases are good for performing solar resource assessment. The results enlighten that the CAMS satellite-based methodology was the one providing the highest correlation values with the minimum bias and RMSE values.

Based on these results, it is proposed to use the CAMS radiation service database to provide solar data for the studies under development in LEAP-RE activities in Kenya.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Table 9 - Statistical parameters obtained for the three ground station used for the solar resource assessment.

Station	Database	Correlation	BIAS [%]	RMSE [%]	BIAS Daily Profile [%]	BIAS Monthly Profile [%]
Homa-Bay	CAMS	0.985	3.85	14.47	6.06	3.84
	PVGIS*	-	-	-	-	-
	POWER	0.961	4.94	22.77	9.11	8.81
Laisamis	CAMS	0.990	3.67	11.86	4.90	3.68
	PVGIS	0.977	3.96	17.75	13.30	14.23
	POWER	0.977	-2.02	15.76	1.42	0.17
Narok	CAMS	0.989	4.06	12.41	4.42	4.05
	PVGIS*	-	-	-	-	-
	POWER	0.976	-5.69	18.36	1.36	-2.76

** PVGIS database returned a time-series of "NaN" values not allowing to use this station.*

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

3. Demand estimation

It has been highlighted that energy system models play a crucial role in the sizing of micro-grids for enabling access to energy in rural areas; nonetheless, they are still far from being capable of satisfying the specific needs that such a problem inherently brings along. In particular, it has been observed the need for considering the problem under a time evolving scenario, able to tackle the uncertainties deriving from the new perspective. To this regard Balderrama (Balderrama et al., 2019) reports how the poor forecast of load evolution of a community in rural Bolivia led to the construction of a mini-grid that only after a few years proved to be undersized due to the new energy uses that evolved in the area. In the introduction of their work Balderrama et al. report other cases of inaccurate demand projections that lead to unexpected situations and wrong mini-grid sizing in the long run. In particular in (Ulsrud et al., 2011) Ulsrud et al. show how in rural India, solar mini-grid projects experienced an increase in energy demand over time for mainly two reasons: users starting to connect more appliances and new connections to the grid were taking place, while others were inscribed in a waiting list, which was due to an increase of the share of population that became able to afford the connection to the grid after its deployment. Diaz et al. in (Díaz et al., 2010) highlight how population growth, higher per-capita consumption and new connections to the mini-grid led to up to double electricity demand in different examples over eight years. Kobayakawa and Kandpal (Kobayakawa & Kandpal, 2015) report clearly how a mini-grid in rural India experienced growth in average daily demand in two years and a half. Finally Riva et al. (Riva et al., 2019) highlight how household electricity consumption is strictly dependent on household income, and as the income of the village can benefit from the access to electricity itself triggering a causal loop bringing an increase in load demand of the village as showed by the same author (Riva, Tognollo, et al., 2018).

All these examples highlight the relevance of properly assessing the load demand that the future mini-grid will have to satisfy, showing how mid-sized mini-grids bring to rapid failure or abandonment of the system. For this reason, proper modelling techniques are required, able to link the estimation of the future demand with the specific needs of the community, and in turn, the mini-grid sizing tools need to be able to grasp this nuances and size the mini-grid accordingly.

Among numerous tools available for the modelling, sizing and optimisation of off-grid energy systems, the most widely adopted is HOMER® (HOMER - Hybrid Renewable and Distributed Generation System Design Software, n.d.), a proprietary software capable of identifying a set of alternative (from the least-cost one to any other feasible) micro-grid configurations as a result of an enumerative optimisation process. Despite its wide range of technological options and features, its closed-source nature does not allow for customisability, hindering context-adaptability and the implementation of new features. Indeed, as regards the accounting for time-evolving electricity demand profiles, the latest release of HOMER® only allows to set a year-by-year series of multipliers to match load forecasts; finer load projection analyses are not possible. Furthermore, the load evolving mode of HOMER® performs a heuristic optimization, not assuring that the result is a global optimum. Similar considerations are valid for other widespread models, such as DER-CAM (DER-CAM, n.d.) and iHOGA (IHOGA, n.d.). Both models have been developed in academia but not released as open-source software and currently they do not allow time-evolving constraints. An open-source tool for micro-grid sizing, based on two-stage stochastic optimisation, is proposed by Balderrama et al. (Balderrama et al., 2019), Micro-gridsPy, which allows adaptation of the model structure towards context-specific formulation requirements depending on the scope of application (Stevanato et al., 2020), but does not allow time-evolving constraints nor enable the expansion of the system capacity over time in response to a non-linearly time-evolving load demand.

Hartvingsson et al. (Hartvigsson et al., 2018) try to overcome this limitation by iterating DER-CAM with a system-dynamics model of appliances diffusion and a load profile

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

generation model, in an attempt to account for feedbacks between socio-economic dynamics and capacity-expansion decisions. Still, the proposed integrated modelling approach does not allow including the non-linear load evolution as a direct input to the optimisation model. Another attempt at controlling the uncertainty associated with boundary conditions evolution over time has been made by Dufo-López et al. (Dufo-López et al., 2016), who combine probability density functions of input parameters variations with stochastic optimisation. Nonetheless, this approach fails to allow for the explicit simulation and the accounting of alternative load evolution patterns. Further examples of a variety of tools or academic algorithms for micro-grids optimisation that yet lack features for load evolution and capacity expansion modelling can be found in dedicated literature reviews (Markovic et al., 2011; Sinha & Chandel, 2014).

On the other hand, the broader field of energy system modelling has experienced in recent years an increasing spread of open-source, customisable energy system models (Pfenninger et al., 2018). Some of these, though not originally conceived for application to small-scale villages, present valuable features in terms of accounting for time-evolving demand and multi-step capacity expansion decisions. For instance, Riva et al. (Riva et al., 2019) couple a long-term energy demand forecasting approach with the energy system optimisation model OSeMOSYS, developed by Howells et al. (Howells et al., 2011), and apply the soft-linked model to rural villages in India. The study highlights how the cost-optimal system design can be significantly influenced by demand variations over time.

In the end, accounting for load evolution within micro-grids optimisation models is not enough. DeCarolis et al. (DeCarolis et al., 2017) collect evidences of experienced discrepancies among load forecasts and real load evolutions. As a matter of fact, modelling load evolution does not represent a significant improvement to the state-of-the-art if not coupled with robust uncertainty management. As pointed out by Pfenninger et al. in (Pfenninger et al., 2014), modelling of uncertainties may be classified as one of the current main challenges for energy modellers, while Yue et al. (Yue et al., 2018) underline how this issue has been poorly addressed or even ignored in the vast majority of the papers they reviewed. In the same work, they highlight stochastic programming, modelling to generate alternatives, Monte Carlo analysis and robust optimization as the most adopted non-deterministic approaches.

Defining load demand is a complex and multifaceted task that can take on different forms and methods depending on the specific context. For example, it is possible to estimate only the demand for electricity, or the demand for different energy carriers, such as space heating or domestic hot water, with different time resolutions and time horizons. For long time horizons, studies focus on the evolution of demand. In fact, it is likely that a successful project will push new residential or commercial users to connect or increase the attractiveness of new devices (Díaz et al., 2010; Ganguly et al., 2020).

In general, there are two families of models for estimating current energy demand:

- *Top-down models*, which consist of setting a target power and energy level and sizing the solution accordingly, are typically adopted in expert-based methods such as in (Louie & Dauenhauer, 2016);

- *Bottom-up models*, which employ surveys to quantify beneficiary needs and desired appliance usage, are usually combined with tools that take into account the stochastic variability of habits over the course of the day (such as, for example, LoadProGen (Mandelli et al., 2016) or RAMP (Lombardi et al., 2019)).

Both categories of models need a large amount of high-resolution data: top-down models require defining energy consumption bands and identifying the factors that assign these bands to different population groups (Daioglou et al., 2012; Ruijven et al., 2011).

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

On the other hand, bottom-up models require a large amount of high-resolution data, since their operating logic assumes that the characteristics and habits of the users are known to the modeler, who can use them to characterize the load demand of the area or category of users.

The research was conducted in the direction of identifying strategies and methodologies to provide reliable inputs for load demand models, ranging from System Dynamics-type approaches (Riva, 2019) up to the most recent seminal research involving Artificial Intelligence techniques, Machine Learning and GIS for input data generation (Allee et al., 2021; Dominguez et al., 2021a; Fobi et al., 2021; VIDA, n.d.).

Despite the considerable research being conducted in this direction, it is still in its infancy and the production of input data for load demand models still relies heavily on specific interviews and surveys with non-electrified communities, which have proved to be a source of unreliable data (Blodgett et al., 2017).

3.1 Literature review and existing tools

A scientific literature review was conducted with the aim of understanding the factors influencing a community's electricity demand, with a focus on demand models to generate inputs for energy planning models (EPMs). The objective of the review is to identify a comprehensive list of potential electricity demand drivers tested in the literature, to be used as a basis for building the appliance ownership database.

A total of 34 articles, including 5 review papers, were reviewed with special focus to developed countries (DCs).

3.1.1 Reviews on demand estimation

The review papers were instrumental in understanding the context of demand modeling for energy planning. As pointed out by Debnath and Mourshed, most of the current energy planning models (EPMs) were created in developed countries, often influenced by developed country assumptions (Debnath & Mourshed, 2018). However, municipalities have different goals when it comes to assessing energy demand. The challenges ahead are greater (e.g. electrification strategies, suppressed demand estimation) and may require *ad hoc* solutions, such as the widespread adoption of decentralized energy systems. These differences are mirrored by the EPMs: requiring specific models, adapted to the peculiarities of the DCs.

Research efforts have been made in recent decades to fill this gap. In their review of case studies on energy planning in remote areas, Riva et al. (Riva, Tognollo, et al., 2018) have proposed a classification in terms of spatial coverage, planning horizon, energy carrier, mathematical model of decision criteria and demand sector. Their analysis revealed a difference in approaches, depending on the spatial coverage: planning at a regional or national level, not usually aimed at the rigorous design of the components of the energy system, tends to use aggregated data to evaluate demand and extrapolate its evolution over time, adopting a classic top-down approach. When dealing with smaller scales, bottom-up models, which are based on field data, are better suited to capturing the socio-economic and cultural dynamics that can influence local demand. However, scarcity of data often limits the applicability of these models and, in particular, forces them to adopt simplistic approaches to long-term demand forecasting, such as assuming a fixed demand over time. The database developed in this work, which aims to become a source of input for community-level EPMs addressing the problem of data scarcity, should therefore include, when possible, also information on the evolution of demand over time.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Kuster et al. (Kuster et al., 2017), reviewing 113 electric load forecasting models, highlighted the problem of high data needs related to the bottom-up approach in forecasting long-term electric demand, stating that one of the disadvantages is represented by the large amount of data and the lack of information in the long run. The document also introduced a classification of the input variables of the models, dividing them into the following classes: socio-economic, environmental (linked to meteorological conditions), construction and occupation (linked to the characteristics of the dwelling) and time index (linked to past question). This last analysis constitutes the starting point for the classification used to classify the drivers derived from the publications. Jones et al. (Jones et al., 2015) conducted an extensive literature review on studies analyzing electricity consumption factors, arriving at the definition of three classes: socio-economic, housing and household appliance factors.

Starting from the classification proposed by Jones et al., the number of classes has been extended to include other drivers, extracted from the literature review. The final classification of the drivers consists of:

1. Socio-economic drivers: a broad category that includes information on socio-economic status for both communities and individual respondents (e.g., population density, household income and composition, business revenues).

2. Dwelling drivers: drivers relating to the characteristics of the households' homes (for example, number of rooms).

3. Appliance Drivers: Appliance-related drivers (e.g., price and power rating).

4. Past Demand Data: A category that includes modeling techniques that rely on past data on electrical demand (for example, past load profiles).

5. Supply drivers: drivers related to the supply side of the electricity system (for example, the number of hours of electricity availability).

6. Alternative Energy Sources Drivers: Drivers related to the characteristics of energy sources other than electricity (for example, the price of kerosene or candles).

7. Geographic Drivers: Drivers related to the location of the community (for example, climate zone and distance from the nearest city).

8. Cultural drivers: drivers related to culture and habits (for example, religion).

In the next section, the publications analyzed are divided according to their study objective (ownership of household appliances, demand for electricity, load profile, evolution of demand over time and other drivers) and each of them is briefly presented to draw insights on the methodology and drivers to adopt in this work. At the end of this section, a table shows the factor classes (according to the classification described above) which the identified drivers fall into.

3.1.2 Literature with focus on appliance ownership

Regression has been used in all publications that have developed models to predict household appliance ownership. In particular, Rao and Ummel (Rao & Ummel, 2017) developed a logistic model and a boosted regression tree, a type of machine learning algorithm, to predict the ownership of refrigerators, washing machines and televisions. For each model, the authors evaluated two sets of covariates: a sparse set, including only income and level of urbanization, and a diverse set, including socio-economic and cultural factors. They concluded that the difference in the input variable groups had a greater impact on the prediction accuracy than the difference between the models. Kurata et al.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

(Kurata et al., 2018) focused on predicting solar home system (SHS) possession, using data from field surveys conducted in Bangladesh. They did this by applying an Ordinary Least Square (OLS) regression, based on the linear least squares method. By differentiating residential from commercial users, they found that the latter method is more sensitive to energy costs. The same type of model was employed by Richmond and Urpelainen (Richmond & Urpelainen, 2019) on data from 5000 Indian households. The study evaluated the influence of time since electrification on three types of variables: ownership of individual appliances, level of ownership (using a level-based breakdown of appliances developed within the study), and number of appliances owned by type. Ordinary least squares (OLS) models were developed to estimate the first two types, while a Poisson model was used for the number of household appliances. The results were characterized by a positive effect of the time variable on all three types of output, thus highlighting the importance of considering the time dynamics in sizing an electrical system.

A survey of home appliances conducted in western Kenya was used by Lee et al. (Lee et al., 2016) to study the relationship between household connection status (national grid, SHS, not connected) and home appliance ownership levels. SHS was found to have no significant impact on ownership of most of the household appliances studied, with the exception of chargers and televisions. However, SHS users were of higher socioeconomic status than households that relied on kerosene or had no electricity at all. Therefore, SHS systems should be seen as a vital step on the energy ladder, which ultimately enables users to achieve better electricity supplies and higher levels of appliance ownership, thanks to the economic development they trigger.

3.1.3 Literature with focus on electricity demand

Different approaches to estimating electricity demand have emerged in the literature. Louw et al. (Louw et al., 2008) estimated the average demand of two rural villages in South Africa: by comparing two sets of survey-based data, they derived a set of heterogeneous input variables whose significance in predicting demand was tested using a log-linear regression model. The relevance of income as a driving factor led the authors to conclude that the use of electricity is a cost-based solution. Regression has also been adopted by Azadeh & Faiz (Azadeh & Faiz, 2011), in parallel with an Artificial Neural Network (ANN), to estimate the annual electricity consumption of Iranian households. During the validation phase, the superiority of the ANN-based model emerged. Another study using linear regression to estimate electricity demand was conducted by Dominguez et al. (Dominguez et al., 2021b), who studied the factors determining electricity consumption in rural Kenyan households, with particular attention to access dynamics. A separate model has been developed to estimate the probability of a household experiencing one of the following three energy transitions:

1. From no access to mains electricity
2. From not having access to the solar home system (SHS)
3. From SHS to mains electricity

The transition approach was also reflected in the choice of an additional dummy input variable, representing whether the household had access to an SHS before being connected to the electricity grid. Blodget et al. (Blodgett et al., 2017) instead used a typical bottom-up approach based on the survey: the average daily consumption was obtained starting from the time windows of possession and use of household appliances declared by the interviewees in eight Kenyan communities. The result was then compared with real consumption data, collected in the same communities after the electricity was supplied, thus highlighting that the estimates based on the survey would have led to an overestimation of demand by 330%. The authors concluded that survey-based sizing methodologies would lead to unacceptable bias and proposed an alternative proxy

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

approach. By conducting an analysis of variance (ANOVA) test, the authors demonstrated that, statistically, consumers from different communities belong to the same sample population. Thus, their consumption data could be interchanged, resulting in a much lower estimation error than the survey-based approach. The proxy methodology of the work is peculiar in that no distance metrics have been adopted to establish the similarities between the communities, thus simplifying the analysis and making it closely linked to the contingent data studied. The adoption of such metrics could broaden the applicability of the approach. Furthermore, as the authors state: "*If the datasets are widely available, these results suggest that mini-grid developers can use them to better predict consumption than the common survey approach*".

3.1.4 Literature with focus on load profile

Similarly to Blodgett et al. (Blodgett et al., 2017), in the study by Hartvigsson & Ahlgren (Hartvigsson & Ahlgren, 2018), the classic survey-based demand estimation was performed with a peculiarity: the interviews were collected in a Tanzanian village it had received electrification more than a decade earlier. Therefore customers, over time, had established patterns of use of household appliances; however, the usage patterns stated during the interviews greatly underestimated the actual load, which was directly measured and compared to the survey-based estimate. Although traditional approaches to demand estimation, based on bottom-up surveys, usually result in system poorly sized, innovative methodologies explored in the scientific literature show promising results. LoadProGen, a model developed by Mandelli et al. (Mandelli et al., 2016), consists of a stochastic generator of load profiles that starts from the same data pool as traditional approaches, i.e. the basket of foreseen appliances and their usage patterns. Thanks to its stochastic nature, LoadProGen can capture the high uncertainty associated with local data. The Remote-Areas Multi-energy systems load profiles (RAMP) model, developed by Lombardi et al. (Lombardi et al., 2019), is based on the same stochastic logic but also includes the estimation of other energy uses. RAMP has been validated on data from a Bolivian village, showing an improvement in performance compared to LoadProGen. However, both models still have the weakness of relying on surveyed data relating to the household appliances intended to be purchased once electricity has been supplied to the community. As already mentioned, this type of information has always proved unreliable in the past, therefore, the coupling of these bottom-up and stochastic load profile generators with models that estimate the ownership of appliances starting from verifiable local data could guarantee predictions more accurate than the question. RAMP was also adopted by Falchetta et al. (Falchetta et al., 2021) as a core component of the M-LED platform, an electricity demand estimation tool that collects multi-sector, bottom-up and highly granular data to generate local monthly load curves which can then be used to derive the less costly electrification strategy for rural communities.

The work done by Hernandez et al. (Hernández et al., 2014) showed how machine learning methods can be applied to the short-term load prediction of a micro-grid. The authors have developed a load profile forecasting procedure using as input only the past two years consumption data and the consumption calendar (to study the relationships between days and between months). Their prediction tool was built in three stages, each applying a different machine learning algorithm between a Self-Organizing Map (SOM) neural network, a k-means algorithm, and a Multi-Layer Perceptron (MLP) neural network. Machine learning was also the focus of a study by Dominguez et al. (Dominguez et al., 2021a) in which a chain of supervised and unsupervised models was trained to estimate the hourly lighting load profiles of rural households in Kenya and Tanzania. Model inputs were publicly available data at the household, village and county levels, combined with satellite imagery. It should be noted that the source used by the authors to estimate the type of lighting devices was also used to build the database which is the focus of this work.

Retrieving data from the load curves of 11 East African micro-grids over a two-year time horizon, Williams and Jaramillo (Williams et al., 2017) sought to identify weekly and

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

monthly seasonal trends in consumption, as well as patterns of long-term growth. No universal result has been achieved in this regard, although it has been possible to state an increase in demand over time for most systems. Further analysis of customers' payment preferences showed that, when left free to choose rate dynamics, customers have a clear tendency towards small and frequent payments. Proper modeling of the tariff scheme should therefore not neglect the frequency of payments. On a different note, the work of Lorenzoni et al. (Lorenzoni et al., 2020) has many aspects in common with this thesis. The authors created a database of load profiles from sixty-one mini-grid projects located in DCs. A clustering algorithm has been applied to the load curves, identifying a set of archetypal profiles; subsequently, an exploratory graphical analysis of the data was performed, to highlight the influence of some potential factors on the shape of the load profile: for example, it emerged that flat curves tended to be associated with large mini-grids, utility-owned and with a long operating history, while peak-dominated profiles characterized small, privately-owned projects.

3.1.5 Literature with focus on demand evolution

In their longitudinal study (Fobi et al., 2018) on the electricity consumption of Kenyan residential customers, Fobi et al. have sought to better understand the evolution of electricity demand over time, thus helping model developers in what has been identified in (Riva, Tognollo, et al., 2018) as the most neglected component of energy planning models. Their results showed an overall increase in consumption over time; however, large discrepancies in trends between urban and rural households have been highlighted (with the former experiencing larger increases), highlighting the need to include urban/rural characterization in future studies. To capture the endogenous relationship between socio-economic variables and electricity demand within a rural Tanzanian community, a System Dynamics model was developed and calibrated in Riva and Colombo (Riva & Colombo, 2020), to then be tested in a subsequent article (Riva, 2020). The model was built on the basis of dynamics that were made explicit using causal loop diagrams in a previous model conceptualization work (Riva, Ahlborg, et al., 2018): however, the overall dynamics can be summarized as a cycle of positive feedback between the electricity supply and different characteristics, mainly socio-economic, of the community.

3.1.6 Other drivers in literature

Most of the works presented concern more than one topic. A wide variety of models were encountered, from regression to clustering and *ad hoc* modeling. Starting with a critique of global energy models, which often operate at too aggregated a level to capture the heterogeneity of energy demand across households, van Ruijven et al. constructed a bottom-up model to estimate complete household energy uses (Ruijven et al., 2011). Among the outputs of the model were the needs for cooking food, water heating and space heating, but also the possession of household appliances and the demand for lighting. Outputs were linked, through integrated energy functions, to input variables both at the household level (*e.g.*, expenditure) and at the community level (e.g., population). The initial model was further developed in a work by Daioglou et al. (Daioglou et al., 2012) and was applied to five DCs, while the original paper only covered the Indian case. Van Ruijven's model was also the starting point for a work by Riva et al. (Riva et al., 2019), in which the dimensioning of an energy system for an Indian rural community has been comprehensively addressed through the soft connection of three different models. First, the evolution of household appliances ownership over 20 years (assumed as the lifetime of the equipment) was obtained by adapting the van Ruijven model, using data from field surveys. The property was then entered into LoadProGen to obtain the total annual load curves for each year of the simulation. Finally, the Open Source Energy Modeling System (OSeMOSYS) model (Howells et al., 2011) was used to obtain, through linear optimization, the least expensive energy supply mix. A completely different methodology was adopted by Fabini et al. (Fabini et al., 2014), who used the k-nearest neighbors regression to predict induced ownership, i.e. the expected increase in household appliances ownership in

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

households receiving electricity. Starting from a set of socio-economic characteristics, the authors defined multidimensional distance metrics to compare the constituencies of Kenya. The metrics were fed into a k-nearest neighbors algorithm to predict property levels in non-electrified districts by assessing their proximity to electrified ones. Finally, using the typical values of the daily electricity consumption for each appliance, the results of ownership were transformed into electricity demand. As the authors state, underpinning the entire process was a key assumption: 'Implicit in this approach is the assumption that locations that share socioeconomic characteristics will also have similar demand for electric services and similar ability to pay for them.' Thus, the authors' methodology can be framed in the proxy approach of (Blodgett et al., 2017) extended to consider multidimensional distance metrics between electrified and non-electrified communities. Limiting the metrics to a set of socio-economic characteristics could, however, lead to a loss of predictive power of the appliance ownership estimation model. Further broadening dimensionality, including also different types of features, could have a positive impact. Machine learning models have also been used by Allee et al. (Allee et al., 2021); their approach can be described as a middle ground between the classic survey-based approach and the proxy approach described in (Blodgett et al., 2017). The authors argued that although using survey data on planned appliance ownership leads to system mid-sizing, other information from field collections could be helpful to apply the proxy approach with a more precise rationale, i.e., again, using some form of distance metric to feed to an estimation model. Starting from this, they developed three different models to predict the electricity demand and the daily load profiles of the sites to be electrified:

1. An *Intercept-only* model, which was trained using only consumption data from already electrified micro-grids.

2. *LASSO*, a parametric linear machine learning model, integrating also information coming from local surveys.

3. A *Random Forest*, a non-parametric tree-based machine learning model, was trained on the same dataset of LASSO.

By demonstrating that the latter two models performed better than Intercept-only, the authors demonstrated that field survey campaigns can still be cost-effective for developers. In addition, they performed variable significance tests to reduce the amount of information to collect, saving time and money. They found that while socio-economic information was of relatively low importance, appliance ownership information was identified by both the LASSO and Random Forest models as critical to estimating demand. In contrast to the choice of previous works to test several explanatory variables, Shibano and Mogi (Shibano & Mogi, 2020) chose to analyze the impact of a single driver, creating an income-based model to estimate household electricity consumption. The ownership of electrical appliances (which was then related to electricity consumption using a Gamma distribution) was modeled using a Gompertz curve, the parameters of which were calibrated via regression.

Williams and Jaramillo (Williams et al., 2018) carried out an analysis similar to that of (Lorenzoni et al., 2020), while adopting a customer-based rather than a system-based point of view. 821 customers were segmented, using a k-means clustering algorithm, in terms of normalized average load curves and average daily electricity consumption. Through an exploratory graphical analysis of the data, an attempt was made to establish a relationship between these variables and the category of customers, *i.e.,* house, company, house and company, and public place. Four case studies of rural micro-grids with the same installed capacity were compared in Bahaj and James (Bahaj & James, 2019), both in terms of load profiles and daily electricity demand. In accordance with (Louw et al., 2008), they pointed out that the dynamics of electricity consumption are cost-based and revolve around the relative difference between consumers' income and the tariff amount. The experience of one of the systems has shown how the adoption of an advance

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

payment system has helped to solve the problem of late payments and abandonments, increasing consumption. The functioning of a Tanzanian micro-grid was analyzed both in the short and long term by Hartvigsson et al. (Hartvigsson et al., 2021). First, load profiles were analyzed to determine a variety of performance metrics such as load factor. Second, electricity charges (linked to electricity demand through tariffs) over 30 months were used to train the regression algorithms. This last phase highlighted a highly seasonal trend in consumption, perhaps linked to the fact that the local economy is heavily based on agriculture. Furthermore, the load profiles of the companies showed great heterogeneity according to the type of company (and therefore the basket of appliances used). The commercial connections should therefore not be neglected in the sizing of the system, but also their field of activity and their seasonality. Table 10 provides a list of references that addresses all the aforementioned drivers.

Table 10 - Classes of drivers identified in literature.

Publication	Socio-economic	Dwelling	Appliance	Past demand	Supply	Alternative energy sources	Geographical	Cultural
Literature with focus on Appliance Adoption								
(K. Lee et al., 2016)					X			
(Rao & Ummel, 2017)	X	X	X		X			X
(Kurata et al., 2018)	X	X	X			X	X	X
(Richmond et al., 2020)	X		X				X	X
Literature with focus on Electricity Demand								
(Louw et al., 2008)	X	X				X		
(Azadeh & Faiz, 2011)	X				X			
(Blodgett et al., 2017)	X							
(Dominguez et al., 2021b)	X				X	X		
Literature with focus on Load Profile								
(Hernández et al., 2014)				X				

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Publication	Socio-economic	Dwelling	Appliance	Past demand	Supply	Alternative energy sources	Geographical	Cultural
(Mandelli, Merlo, et al., 2016)	X		X					
(Williams et al., 2017)				X	X			
(Hartvigsson & Ahlgren, 2018)			X					
(Lombardi, Balderrama, et al., 2019)	X		X					
(Lorenzoni et al., 2020)	X			X	X		X	
(Dominguez et al., 2021a)	X	X	X	X	X			
(Falchetta et al., 2021)	X	X	X	X	X			
Literature with focus on Demand Evolution								
(Fobi et al., 2018)	X				X			
(Riva, Ahlborg, et al., 2018)	X		X	X	X	X	X	X
(Riva & Colombo, 2020)	X		X	X	X	X	X	X
(Riva, 2020)	X		X	X	X	X	X	X
Literature with Multiple Foci								
(van Ruijven et al., 2011)	X	X				X	X	
(Daioglou et al., 2012)	X	X					X	
(Fabini et al., 2014)	X	X	X		X	X		
(Williams et al., 2018)	X			X				
(Bahaj & James, 2019)	X				X			

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Publication	Socio-economic	Dwelling	Appliance	Past demand	Supply	Alternative energy sources	Geographical	Cultural
(Riva, Gardumi, et al., 2019)	X	X	X					
(Shibano & Mogi, 2020)	X							
(Allee et al., 2021)	X	X	X			X		X
(Hartvigsson et al., 2021)	X			X				
Total								
	25	10	14	10	14	9	8	7

The class of socio-economic drivers emerges as the most adopted in the literature reviewed, followed by drivers related to the techno-economic parameters of the appliances available to users and to the technology that supplies electricity to users. While cultural drivers seem to be the least considered, probably due to the complexity of collecting and classifying this class of drivers and the engineering background that often led to the exclusion of this analysis category from previous works (Sovacool, 2014).

In the appliance ownership literature, as well as in the load profile literature, almost all classes of drivers are involved in the studies. Furthermore, in the appliance ownership group, this comprehensive coverage is also found within the documents, excluding (K. Lee et al., 2016). This is a difference from the Load Profile group, where most items revolved around a small number of categories. Similarly, the Demand Evolution group papers show near-total coverage of factor categories.

3.1.7 Innovative methodologies for demand estimation

Kuster et al. (Kuster et al., 2017) reviewed the literature on demand estimation and highlighted that regression, neural networks, machine learning and bottom-up methodologies represent the major classes of approaches for estimating demand, as highlighted in Figure 50. More recent studies have enriched the field and attempted to increase accuracy of methodologies, yet relying on the same class of artificial intelligence (AI) techniques as mentioned in the figure. However, as recent improvements, it is worth mentioning the recent trend of creating data-driven models (Lorenzoni et al., 2020), social-dynamics based approaches (Riva & Colombo, 2020) and GIS-enabled spatial methodologies (Falchetta et al., 2021), which do not fit in the classification by Kuster et al. (Kuster et al., 2017). The literature reviewed in the previous sections, instead, generally employed regression methodologies, e.g. k-means (Williams et al., 2018), LASSO or Random Forest (Blodgett et al., 2017), Neural Network approaches (Dominguez et al., 2021a), or survey-based bottom-up models (Lombardi et al., 2019).

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

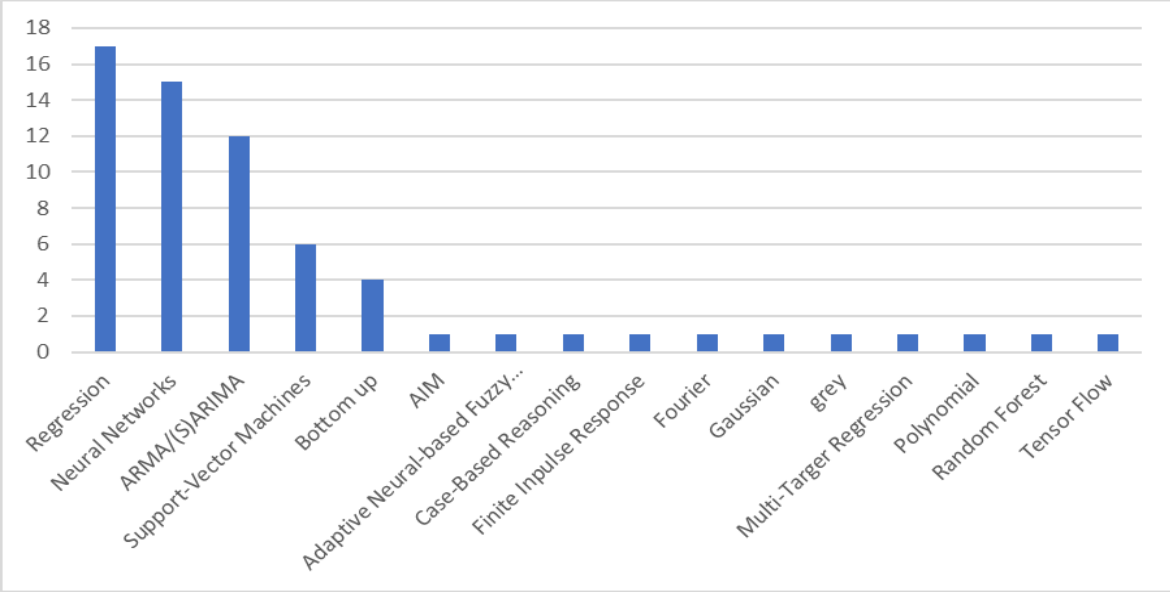


Figure 50 - Methodologies for demand estimation, adapted from (Kuster et al., 2017).

The major input parameters that load assessment models have used or reported to be relevant for energy assessment are reported in Table 11. The table classifies the major driver considered in the literature for demand estimation and classifies them by class, in agreement with the classes referred in the previous sections. Moreover, it is also reported what entity that input quantity is referred to, be them Village (V), Household (H) and Appliance (A). These inputs can be used as reference for calibrating surveys for energy assessment to be used by demand estimation methods, as reported in Figure 50.

Table 11 - Major driver list to be used for demand estimation studies, grouped by type of driver and entity the driver is linked to (V: Village, H: Household and A: Appliance).

Id	Driver name	Driver class	Entity	References
1	Connection type or fee	Supply	V	(Allee et al., 2021, Dominguez et al., 2021a, Falchetta et al., 2021, Lee et al., 2016, Riva, 2020, Riva and Colombo, 2020)
2	Electricity price	Supply	H	(Azadeh and Faiz, 2011, Bahaj and James, 2019, Dominguez et al., 2021b, Riva, 2020, Riva and Colombo, 2020)
3	Installed power capacity of the system	Supply	V, H	(Lorenzoni et al., 2020)
4	Hours of available electricity	Supply	H	(Dominguez et al., 2021a, Rao and Ummel, 2017, Riva, 2020, Riva and Colombo, 2020, Riva et al., 2018)
5	Pre-paid/post-paid tariff (tariff scheme)	Supply	V, H	(Bahaj and James, 2019, Lorenzoni et al., 2020)

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Id	Driver name	Driver class	Entity	References
6	Tariff type, method or frequency	Supply	V, H	(Lorenzoni et al., 2020, Pinomaa, 2021, Williams et al., 2018)
7	Time passed since the electrification	Supply	V, H	(Fobi et al., 2018, Louw et al., 2008, Richmond et al., 2020, Williams et al., 2017)
8	System management (business model)	Supply	V	(Lorenzoni et al., 2020)
9	Number of connections of the system	Supply	V	(Lorenzoni et al., 2020)
10	Prospect of grid access in the next year	Supply	V, H	(Kurata et al., 2018)
11	Share of households by energy access (grid, SHS, etc.)	Supply	V	(Dominguez et al., 2021b)
12	Capacity building implemented	Supply	V, H	(Morganti, 2021, Riva, 2020, Riva and Colombo, 2020, Riva et al., 2018)
13	Urban/rural location (GPS location proxy)	Socio-economic	V	(Falchetta et al., 2021, Fobi et al., 2018, Rao and Ummel, 2017)
14	Recently affected by a natural disaster (GPS location proxy)	Socio-economic	V	(Kurata et al., 2018)
15	Employment (Working hours proxy)	Socio-economic	H	(Fabini et al., 2014, Falchetta et al., 2021)
16	Nighttime lights from satellite imagery (GPS location proxy)	Socio-economic	V	(Dominguez et al., 2021a, Falchetta et al., 2021)
17	Presence of a household business (Working hours proxy)	Socio-economic	H	(Dominguez et al., 2021b, Riva, 2020, Riva and Colombo, 2020, Riva et al., 2018)
18	Size of business activity in terms of employees	Socio-economic	H	(Kurata et al., 2018)
19	Age of household head	Socio-economic	H	(Allee et al., 2021, Dominguez et al., 2021a,b, Kurata et al., 2018, Rao and Ummel, 2017)
20	Education level of household head	Socio-economic	H	(Allee et al., 2021, Dominguez et al., 2021a, Kurata et al., 2018, Rao and Ummel, 2017)
21	Income of the household	Socio-economic	H	(Allee et al., 2021, Bahaj and James, 2019, Falchetta et al., 2021, Kurata et al., 2018, Louw et al., 2008, Rao and Ummel, 2017, Riva

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Id	Driver name	Driver class	Entity	References
				et al., 2018, Shibano and Mogi, 2020)
22	Number of people/adults/youngsters/elderly in the household	Socio-economic	H	(Allee et al., 2021, Dominguez et al., 2021b, Kurata et al., 2018, Louw et al., 2008, Rao and Ummel, 2017, Richmond et al., 2020, Riva et al., 2019)
23	Number of working people within the household	Socio-economic	H	(Allee et al., 2021, Fabini et al., 2014)
24	Size of business activity in terms of revenues	Socio-economic	H	(Kurata et al., 2018)
25	Age of business activity	Socio-economic	H	(Stevanato et al., 2019)
26	Number of years the household lived in the community	Socio-economic	H	(Allee et al., 2021)
27	Population density	Socio-economic	V	(Daioglou et al., 2012, Dominguez et al., 2021a)
28	Ratio of females within the household	Socio-economic	H	(Kurata et al., 2018)
29	Ratio of household members with high education level	Socio-economic	H	(Kurata et al., 2018)
30	Business hours of micro-enterprise	Socio-economic	H	(Kurata et al., 2018)
31	Cooking fuel usages	Socio-economic	A	(Allee et al., 2021)
32	Information on the owner of micro-enterprise	Socio-economic	H	(Kurata et al., 2018)
33	Monthly energy expenditure of household	Socio-economic	H	(Allee et al., 2021)
34	Ownership of a SHS before grid connection	Socio-economic	H	(Dominguez et al., 2021b)
35	Access to credit in the past	Socio-economic	H	(Kurata et al., 2018, Louw et al., 2008, Riva, 2020, Riva and Colombo, 2020, Riva et al., 2018)
36	Dimension of land owned	Socio-economic	V	(Falchetta et al., 2021, Kurata et al., 2018)
37	Monthly household expenditure	Socio-economic	H	(Daioglou et al., 2012, Richmond et al., 2020, Riva, 2020, Riva and Colombo, 2020, Riva et al., 2018, van Ruijven et al., 2011)

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Id	Driver name	Driver class	Entity	References
38	Ownership of large livestock	Socio-economic	H	(Dominguez et al., 2021a,b)
39	Ownership of motorized vehicle	Socio-economic	H	(Allee et al., 2021, Dominguez et al., 2021a,b, Rao and Ummel, 2017)
40	Ownership of small livestock	Socio-economic	H	(Dominguez et al., 2021a,b)
41	Presence of street lighting in the neighborhood	Socio-economic	V, H	(Dominguez et al., 2021a)
42	Respondent category (Household, business, etc.)	Socio-economic	H	(Hartvigsson et al., 2021, Lombardi et al., 2019, Lorenzoni et al., 2020, Mandelli et al., 2016, Williams et al., 2018)
43	Seasonality of business activity	Socio-economic	H	(Hartvigsson et al., 2021)
44	Sex of household's head	Socio-economic	H	(Allee et al., 2021, Dominguez et al., 2021a,b, Kurata et al., 2018, Rao and Ummel, 2017, Richmond et al., 2020)
45	Size of public service in terms of area occupied	Socio-economic	H	(Colombelli, 2019, F., 2021)
46	Socio-economic status of household head	Socio-economic	H	(Allee et al., 2021, Dominguez et al., 2021a,b, Kurata et al., 2018)
47	Type of business activity	Socio-economic	H	(Allee et al., 2021, Hartvigsson et al., 2021, Lombardi et al., 2019, Mandelli et al., 2016)
48	Type of public service	Socio-economic	H	(Falchetta et al., 2021, Lombardi et al., 2019, Mandelli et al., 2016)
49	Monthly electricity consumption	Past-demand data	V, H, A	(Morganti, 2021)
50	Monthly electricity expenditure	Past-demand data	H	(van Ruijven et al., 2011)
51	Climate zone	Geographical	V	(Lorenzoni et al., 2020)
52	Distance from the nearest city	Geographical	V	(Falchetta et al., 2021, Richmond et al., 2020)
53	Altitude	Geographical	V	This deliverable.
54	Agriculture-related information	Geographical	H	(Falchetta et al., 2021)
55	Floorspace of the dwelling	Dwelling	H	(Daioglou et al., 2012, Riva et al., 2019, van Ruijven et al., 2011)
56	Number of rooms of the	Dwelling	H	(Allee et al., 2021, Daioglou et al.,

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Id	Driver name	Driver class	Entity	References
	dwelling			2012, Dominguez et al., 2021b, Kurata et al., 2018, Louw et al., 2008, Rao and Ummel, 2017, Riva et al., 2019)
57	Ownership status of the dwelling	Dwelling	H	(Allee et al., 2021, Rao and Ummel, 2017)
58	Quality of the dwelling (roof/wall material, etc.)	Dwelling	H	(Allee et al., 2021, Dominguez et al., 2021a,b, Fabini et al., 2014, Rao and Ummel, 2017)
59	Marital status of household head	Cultural	H	(Allee et al., 2021, Kurata et al., 2018)
60	Religion or Race	Cultural	V, H	(Rao and Ummel, 2017, Richmond et al., 2020)
61	Number of hours studied per night by school children	Cultural	H	(Allee et al., 2021)
62	Bedtime and waketime	Cultural	H	(Allee et al., 2021)
63	Presence of a smoking person within the household	Cultural	H	(Kurata et al., 2018)
64	Price of appliance	Appliance	A	(Azadeh and Faiz, 2011, Rao and Ummel, 2017)
65	Affordability of appliance	Appliance	V, H	(Rao and Ummel, 2017)
66	Planned appliances	Appliance	H	(Allee et al., 2021)
67	Price of energy sources	Alternative sources	A	(Allee et al., 2021, Dominguez et al., 2021b, Fabini et al., 2014, Kurata et al., 2018, Louw et al., 2008, Riva and Colombo, 2020, Riva et al., 2018)
68	Main lighting fuel	Alternative sources	H	(Fabini et al., 2014)
69	Type of main cooking fuel	Alternative sources	H	(Allee et al., 2021, Dominguez et al., 2021a,b)

3.2 Data Scarcity and load demand estimation

The context of rural areas of developing countries, to which SETaDiSMA refers to, is particularly affected by the issue of data paucity. It is, in general, an extremely complex task to retrieve data from the field to estimate the future energy demand of rural communities, and there is also a general lack of available tools from the scientific and practitioners' community. The scientific community is well aware of this issue and of the relevance of properly estimating load demand given its impact on system sizing and long-term sustainability. As a result, the studies analyzed in Section 3.1 were developed and produced a large numer of proposed methodologies and useful insights for future steps.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

In fact, bottom-up approaches and models for characterization of load demand in rural areas exist (*e.g.*,: (Lombardi et al., 2019; Mandelli et al., 2016), but they rely on the assumption that high resolution data on appliances ownership and usage patterns exist for the study area. This is unfortunately often not true in developing countries, as Debnath and Mourshed highlight, data paucity is one of the main challenges for energy modelling in developing countries (Debnath & Mourshed, 2018). Secondly, often pre-electrification studies are used to feed such models, but they proved to be empirically inaccurate, given the intrinsic problem in the use of the tool itself, that aims at predicting uses of electricity by surveying individuals that never experienced electricity before (Blodgett et al., 2017).

This document, building on the observed literature, proposes two different approaches that can be adopted to estimate the load demand of rural communities of developing countries yet to be connected to electricity services, with two different levels of computational and data intensity.

3.2.1 Archetypes of load demand for rural communities

This chapter proposes a first approach for improving load demand characterization in off-grid energy planning. It aims at developing a simple but solid methodology, with a geographical applicability to the entire Sub-Saharan Africa that proves to be of easy access for the entire community of off-grid energy planners, including government planning bodies and NGOs. The proposed approach is a set of load archetypes for household, health facilities and school categories of users, differentiated by wealth tier, latitude and climate zone. The archetypes can be applied for off-grid system planning at local level (e.g.: sizing of a single mini-grid), national strategy development (e.g.: prioritization of areas for infrastructure development) and regional level (e.g.: Sub Saharan level analysis for policy support), and examples of applications at all levels are proposed.

In order to develop a technique that allows to cover the entire Sub-Saharan Africa (SSA), we developed an archetype-based load estimation approach, synthesized in Figure 51, to be integrated with mini-grid sizing models.

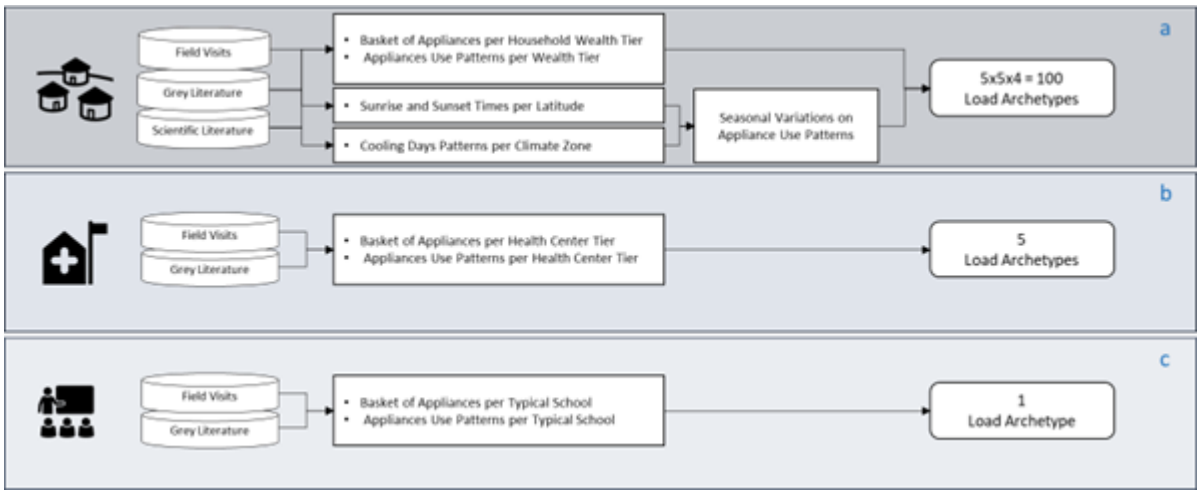


Figure 51 - Schematic framework of the methodology for the creation of the load archetypes for a) the households, b) the health centers and c) the school.

For the Households, (Figure 51a – top), we identify three main drivers of variation of demand, namely *i)* wealth level of the household, *ii)* latitude of the village and *iii)* cooling days of the area. Each of the three drivers will independently determine a variation in the load curve. The same concept was developed in seminal form in (Falchetta et al., 2021), and applied to the sole country of Kenya, reaching a satisfying level of coherence of the methodology's outputs with the existing literature.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

In (Falchetta et al., 2021) only the driver i) wealth level of the household was considered, determining a variation in the basket of appliances used by the 5 different categories of users. Here we expand the concept in order to cover a broader spectrum of users across the entire sub-continent.

A. ***Wealth Tier of the household***: the users' categories are divided according to their wealth level, and each level is considered to own a different basket of appliances, created starting from a systematic review of the literature on electricity use in rural SSA (Ciller & Lumbreras, 2020; Dagnachew et al., 2018; Huld et al., 2017; Kotikot et al., 2018; Zeyringer et al., 2015). The wealthier classes own more energy intensive appliances and in larger number, are inspired from but not based on the World Bank's Multi-Tier Framework (ESMAP, 2015). Such parameters are represented with columns in Table 12.

B. ***Latitude of the village***: as the approach needs to be valid in all SSA, villages at different latitudes need to be modelled, as the latitude will determine different sunrise and sunset hours, this parameter determines a seasonal change in appliances times of use. Such parameters are represented with lines in Table 12.

C. ***Cooling days of the area***: for the same reason as in point ii), different climates can be found in different areas of SSA, and the need for cooling in the households will happen at different moments of the years. As suggested by Falchetta and Mistry (Falchetta & Mistry, 2021) cooling needs will be a key load in the near future of the continent, and this parameter determines the seasonality of their use in the households. Such parameters are represented by NC (No Cooling Days), AY (All Year Cooling), OM (Cooling Days from October to March) and AS (Cooling Days from April to September) in Table 12.

Table 12 - Archetypes of load demand developed, combining the three dimensions of latitude, wealth level and climate.

	Wealth 1	Wealth 2	Wealth 3	Wealth 4	Wealth 5
Lat [10;20]	NC|AY|OM|AS	NC|AY|OM|AS	NC|AY|OM|AS	NC|AY|OM|AS	NC|AY|OM|AS
Lat [10;-10]	NC|AY|OM|AS	NC|AY|OM|AS	NC|AY|OM|AS	NC|AY|OM|AS	NC|AY|OM|AS
Lat [-10;-20]	NC|AY|OM|AS	NC|AY|OM|AS	NC|AY|OM|AS	NC|AY|OM|AS	NC|AY|OM|AS
Lat [-20;-30]	NC|AY|OM|AS	NC|AY|OM|AS	NC|AY|OM|AS	NC|AY|OM|AS	NC|AY|OM|AS
Lat [< -30]	NC|AY|OM|AS	NC|AY|OM|AS	NC|AY|OM|AS	NC|AY|OM|AS	NC|AY|OM|AS

For the Health Centres, (Figure 51b – center), based on past work (Falchetta et al., 2021), we develop 5 different Tiers of Health facilities, from Rural Dispensary (Tier 1) up to Sub-County Hospital (Tier 5). For the School (Figure 51c – bottom), also based on the same previous study (Falchetta et al., 2021), we propose the archetypical load of a rural primary school.

Through the proposed approach we developed 100 (5x5x4) archetypes of household users, characterized by different sets of appliances (wealth parameter), seasonal variations in the

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

time of use of the appliances (latitude parameter) and different seasonal use of ambient cooling devices (climate zone parameter): 5 Health Facilities archetypical loads based on the kind of Health Centre, and 1 archetypical load for a rural primary school. Such user classes are then used to feed the bottom-up stochastic load curve generator model RAMP (Lombardi, Balderrama, et al., 2019).

3.2.2 Database of observed load demand trends

Improving on the archetype concept, it is possible to build a database of observed consumption patterns in recently electrified communities to better inform the load prediction, based on the proximity approach proposed by (Fabini et al., 2014) and expanded to the purpose of this research. The database can have the double role of serving as data source per se, providing harmonized and reliable data to energy planners and can represent the training basis for load demand prediction algorithms.

The database can be developed correlating the adoption and use of electric appliances in households, small businesses and public services in recently electrified areas with a set of socioeconomic parameters, geographical and climatic factors, like the ones provided in Table 12 of this document. It can serve as informative platform and future reference for linking the context of intervention with a set of expected appliances and their patterns of use by the different users. The database can furthermore serve as basis for training an Algorithm for Load Estimation of non-served areas, taking advantage of the identified correlations between drivers of adoption and appliance ownership and use. This will fill the gap of the missing input data for bottom-up load estimation tools.

Following the statement of Blodgett et al. (Blodgett et al., 2017) on the research potential that could be offered by large open databases containing information on electrification case studies, the first goal of this work was to set-up and structure a first version of such a database. To widen its applicability, the database should be constructed by harmonizing different data sources, thus allowing for a wide range of comparisons and analyses. Notably, the source heterogeneity will allow practitioners to assess the validity of their research results among contexts that may differ in terms of spatial scale, type of intervention, geographic location, socio-economic and cultural characteristics of the studied communities. To be useful, the database should therefore contain a variety of information on electrified users from DCs. Its open-access nature, besides, will allow future practitioners and developers to enrich it with new data, thus widening its potential.

The product will have many characteristics in common with the database developed by Lorenzoni et al. (Lorenzoni et al., 2020) on micro-grids. However, some key differences should be pointed out:

- Rather than being a system-based tool, the database will be customer-based. In this way, even when only uncomplete information about a given community will be available, the aggregation of the data with other instances from the database will allow performing analyses. Panel surveys collecting only sparse samples from single communities will therefore also be included.

- Following the customer-based approach, a differentiation between types of users (household/business/public service) will be present.

- To widen the scope of the tool, the type of connections analyzed will not be limited to micro-grids but will also consider grid-connected and SHS users.

- The study focus will be different, shifting from load profiles to appliance ownership.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

- The database will focus on Sub-Saharan Africa, neglecting at this stage other world regions.

The choice of pivoting the database on appliance data (i.e., ownership, number, usage patterns), is justified by the fact that they constitute the key information needed for bottom-up survey-based modelling approaches. When practitioners and developers have tried to assess appliance info through pre-electrifications surveys, the results have consistently proven to be unreliable (Blodgett et al., 2017; Hartvigsson & Ahlgren, 2018). However, not all the information collected on-field has the same degree of uncertainty: e.g., the household size or the number of rooms can be asserted with more certitude than the future time windows of usage of an appliance that, at the moment of the survey, may not even have been purchased yet. The database could therefore be employed to establish statistical relationships between those highly verifiable data and the target appliance variable, using data coming from already electrified customers. In this sense, the database's main field of use will be using the proxy approach described by Blodgett et al. (Blodgett et al., 2017). Similarly to the work conducted by Fabini et al. (Fabini et al., 2014), ad-hoc metrics could be employed to assess the multidimensional distance between users within the database and soon-to-be-electrified ones. Thus, the statistical relationships obtained within the database could be extended to obtain reliable appliance ownership estimates for not electrified users. The presence, within the database, of heterogeneous information from the eight different factor classes will allow to explore more potential correlations and increase the multidimensionality of the metrics with respect to the study by Fabini et al. which considered only socio-economic variables.

The developed archetypes, integrated in a mini-grid sizing tool, are openly available on the GitHub Repository: https://github.com/SESAM-Polimi/Micro-gridsPy-SESAM/tree/Micro-gridsPy-2.0

3.3 Demand Validation

3.3.1 Archetypes: validation of appliance adoption

The validation of the constructed demand archetypes has been conducted relying on data collected in Rwanda for 60 rural households from 7 villages. Along with a detailed list of owned electrical appliances by each user, income data has been collected. This allows to perform a wealth-specific validation of archetypical appliance adoption against real observations. In order to link each household to one of the five wealth tiers composing the archetypes, statistical data about rural income at national level have been retrieved (EICV 3 THEMATIC REPORT - Income | National Institute of Statistics Rwanda, 2012). These data provided a division by income quintiles based on a sample of 2253 observations, of which only 10% related to urban households living in Kigali city. Assuming that rural households are the most represented in the sample, the quintile division has been linked to the archetypical division in wealth tiers, Table 13. Each of the 60 households has been then matched to an archetypical wealth tier. The resulting distribution of households is reported in Table 14.

Table 13 - Division of income quintiles for Rwanda (EICV 3 THEMATIC REPORT - Income | National Institute of Statistics Rwanda, 2012].

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

	Consumption	Income	Ratio	No. of HHs (000s)
All Rwanda	308,993	289,337	1.07	2,253
Kigali City	936,761	950,342	0.99	223
Southern Province	248,610	184,651	1.35	549
Western Province	220,385	217,873	1.01	528
Northern Province	232,378	247,613	0.94	411
Eastern Province	255,659	223,945	1.14	542
Q1	119,027	81,689	1.46	600
Q2	152,697	114,127	1.34	468
Q3	197,380	147,480	1.34	429
Q4	269,095	216,627	1.24	390
Q5	992,427	772,757	1.28	367

Table 14 - Distribution of households according to collect income data and archetypical wealth tiers.

Wealth tier (WT)	Number of Households
1	48
2	5
3	6
4	0
5	0

The cooling behavior assumed is "No Cooling Days", i.e. no fans are assumed to be owned in all income tiers: this choice is justified by the analysis of hourly temperature profiles for the year 2020, retrieved from PVGIS (Šúri et al., 2005), for three villages of the sample. Yearly average temperatures are 22°C for Kabuye village, 21°C for Rutenderi and 18°C for Nyakabanda. Temperature surpasses 30°C only for Kabuye village for 17 days in the year (Figure 52).

In order to compare the adoption of the *i*-th appliance between the archetypical and observed *j*-th household, the mean bias error (MBE_i) on the number *N* of adopted units has been calculated for each wealth tier *T* as:

$$MBE_{i,T} = \frac{1}{n}\sum_{j=1}^{n}(N_{ij,T}^{arch} - N_{ij,T}^{obs}) \quad (13)$$

Negative values of $MBE_{i,T}$ highlight an underestimation in the archetypical number of the *i*-th appliance for tier *T*. Since observed data highlight the presence of appliances not considered by the archetypes (e.g. blender, microwave oven, stove electric coil), this leads systematically to negative MBEs for such appliances.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

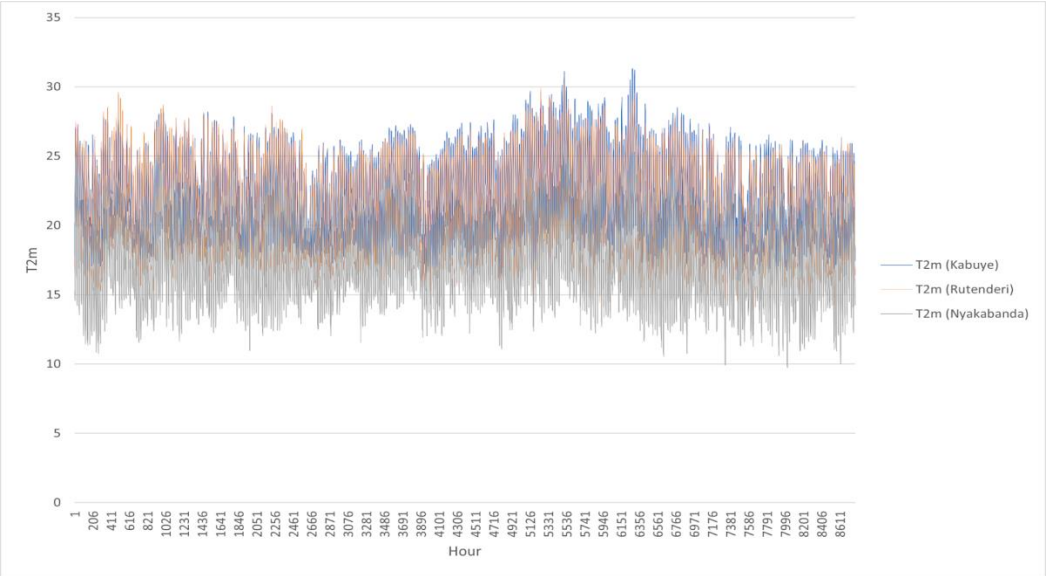


Figure 52 - Hourly temperature profile at 2 meters for Kabuye village, Rutenderi and Nyakabanda.

The obtained results are plotted in Figure 53 and show in general underestimation of the number of appliances, notably light bulbs and radios. Archetypes indeed estimate on average 2 light bulbs less than observed for tier 1 and 1.5 for tier 3. Radios are underestimated by 1 unit for tier 1 and 0.5 units for tier 3. Tiers 2 and 3 show the adoption of appliances not included by the archetypes: irons, kettles, pressure cookers, blenders, microwaves and electric coils. For tier 3, the underestimation of irons (-0.7) and electric pressure cookers (-0.5) is relevant due to the high power consumption of such appliances. These results however in a low impact due to the 10% share of tier 3 households in the sample.

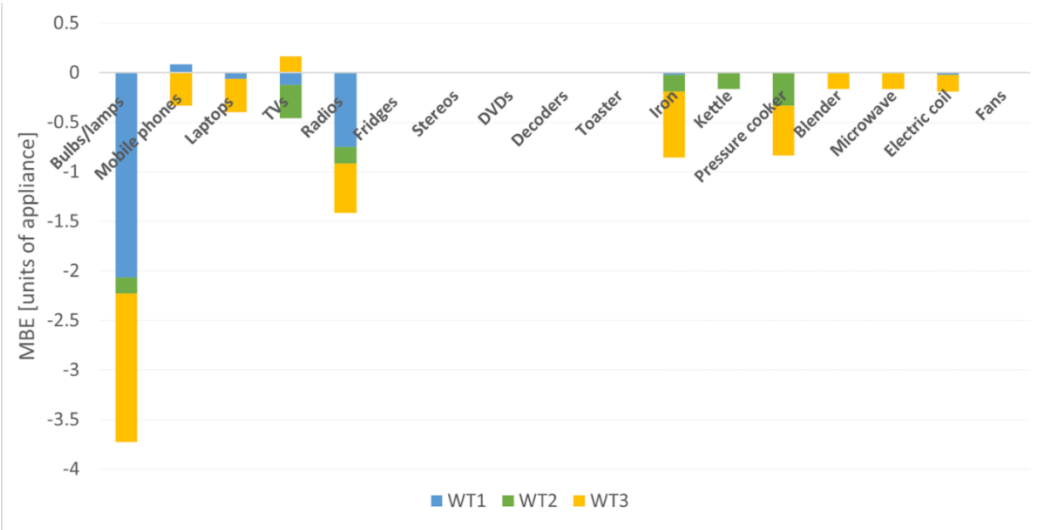


Figure 53 - MBE between archetypical and observed appliance adoption according to the different wealth tier assumed in this work.

In conclusion, even though the inclusion of non-predicted appliances turns out to be necessary, the approximation provided by the archetypes can be considered acceptable for the analysed sample. A larger dataset is however required to support and have a more robust validation.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

4. Final remarks

The work presented in this report is part of the activities under development in SETaDiSMA project and contributes to the identification of the most adequate databases for wind and solar power estimation as well as the most adequate methodology for load demand estimation. This information is crucial to properly design and plan a micro-grid investment and to propose a first guideline to optimally size micro-grid projects given the socio-economic information of the site, and perform validation on real case studies.

The main databases publicly available to support the wind and solar resource assessment procedures on the regions under study were presented, as well as validation of these databases using the few ground-based stations for which data was available. Seven locations were analyzed in the case of wind resource assessment. Results show that database's quality highly depends on the terrain and roughness classes. The obtained results are expected to occur since databases uses have a coarse spatial resolution to represent the variability of the wind speed and direction over complex terrain or locals with medium/high roughness. Therefore, ideally, to properly address the wind power integration into mini-grids/national power systems, databases for wind speed and direction characterization with hourly (or sub-hourly) resolution and high spatial resolution covering the African continent should be made publicly available. Nevertheless, results suggest the NASA-MERRA2 database provide a slightly better representation of the wind resource in comparison with ECMWF-ERA5.

For the solar resource assessment, based on the three stations analyzed as well as in the literature review, it is concluded that the databases' publicly available already provide GHI data with high accuracy. In specific, for the regions under analysis, the Copernicus Atmospheric Monitoring Service (CAMS) database presented the more accurate results and is, therefore, the most suitable for the work under development.

From the review conducted in this framework that the most adopted techniques for load estimation are two. i) Top-down approaches, in which the load is assigned to categories of users according to fixed parameters, usually wealth related. ii) Bottom-up approaches, in which the load curve is formulated based on specific characteristics of the user, by defining the energy consumption behaviour of every specific user.

In the context of this work, a ready-to-use tool for load estimation is proposed, validated and made available through a GitHub repository, in the form of load consumption archetypes. The archetypes are conceived for energy modelling purposes specifically, and are in fact presented in an integration with a mini-grid sizing tool. The purpose of the archetypes is that of providing a tool for load estimation in the case of scarcity of data on the context of operation. A more data-intensive approach is also presented, in the form of a database correlating socio-economic data with appliance adoption and use patterns, this database would have the scope of allowing to generate inputs for load curve simulation tools based on common socio-economic and cultural traits among villages.

More available data from the visited mini-grids in the future might allow to better validate and formulate both the approaches proposed.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

References

Africa Union Comission. (2015). Agenda 2063. Tech. Rep. Retrieved from https://au.int/sites/default/files/documents/41480-doc-2nd_Continental_Progress_Report_on_Agenda_2063_English.pdf

Allee, A., Williams, N. J., Davis, A., and Jaramillo, P. (2021). Predicting initial electricity demand in off-grid Tanzanian communities using customer survey data and machine learning models. Energy for Sustainable Development, 62, 56-66. https://doi.org/10.1016/j.esd.2021.03.008

Azadeh, A., and Faiz, Z. S. (2011). A meta-heuristic framework for forecasting household electricity consumption. Applied Soft Computing, 11(1), 614-620. https://doi.org/10.1016/j.asoc.2009.12.021

Bahaj, A. S. and Patrick A. B. James. Electrical mini-grids for development: Lessons from the field. Proceedings of the IEEE, 107(9):1967–1980, 9 2019. ISSN 0018-9219, 1558-2256. doi: 10.1109/JPROC.2019.2924594.

Balderrama, S., Lombardi, F., Riva, F., Canedo, W., Colombo, E., and Quoilin, S. (2019). A two-stage linear programming optimization framework for isolated hybrid micro-grids in a rural context: The case study of the "El Espino" community. Energy, 188, 116073. https://doi.org/10.1016/j.energy.2019.116073

Berrisford, P., Dee, D. P., Poli, P., Brugge, R., Fielding, K., Fuentes, M., … Simmons, A. (2011). The ERA-Interim archive Version 2.0. ERA Report Series. Shinfield Park, Reading: ECMWF.

Blanc, P., and Wald, L. (2015). L'estimation du rayonnement solaire au sol par al nouvelle méthode Heliosat-4. Lá Météorologie, 90, 53-61.

Blodgett, C., Dauenhauer, P., Louie, H., and Kickham, L. (2017). Accuracy of energy-use surveys in predicting rural mini-grid user consumption. Energy for Sustainable Development, 41, 88-105. https://doi.org/10.1016/j.esd.2017.08.002

Bosilovich, M., Akella, S., Coy, L., Cullather, R., Draper, C., Gelaro, R., … Suarez, M. (2015). MERRA-2?: Initial Evaluation of the Climate. In NASA/TM-2015-104606/Vol. 43, NASA Technical Report Series on Global Modeling and Data Assimilation (p. 139).

Brower, M. C., Bailey, B. H., Doane, J., and Eberhard, M. J. (2012). Wind Resource Assessment: A Practical Guide to Developing a Wind Project - Michael Brower. (M. C. Brower, Ed.). Wiley.

Ciller, P., and Lumbreras, S. (2020). Electricity for all: The contribution of large-scale planning tools to the energy-access problem. Renewable and Sustainable Energy Reviews, 120, 109624. https://doi.org/10.1016/j.rser.2019.109624

Colombelli, Fabrizio. Technical and socio-economic sustainability of rural electrification in Tanzania : the case of Ninga small hydropower plant project. PhD thesis, 2019. URL http://hdl.handle.net/10589/148847. [Online; accessed 2022-03-24].

Couto, A., Costa, P., Cardoso, J., Simões, T., Rodrigues, C., and Cuamba, B. (2022). Standalone solar cooking appliance design metrics - v2. LEAP-RE: WP 10 PURAMS Deliverable.

Dagnachew, A. G., Lucas, P. L., Hof, A. F., and Vuuren, D. P. van. (2018). Trade-offs and synergies between universal electricity access and climate change mitigation in Sub-Saharan Africa. Energy Policy, 114(December 2017), 355-366. https://doi.org/10.1016/j.enpol.2017.12.023

Daioglou, V., Bas J. van Ruijven, and Detlef P. van Vuuren. Model projections for household energy use in developing countries. Energy, 37(1):601–615, 1 2012. ISSN 03605442. doi: 10.1016/j.energy.2011.10.044.

Daioglou, V., van Ruijven, B. J., and van Vuuren, D. P. (2012). Model projections for household energy use in developing countries. Energy, 37(1), 601-615. https://doi.org/10.1016/j.energy.2011.10.044

Debnath, K. B., and Mourshed, M. (2018). Challenges and gaps for energy planning models in the developing-world context. Nature Energy, 3(3), 172-184. https://doi.org/10.1038/s41560-018-0095-2

DeCarolis, J., Daly, H., Dodds, P., Keppo, I., Li, F., McDowall, W., Pye, S., Strachan, N., Trutnevyte, E., Usher, W., Winning, M., Yeh, S., and Zeyringer, M. (2017). Formalizing best practice for energy

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

system optimization modelling. Applied Energy, 194, 184-198. https://doi.org/10.1016/j.apenergy.2017.03.001

DER-CAM. (n.d.). https://building-micro-grid.lbl.gov/projects/der-cam

Díaz, P., Arias, C. A., Peña, R., and Sandoval, D. (2010). FAR from the grid: A rural electrification field study. Renewable Energy, 35(12), 2829-2834. https://doi.org/10.1016/j.renene.2010.05.005

Doddy Clarke, E., and Sweeney, C. (2022). Solar energy and weather. Weather, 77(3), 90-91. https://doi.org/10.1002/wea.4124

Doelling, D. R., Loeb, N. G., Keyes, D. F., Nordeen, M. L., Morstad, D., Nguyen, C., … Sun, M. (2013). Geostationary Enhanced Temporal Interpolation for CERES Flux Products. Journal of Atmospheric and Oceanic Technology, 30(6), 1072-1090. https://doi.org/10.1175/JTECH-D-12-00136.1

Dominguez, C., Orehounig, K., and Carmeliet, J. (2021a). Estimating hourly lighting load profiles of rural households in East Africa applying a data-driven characterization of occupant behavior and lighting devices ownership. Development Engineering, 6, 100073. https://doi.org/10.1016/j.deveng.2021.100073

Dominguez, C., Orehounig, K., and Carmeliet, J. (2021b). Understanding the path towards a clean energy transition and post-electrification patterns of rural households. Energy for Sustainable Development, 61, 46-64. https://doi.org/10.1016/j.esd.2021.01.002

Dufo-López, R., Cristóbal-Monreal, I. R., and Yusta, J. M. (2016). Optimisation of PV-wind-diesel-battery stand-alone systems to minimise cost and maximise human development index and job creation. Renewable Energy, 94, 280-293. https://doi.org/10.1016/j.renene.2016.03.065

EICV 3 THEMATIC REPORT - Income | National Institute of Statistics Rwanda. (2012, September). https://www.statistics.gov.rw/publication/eicv-3-thematic-report-income

ESMAP. (2015). Beyond Connections Energy Access Redefined. World Bank. https://openknowledge.worldbank.org/handle/10986/24368

ESMAP. (2022). Mini grids for hald a billion people - Market outlook and handbook for decision makers. Tech. Rep. Retrieved from https://pdf.usaid.gov/pdf_docs/PA00X8SK.pdf

Fabini, D. H., Baridó, D. P. de L., Omu, A., and Taneja, J. (2014). Mapping Induced Residential Demand for Electricity in Kenya. Proceedings of the Fifth ACM Symposium on Computing for Development, 43-52. https://doi.org/10.1145/2674377.2674390

Falchetta, G., and Mistry, M. N. (2021). The role of residential air circulation and cooling demand for electrification planning: Implications of climate change in sub-Saharan Africa. Energy Economics, 99, 105307. https://doi.org/10.1016/j.eneco.2021.105307

Falchetta, G., Stevanato, N., Moner-Girona, M., Mazzoni, D., Colombo, E., and Hafner, M. (2021). The M-LED platform: Advancing electricity demand assessment for communities living in energy poverty. Environmental Research Letters, 16(7), 074038. https://doi.org/10.1088/1748-9326/ac0cab

Fobi, S., Deshpande, V., Ondiek, S., Modi, V., and Taneja, J. (2018). A longitudinal study of electricity consumption growth in Kenya. Energy Policy, 123(August), 569-578. https://doi.org/10.1016/j.enpol.2018.08.065

Fobi, S., Kocaman, A. S., Taneja, J., and Modi, V. (2021). A scalable framework to measure the impact of spatial heterogeneity on electrification. Energy for Sustainable Development, 60, 67-81. https://doi.org/10.1016/j.esd.2020.12.005

Ganguly, R., Jain, R., Sharma, K. R., and Shekhar, S. (2020). Mini grids and enterprise development: A study of aspirational change and business outcomes among rural enterprise owners in India. Energy for Sustainable Development, 56, 119-127. https://doi.org/10.1016/j.esd.2020.04.004

Hammer, A., Heinemann, D., Hoyer, C., Kuhlemann, R., Lorenz, E., Müller, R., and Beyer, H. G. (2003). Solar energy assessment using remote sensing technologies. Remote Sensing of Environment, 86(3), 423-432. https://doi.org/10.1016/S0034-4257(03)00083-X

Hartvigsson, E., and Ahlgren, E. O. (2018). Comparison of load profiles in a mini-grid: Assessment of performance metrics using measured and interview-based data. Energy for Sustainable Development, 43, 186-195. https://doi.org/10.1016/j.esd.2018.01.009

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Hartvigsson, E., Ehnberg, J., Ahlgren, E. O., and Molander, S. (2021). Linking household and productive use of electricity with mini-grid dimensioning and operation. Energy for Sustainable Development, 60, 82-89. https://doi.org/10.1016/j.esd.2020.12.004

Hartvigsson, E., Stadler, M., and Cardoso, G. (2018). Rural electrification and capacity expansion with an integrated modeling approach. Renewable Energy, 115(2018), 509-520. https://doi.org/10.1016/j.renene.2017.08.049

Hernández, L., Baladrón, C., Aguiar, J. M., Calavia, L., Carro, B., Sánchez-Esguevillas, A., Pérez, F., Fernández, Á., and Lloret, J. (2014). Artificial Neural Network for Short-Term Load Forecasting in Distribution Systems. Energies, 7(3), Article 3. https://doi.org/10.3390/en7031576

HOMER - Hybrid Renewable and Distributed Generation System Design Software. (n.d.). Retrieved June 7, 2022, from https://www.homerenergy.com/index.html

Howells, M., Rogner, H., Strachan, N., Heaps, C., Huntington, H., Kypreos, S., Hughes, A., Silveira, S., DeCarolis, J., Bazillian, M., and Roehrl, A. (2011). OSeMOSYS: The Open Source Energy Modeling System. An introduction to its ethos, structure and development. Energy Policy. https://doi.org/10.1016/j.enpol.2011.06.033

Huld, T., Moner-Girona, M., and Kriston, A. (2017). Geospatial Analysis of Photovoltaic Mini-Grid System Performance. Energies, 10(218). https://doi.org/10.3390/en10020218

IHOGA. (n.d.). https://ihoga.unizar.es/en/

Jayachandran, M., Gatla, R. K., Rao, K. P., Rao, G. S., Mohammed, S., Milyani, A. H., … Geetha, S. (2022). Challenges in achieving sustainable development goal 7: Affordable and clean energy in light of nascent technologies. Sustainable Energy Technologies and Assessments, 53, 102692. https://doi.org/10.1016/j.seta.2022.102692

Jones, R. V., Fuertes, A., and Lomas, K. J. (2015). The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings. Renewable and Sustainable Energy Reviews, 43, 901-917. https://doi.org/10.1016/j.rser.2014.11.084

Kobayakawa, T., and Kandpal, T. C. (2015). Analysis of electricity consumption under a photovoltaic micro-grid system in India. Solar Energy, 116, 177-183. https://doi.org/10.1016/j.solener.2015.04.001

Kotikot, S. M., Ajinjeru, C., Odukomaiya, A., and Omitaomu, O. A. (2018). Geospatial Framework for Estimating Household Electricity Demand for Urban Infrastructure Planning in Select African Countries. 2018 IEEE PES/IAS PowerAfrica, PowerAfrica 2018, 613-618. https://doi.org/10.1109/PowerAfrica.2018.8521122

Kurata, M., Matsui, N., Ikemoto, Y., and Tsuboi, H. (2018). Do determinants of adopting solar home systems differ between households and micro-enterprises? Evidence from rural Bangladesh. Renewable Energy, 129, 309-316. https://doi.org/10.1016/j.renene.2018.06.015

Kuster, C., Rezgui, Y., and Mourshed, M. (2017). Electrical load forecasting models: A critical systematic review. Sustainable Cities and Society, 35, 257-270. https://doi.org/10.1016/j.scs.2017.08.009

Lee, K., Miguel, E., and Wolfram, C. (2016). Appliance ownership and aspirations among electric grid and home solar households in rural Kenya. American Economic Review, 106(5), 89-94. https://doi.org/10.1257/aer.p20161097

Lombardi, F., Balderrama, S., Quoilin, S., and Colombo, E. (2019). Generating high-resolution multi-energy load profiles for remote areas with an open-source stochastic model. Energy, 177, 433-444. https://doi.org/10.1016/j.energy.2019.04.097

Lorenzoni, L., Cherubini, P., Fioriti, D., Poli, D., Micangeli, A., and Giglioli, R. (2020). Classification and modeling of load profiles of isolated mini-grids in developing countries: A data-driven approach. Energy for Sustainable Development, 59, 208-225. https://doi.org/10.1016/j.esd.2020.10.001

Louie, H., and Dauenhauer, P. (2016). Effects of load estimation error on small-scale off-grid photovoltaic system design, cost and reliability. Energy for Sustainable Development, 34, 30-43. https://doi.org/10.1016/j.esd.2016.08.002

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Louw, K., Conradie, B., Howells, M., and Dekenah, M. (2008). Determinants of electricity demand for newly electrified low-income African households. Energy Policy, 36(8), 2812-2818. https://doi.org/10.1016/j.enpol.2008.02.032

Mandelli, S., Merlo, M., and Colombo, E. (2016). Novel procedure to formulate load profiles for off-grid rural areas. Energy for Sustainable Development, 31, 130-142. https://doi.org/10.1016/J.ESD.2016.01.005

Mandelli, Stefano, Jacopo Barbieri, Riccardo Mereu, and Emanuela Colombo. Off-grid systems for rural electrification in developing countries: Definitions, classification and a comprehensive literature review. Renewable and Sustainable Energy Reviews, 58:1621–1646, 5 2016. ISSN 13640321. doi: 10.1016/j.rser.2015.12.338.

Markovic, D., Cvetkovic, D., and Masic, B. (2011). Survey of software tools for energy efficiency in a community. Renewable and Sustainable Energy Reviews, 15(9), 4897-4903. https://doi.org/10.1016/j.rser.2011.06.014

Marques, F., Páscoa, S., Carvalho, M. J., and Cardoso, J. P. (2020). Solar Irradiance (Global, Direct and Diffuse) Quality Control Methodologies Review: Application to Time Series Measured At LES/LNEG, Lisboa, Portugal. In Proceedings of the ISES EuroSun 2020 Conference - 13th International Conference on Solar Energy for Buildings and Industry (p. 11). Freiburg, Germany: International Solar Energy Society. https://doi.org/10.18086/eurosun.2020.09.04

Mbinkar, E. N., A. Asoh, D., Tchuidjan, R., and Baldeh, A. (2021). Design of a Photovoltaic Mini-Grid System for Rural Electrification in Sub-Saharan Africa. Energy and Power Engineering, 13(03), 91-110. https://doi.org/10.4236/epe.2021.133007

Mueller, R., Behrendt, T., Hammer, A., and Kemper, A. (2012). A new algorithm for the satellite-based retrieval of solar surface irradiance in spectral bands. Remote Sensing, 4(3), 622-647. https://doi.org/10.3390/rs4030622

National Centers for Environmental Prediction NOAA, U. S. N. W. S. (2000). NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999. https://doi.org/10.5065/D6M043C6

National Centers for Environmental Prediction NOAA, U.S. Department of Commerce, N. W. S. (2015). NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive. https://doi.org/10.5065/D65D8PWK

NCEP/DOE. (2000). NCEP/DOE Reanalysis 2 (R2).

Pfenninger, S., and Staffell, I. (2016). Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. Energy, 114, 1251-1265. https://doi.org/10.1016/j.energy.2016.08.060

Pfenninger, S., Hawkes, A., and Keirstead, J. (2014). Energy systems modeling for twenty-first century energy challenges. Renewable and Sustainable Energy Reviews, 33, 74-86. https://doi.org/10.1016/j.rser.2014.02.003

Pfenninger, S., Hirth, L., Schlecht, I., Schmid, E., Wiese, F., Brown, T., Davis, C., Gidden, M., Heinrichs, H., Heuberger, C., Hilpert, S., Krien, U., Matke, C., Nebel, A., Morrison, R., Shivakumar, A., Staffell, I., Trondle, T., and Wingenbach, C. (2018). Opening the black box of energy modelling: Strategies and lessons learned. Energy Strategy Reviews, 19, 63-71. https://doi.org/10.1016/j.esr.2017.12.002

Rao, N. D., and Ummel, K. (2017). White goods for white people? Drivers of electric appliance growth in emerging economies. Energy Research and Social Science, 27, 106-116. https://doi.org/10.1016/j.erss.2017.03.005

Richmond, J., and Urpelainen, J. (2019). Electrification and appliance ownership over time: Evidence from rural India. Energy Policy, 133(October 2018), 110862. https://doi.org/10.1016/j.enpol.2019.06.070

Richmond, Jennifer, Shalu Agrawal, and Johannes Urpelainen. Drivers of household appliance usage: Evidence from rural india. Energy for Sustainable Development, 57:69–80, 8 2020. ISSN 09730826. doi: 10.1016/j.esd.2020.05.004.

Riva, F. (2019). Modelling Endogenous Complexities In Rural Electrification: On The Local Dynamics Of Growth And The Planning Of Off-Grid Systems. Politecnico di Milano.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Riva, F. (2020). When complexity turns into local prosperity: A system dynamics approach to meeting the challenges of the rural electricity-development nexus. Energy for Sustainable Development, 59, 226-242. https://doi.org/10.1016/j.esd.2020.10.009

Riva, F., and Colombo, E. (2020). System-dynamics modelling of the electricity-development nexus in rural electrification based on a Tanzanian case study. Energy for Sustainable Development, 56, 128-143. https://doi.org/10.1016/j.esd.2020.04.001

Riva, F., Ahlborg, H., Hartvigsson, E., Pachauri, S., and Colombo, E. (2018). Electricity access and rural development: Review of complex socio-economic dynamics and casual diagrams for more appropriate energy modelling. Energy for Sustainable Development, 43, 203-223. https://doi.org/10.1016/j.esd.2018.02.003

Riva, F., Gardumi, F., Tognollo, A., and Colombo, E. (2019). Soft-linking energy demand and optimisation models for local long-term electricity planning: An application to rural India. Energy, 166, 32-46. https://doi.org/10.1016/j.energy.2018.10.067

Riva, F., Tognollo, A., Gardumi, F., and Colombo, E. (2018). Long-term energy planning and demand forecast in remote areas of developing countries: Classification of case studies and insights from a modelling perspective. Energy Strategy Reviews, 20, 71-89. https://doi.org/10.1016/j.esr.2018.02.006

Ruijven, B. J. van, Vuuren, D. P. van, Vries, B. J. M. de, Isaac, M., Sluijs, J. P. van der, Lucas, P. L., and Balachandra, P. (2011). Model projections for household energy use in India. Energy Policy, 39(12), 7747-7761. https://doi.org/10.1016/j.enpol.2011.09.021

Saad, A., M. Amen, M., Refaat, M., and Morad, A. (2012). Utilization High Spatial Satellite Images Reflectance's to Estimate the Nautical Charts for the Nile River. Australian Journal of Basic and Applied Sciences, (1991-8178), 8.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., … Goldberg, M. (2010). NCEP Climate Forecast System Reanalysis (CFSR) Selected Hourly Time-Series Products, January 1979 to December 2010. Boulder, CO: Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D6513W89

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., … Becker, E. (2011). NCEP Climate Forecast System Version 2 (CFSv2) 6-hourly Products. Boulder, CO: Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D61C1TXF

Schunke, J., Laux, P., Bliefernicht, J., Waongo, M., Sawadogo, W., and Kunstmann, H. (2021). Exploring the Potential of the Cost-Efficient TAHMO Observation Data for Hydro-Meteorological Applications in Sub-Saharan Africa. Water, 13(22), 3308. https://doi.org/10.3390/w13223308

SEforALL, and BloombergNEF. (2019). State of the Global Mini-grids Market Report 2020. Report, 53(9), 1689-1699.

Sengupta, M., Habte, A., Wilbert, S., Gueymard, C., and Remund, J. (2021). Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications: Third Edition. NREL/TP-5D00-77635. Retrieved from https://www.osti.gov/servlets/purl/1778700/

Shibano, K., and Mogi, G. (2020). Electricity Consumption Forecast Model Using Household Income: Case Study in Tanzania. Energies, 13(10), Article 10. https://doi.org/10.3390/en13102497

Sinha, S., and Chandel, S. S. (2014). Review of software tools for hybrid renewable energy systems. Renewable and Sustainable Energy Reviews, 32, 192-205. https://doi.org/10.1016/j.rser.2014.01.035

Sovacool, B. K. (2014). What are we doing here? Analyzing fifteen years of energy scholarship and proposing a social science research agenda. Energy Research and Social Science, 1, 1-29. https://doi.org/10.1016/j.erss.2014.02.003

Staffell, I., andand Pfenninger, S. (2016). Using bias-corrected reanalysis to simulate current and future wind power output. Energy, 114, 1224-1239. https://doi.org/10.1016/j.energy.2016.08.068

Stevanato, N., Francesco Lombardi, Emanuela Colmbo, Sergio Balderrama, and Sylvain Quoilin. Two-stage stochastic sizing of a rural micro-grid based on stochastic load generation. pages 1–6, Milan, Italy, 6 2019. 2019 IEEE Milan PowerTech, IEEE. ISBN 978-1-5386-4722-6. doi: 10.1109/PTC.2019.8810571

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Stevanato, N., Rinaldi, L., Pistolese, S., Subieta, S. L. B., Quoilin, S., and Colombo, E. (2020). Modeling of a village-scale multi-energy system for the integrated supply of electric and thermal energy. Applied Sciences (Switzerland), 10(21), 1-20. https://doi.org/10.3390/app10217445

Šúri, M., Huld, T. A., and Dunlop, E. D. (2005). PV-GIS: A web-based solar radiation database for the calculation of PV potential in Europe. International Journal of Sustainable Energy, 24(2), 55-67. https://doi.org/10.1080/14786450512331329556

Ulsrud, K., Winther, T., Palit, D., Rohracher, H., and Sandgren, J. (2011). The Solar Transitions research on solar mini-grids in India: Learning from local cases of innovative socio-technical systems. Energy for Sustainable Development, 15(3), 293-303. https://doi.org/10.1016/j.esd.2011.06.004

USAID, African Union, and Power Africa. (2021). Unlocking Africa'S Mini-Grid Market Final Report Usaid Scaling Up Renewable Energy Program (Sure). Tech. Rep. Retrieved from https://pdf.usaid.gov/pdf_docs/PA00X8SK.pdf

van Ruijven, Bas J., Detlef P. van Vuuren, Bert J.M. de Vries, Morna Isaac, Jeroen P. van der Sluijs, Paul L. Lucas, and P. Balachandra. Model projections for household energy use in india. Energy Policy, 39(12):7747–7761, 2011. doi: 10.1016/j.enpol.2011.09.0.

VIDA. (n.d.). Retrieved August 16, 2022, from https://www.villagedata.io/

Wild, M., Ohmura, A., Schär, C., Müller, G., Folini, D., Schwarz, M., … Sanchez-Lorenzo, A. (2017). The Global Energy Balance Archive (GEBA) version 2017: a database for worldwide measured surface energy fluxes. Earth System Science Data, 9(2), 601-613. https://doi.org/10.5194/essd-9-601-2017

Williams, N. J., Jaramillo, P., Campbell, K., Musanga, B., and Lyons-Galante, I. (2018). Electricity Consumption and Load Profile Segmentation Analysis for Rural Micro Grid Customers in Tanzania. 2018 IEEE PES/IAS PowerAfrica, PowerAfrica 2018, 360-365. https://doi.org/10.1109/PowerAfrica.2018.8521099

Williams, N. J., Jaramillo, P., Cornell, B., Lyons-Galante, I., and Wynn, E. (2017). Load characteristics of East African micro-grids. 2017 IEEE PES PowerAfrica, 236-241. https://doi.org/10.1109/PowerAfrica.2017.7991230

Yue, X., Pye, S., DeCarolis, J., Li, F. G. N., Rogan, F., and Gallachóir, B. Ó. (2018). A review of approaches to uncertainty assessment in energy system optimization models. Energy Strategy Reviews, 21, 204-217. https://doi.org/10.1016/j.esr.2018.06.003

Zebra, E. I. C., van der Windt, H. J., Nhumaio, G., and Faaij, A. P. C. (2021). A review of hybrid renewable energy systems in mini-grids for off-grid electrification in developing countries. Renewable and Sustainable Energy Reviews, 144, 111036. https://doi.org/10.1016/j.rser.2021.111036

Zeyringer, M., Pachauri, S., Schmid, E., Schmidt, J., Worrell, E., and Morawetz, U. B. (2015). Analyzing grid extension and stand-alone photovoltaic systems for the cost-effective electrification of Kenya. Energy for Sustainable Development, 25(2015), 75-86. https://doi.org/10.1016/j.esd.2015.01.003

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.