Energy for Sustainable Development 82 (2024) 101520

Contents lists available at ScienceDirect

Energy for Sustainable Development

journal homepage: www.journals.elsevier.com/energy-for-sustainable-development

GIS-facilitated procedure for optimal rural electrification planning: A case study in Naeder, Ethiopia

Aleksandar Dimovski *, Zahra Pezham, Mohammad Ahmadi, Lorenzo Maria Filippo Albertini, Darlain Irenee Edeme, Marco Merlo

Department of Energy, Politecnico di Milano, Via Lambruschini, 4, Milano, Italy

ARTICLE INFO

Keywords:
Rural electrification
Geospatial planning
Multi-Tier Framework
Machine learning
MILP
Energy demand

ABSTRACT

Although the global electrification rate has reached 91 %, roughly 730 million people still live without reliable and affordable access to electricity, experiencing the first regression since 2013, following the COVID-19 crisis. In this context, this paper aims to define a comprehensive electrification strategy through an innovative model based on open-source data and machine learning algorithms, able to reduce the time and resource-consuming on-field campaign that is generally needed for gathering data, and subsequently define the electrification strategy. Following the location of human settlements and their socio-economic characterizations carried out by a novel open-source tool proposed within this paper named VANIA (Village ANalytics in Africa), the energy demand and hourly demand profile of each community are estimated through the application of machine learning techniques based on MTF (Multi-Tier Framework) surveys and a stochastic bottom-up model for load profile generation. The approach is designed to manage the complex nonlinear relationship between the energy needs of a community and its socio-demographic parameters. Then, taking the communities' demand profile as input, a GIS-facilitated procedure is utilized to optimize the electrification strategy for the territory under investigation, proposing the least-cost electrification solution. The final electrification plan focuses on long-term solutions enabling growth over time in which each community can be either connected to the national grid or supplied by an off-grid system. Ultimately, to demonstrate the approach and showcase its operational capabilities, the methodology is utilized for the electrification planning of the Naeder province in Tigray, Ethiopia, characterized by a predominantly lacking electrification status and low energy demand. The suggested solution advocates for the cost-efficient electrification of approximately 11,560 households clustered in 50 communities. Considering consolidated economic parameters and a perceived cost of electricity of 110 €/MWh showed a preference toward grid extension, with 39 out of 50 communities connected to the national grid. Finally, sensitivity analysis on the cost of energy showed that regardless of the value, 3 communities should be electrified with a microgrid, whereas for values upward of 130 €/MWh the microgrid starts becoming the more lucrative option, and at 145 €/MWh an extension is not economically justified.

Introduction

Various literature sources have already found a clear correlation between access to electricity and improvement in socio-economic indicators (Mandelli et al., 2016). As such, the seventh goal of the 2030 Agenda for Sustainable Development aims to "Ensure access to affordable, reliable, sustainable and modern energy for all" through an equitable distribution of resources. However, even though the global

Abbreviations: BESS, Battery Energy Storage System; CoE, Cost of Electricity; DG, Diesel Generator; DSO, Distribution System Operator; ESMAP, Energy Sector Management Assistance Program; GHI, Global Horizontal Irradiation; GIS, Geographical Information Systems; GISELE, GIS for rural ELEctrification; HDI, Human Development Index; HOMER, Hybrid Optimization of Multiple Energy Resources; IEA, International Energy Agency; LCOE, Levelized Cost of Electricity; LV, Low Voltage; MAE, Mean Average Error; MILP, Mixed-Integer Linear Programming; ML, Machine Learning; MTF, Multi-Tier Framework; MV, Medium Voltage; NPC, Net Present Cost; NREL, National Renewable Energy Laboratory; O&M, Operation and Maintenance; OnSSET, Open-Source Spatial Electrification Tool; OSM, Open Street Map; PV, Photovoltaic; REM, Reference Electrification Model; RES, Renewable Energy Sources; RF, Random Forest; RMSE, Root Mean Square Deviation; RNM, Reference Network Model; VANIA, Village ANalytics In Africa.

* Corresponding author.
E-mail address: aleksandar.dimovski@polimi.it (A. Dimovski).

https://doi.org/10.1016/j.esd.2024.101520
Received 2 May 2024; Received in revised form 18 July 2024; Accepted 19 July 2024
Available online 5 August 2024

Table 1
Summary of GIS electrification planner tools.

Tool	Geospatial planning	Grid extension	Clustering approach	Energy system model	Open Source Code	Electric system optimization	Penalty cost factor	Weakness
Data for Good	x	–	–	–	x	–	–	Only six countries - MV lines; Basic implementation of a routing algorithm without a comprehensive planning framework
OnSSET	x	x	x	x	x	–	x	1 km × 1 km resolution; Improper assessment of grid extension and no modeling of electric network
Network planner	x	x	–	–	–	–	–	Mini grid modeling based only on diesel generators; Improper modeling of the electric network
GEOSIM	x	x	–	x	–	–	–	For stand-alone systems only PV technology is considered; Lack of transparency on the algorithms utilized
HOMER	–	–	–	x	–	–	–	The distribution grid is not optimized or modeled; Lack of large-scale analysis
REM/RNM	x	x	–	x	–	–	x	Energy sources are limited; Focus on urban areas; Utilization of only heuristic algorithms
GISELE	x	x	x	x	x	x	x	Needs in input an estimation of the energy demand; Clustering algorithm without a direct way of determining the parameters; Limited scalability

electrification rate reached 91 % of the worldwide population in 2020, around 730 million people still live without access to electricity (IEA, 2022a), with over 77 % of them being in Sub-Saharan Africa. Henceforth, the African Union adopted the “Agenda 2063: The Africa We Want”, the continent’s strategic framework that aims to deliver inclusive and sustainable development (Tella, 2018; Garfias Royo et al., 2022). However, the COVID-19 pandemic has interrupted the favorable trajectory of energy access, leading to the first recorded setback since 2013 (WHO Media inquiries, 2022). This necessitates the redirection of investment flows to reinstate momentum, particularly for off-grid systems, where the cumulative investment falls significantly short of the estimated requirement for achieving universal energy access (IEA, 2022b). However, errors arising from survey-based bottom-up approaches (Riva et al., 2019) and lack of reliable information could be barriers for governments, international organizations, and local enterprises (Avila et al., 2017) inclined to deploy new infrastructures. On the other hand, owing to the elevated investment expenses related to electrical infrastructure, planners and investors necessitate tools that can enable them to formulate strategies for reducing electrification costs (Loken, 2007; Cinelli et al., 2022; Environmental Research: Infrastructure and Sustainability, 2021). In general, field visits remain essential, rendering the process time- and resource-intensive, particularly when addressing expansive and distant regions, as exemplified by the case of FUNAE, a public institution that works on rural electrification in Mozambique (Uamusse et al., 2017). Under such circumstances, the utilization of open-source GIS data and procedures could expedite and enhance the decision-making process (Dimovski et al., 2023).

GIS for rural electrification overview

Energy access planning, especially in developing countries, faces various difficulties when trying to estimate load profiles and energy trends across the years (The World Bank, 2013). As the population of sub-Saharan Africa continues to expand rapidly (The World Bank, 2022), the task of augmenting electricity accessibility has proven to be complex and financially demanding. It is within this context that the integration of open-source and GIS data could signify a significant stride in the analysis and characterization of remote regions (Isihak et al., 2022; Blechinger et al., 2019), which applies to both load estimation and the configuration of electrification solutions. In order to obtain a reliable load estimation in rural areas, different algorithms are proposed in the literature, based on different approaches, such as the stochastic bottom-up procedure proposed in RAMP (Lombardi et al., 2019), the end-use technique (Mwakitalima & King’ondu, 2015), and socio-demographic comparison (Caquilpan et al., 2017), or by using archetypes of daily energy demand profiles (Setiawan et al., 2009). Recently, computer vision algorithms (Correa et al., 2021) and ML algorithms (Sarhani & El Afia, 2015) have been gaining momentum. Indeed, these approaches that are potentially able to find nonlinear correlations between real-life data coming from surveys on one side, and socio-economic parameters on the other side represent a viable way for assessing the energy requirements. Pursuant to this objective, ESMAP initiated a data collection initiative known as the MTF framework in 2015, aimed at gathering extensive country-scale data and providing analysis to portray a country’s energy situation, including the categorization of different levels of energy access into six distinct Tiers. This dataset holds the potential to serve as input indicators for a ML-driven comprehensive energy assessment analysis on a wide scale.

Table 1 summarize the features of the main procedures presented in the literature: (Gershenson et al., 2019) utilizing Data for Good (Meta, 2022) applies a routing algorithm to a night-time lighting dataset from NASA to estimate the transmission and distribution network across the entire World, particularly in the African continent; OnSSET (Mentis et al., 2017) calculates the LCOE and the resulting least-cost electrification strategy (Khavari et al., 2021); Network planner (Natali & Carbajal, 2017; Cader et al., 2016) attempts to design electrification solution from the level of a single community to the level of an entire nation (“Network Planner”, n.d.; GEOSIM IED, n.d.) is a GIS-based software developed for swiftly constructing highly interactive rural electrification planning scenarios which enables georeferenced electrification planning within a particular framework when load forecasts are feasible and diesel, biomass, and hydro mini-grids can be modeled (Watchueng et al., 2010). HOMER (HOMER Energy LLC, n.d.; Farret & Godoy, 2005) is a software developed by NREL which conducts NPC analysis and identifies the least-cost hybrid generation portfolio for a pre-defined area of interest. The RNM (Khaitan & Gupta, 2013; Mateo Domingo et al., 2011) uses a heuristic branch-exchange algorithm to efficiently create various grid configurations aiming to reduce the overall costs, considering investment, O&M and reliability. It was developed to conduct analysis in Spain and in the US that would guide the regulatory bodies in properly setting the revenue cap of local DSOs. On the basis of RNM as a network designer, another tool named REM was developed (Ciller et al., 2019; Amatya et al., 2019). It is a comprehensive least-cost electrification design tool that was specifically developed for brown- and green-field electrification planning analysis. It identifies the most cost-effective electrification solution on a household-

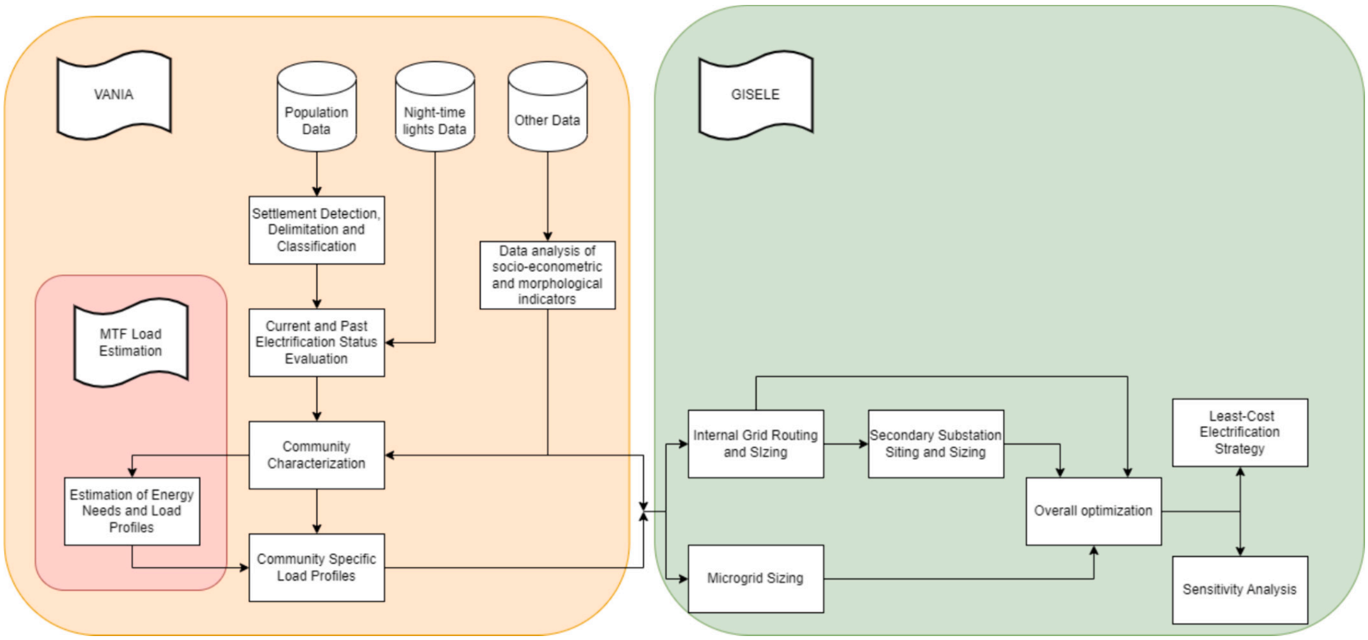


Fig. 1. Overall flowchart of the proposed methodology.

level choosing between stand-alone systems, isolated microgrids with local generation, and distribution network extension. Finally, a research group at Politecnico di Milano has introduced a new tool named GISELE that serves to delineate the optimal electrification strategy in a specified area. This involves a comparative analysis between stand-alone microgrid systems and public grid extension, effectively leveraging GIS data, graph theory, and MILP modeling techniques (Dimovski et al., 2023; Corigliano et al., 2020). However, even though the tool has had a significant evolution over the recent years, it should be noted that it lacks features for a wide-scale energy assessment, which requires an external procedure to quantify the energy needs of the identified communities.

The literature review clearly points out the complexity of the problem and the deficiency in solutions capable of adequately addressing the overall issue. In particular, the authors highlight the lack of input data or, equivalently, the substantial time and economic effort required to collect data.

The objectives of this paper revolve around introducing an efficient procedure that tackles time and resource constraints to promote a comprehensive electrification strategy. First, open source and GIS data are acquired and processed to locate communities and characterize them with socio-economic indicators. Second, using the georeferenced surveys from the MTF, a bottom-up energy assessment analysis is conducted for a subset of the communities in the region of interest for which surveys were available. Then, a ML algorithm is proposed to capture the non-linear relationships between the communities' indicators and their energy needs, and subsequently extrapolate the energy demand results to the entire area under analysis. Finally, the energy assessment is utilized to support cost-effective electrification planning prioritizing the detailed topological and geospatial design of the electric grid. This includes both communities that are supposed to be connected to the national grid, as well as those operating autonomously as isolated off-grid systems. Additionally, the optimal hybrid generation portfolio is obtained for the latter.

Proposed methodology

This chapter will go through the three steps of the proposed framework depicted in Fig. 1. The first two are novelties proposed in this paper, that are integrated with electrification planning tool GISELE, aiming to illustrate the effectiveness of the proposed framework in electrification planning.

In particular, the first step is Village Analytics in Africa (VANIA), a spatial analyst software tool devoted to gathering and processing open-source data for an area under investigation. It acquires and processes open-source databases to obtain country-scale socio-demographic information, creates clusters of population (i.e. communities) and subsequently characterizes them, effectively enhancing the electrification efforts with a diverse set of detailed data.

The second part is an energy estimation procedure based on the MTF framework, which aims to estimate the energy demand of communities of interest. Using the socio-demographic characteristics obtained from VANIA and surveys collected within the MTF, it proposes an ML approach to find complex nonlinear correlations between the energy consumption of a community and its socio-economic characteristics.

Finally, as previously discussed, given the communities and energy demand from the previous steps, GISELE is utilized to propose the optimal electrification strategy (Dimovski et al., 2023). The next subsections aim to provide sufficient input on the novelties that this paper introduces, as well as a brief overview of the most important aspects of the tool. Finally, it is fully available open-source on the platform GitHub under an APACHE 2.0 license, accessible on (Gisele, 2022).

VANIA

VANIA is a Python-based procedure that performs three complementary actions:

1) Locates communities as clusters within a territory of interest using a novel density-based clustering approach.
2) Processes currently available open-source datasets to perform socio-economic parametrization of the communities.
3) Labels the communities based on their presumed electrification status, effectively locating the ones that would require access to electricity.

The clustering procedure utilizes an iterative implementation of DBSCAN, a renowned density-based algorithm, the main advantage of which is its suitability for analyzing large datasets with varying sample density (Ester et al., 1996).

The first step is the definition of the community borders, and for this

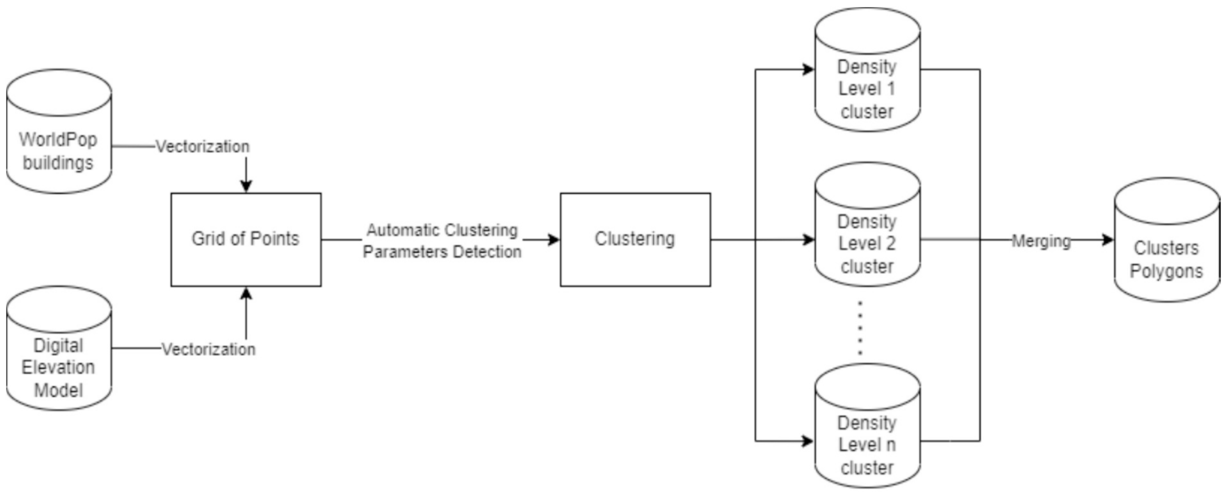


Fig. 2. Overview of the clustering procedure.

goal the proposed procedure takes as input a dataset that implies the spatial distribution of the population. Different datasets are investigated, such as WorldPop's 100-m resolution buildings count dataset and its, 100-m resolution estimated population dataset (Tatem, 2017) and Google's Open Buildings dataset (Google, 2022). The objective is to consider the variability of population density even within the same administrative unit weighing a certain attribute of points referring to the population distribution in space. For the two datasets referring to buildings, the number of buildings or people inside the pixel is used for the WorldPop raster, while the roof area is used for the Google vector. The reason why the authors also test the result with the WorldPop population dataset is evident when thinking about urban areas, where the knowledge of both buildings and population density is relevant. Thus, the proposed algorithm segments the population into clusters with varying densities on a case-by-case basis, relying on the count of building footprints and the population. The detection of possible outliers is composed as described below:

1. Calculation of the Population Mean.
2. Calculation of the Population Standard Deviation.
3. Calculation of Z score as:

$$Z = \frac{(X - \mu)}{\sigma} \tag{1}$$

where:
X = observation, which in this case is the actual population;
μ = Population Mean;

Table 3
Processed attributes for each community through VANIA.

Number of administrative areas for each level present in the cluster	Km of electric grids (categorized by type: existing, planned, etc.)
Categorized OSM locations inside the cluster	% of buildings close (500 m) to the electric grids (categorized by type)
Area and extension of the cluster	Km of roads within the cluster
Estimated (distributed) population through Facebook's dataset	% of buildings close (500 m) to the roads
Estimated Buildings distribution through Worldpop's dataset	Number of health centers and Number of health centers per 1000 inhabitants
Population and building densities	Close to (500 m) grid health centers
% of urban and rural areas inside the cluster	Number of Schools and Number of Schools per 1000 inhabitants
% of the area with night lights	Close to (500 m) grid Schools
% of buildings in areas with night lights	% of area inside the cluster for each landcover category
Decision whether or not to include cluster in the GISELE modeling	Rivers in a (15 km) buffer zone with respect to the cluster
Estimation of the load barycenter of the cluster	Closest rivers' point and highest flow rate river point
Max, min, average elevation	Average GHI value in the cluster
Highest and average wind speed value in the cluster	Max areas [km^2] required for PV and Wind Power based on peak load
Potential location for PV plant (assuming only off-grid system)	Categorization of the electrification status
Potential location for wind plant (assuming only off-grid system)	

Table 2
Name, sources and format of considered dataset.

Name	Format	Source	Name	Format	Source
Administrative	MultiPolygon Vector	UN Agencies	Networks	Vectors	World Bank
Cell towers	Points	OpenCell	Night Lights	Raster	NASA
Clustering (Sub)	MultiPolygon Vector	KTH	Population	Raster	Facebook
Crops	Raster	Harvard	Population Growth	Raster	Columbia
Development Potential	Raster	Columbia	Poverty	Raster	WorldPop
Distance to city	Raster	–	Protected areas	Vector	Protected Planet
Elevation	Raster	RCMRD	Relative Wealth	Points	Facebook
Food Insecurity	Raster	Columbia	Rivers	Vector	Hydroshed
GHI	Raster	Global Solar Atlas	Roads	Vector	OSM
HDI	Shapefile	UN	Schools	Points	OSM
Hospitals	Points	OCHA	Singularities	Vectors	Multiple
Landcover	Raster	ESA	Substations	Vectors	Multiple
Literacy	Raster	WorldPop	Urban	Raster	WorldPop
Locations	Vectors	OpenStreetMap	Wind	Raster	Global Wind Atlas
MTF	Stata Dataset	World Bank			

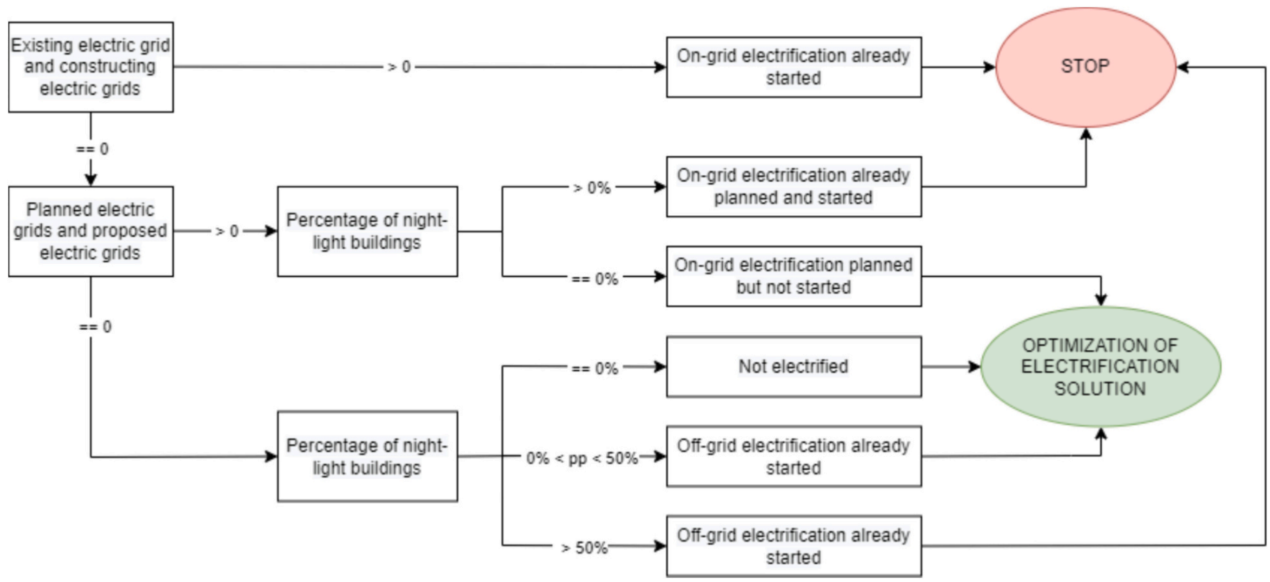


Fig. 3. Definition of electrification rate and type procedure.

σ = Population Standard Deviation.

4. Defining the threshold as the standard deviations from the mean, beyond which a point is classified as an outlier.
5. Extraction of outliers as those points for which the Z score is greater than the previously defined threshold.

The entire clustering procedure can be described through different iterative steps displayed on Fig. 2. The initial phase involves the vectorization of buildings and population through i-level density clustering. Furthermore, after the initial inputs regarding buildings, population, and elevation have been converted into vector format, the tool generates a geo-referenced regular grid of points. Once the boundaries of each community are defined, the procedure attributes multiple indicators for characterizing each community through economic, social, environmental, and infrastructural information. Based on the availability of various datasets, the outcome is a comprehensive database that pertains to distinct, defined communities; data for each community is collected from the sources documented in Table 2.

During the second step, the information gathered for each cluster are processed and, through the combination of the geospatial data and the datasets, the procedure defines different attributes, as detailed in Table 3.

One of the key features of VANIA is the assessment of the electrification status of each community, which is based on a two-step procedure:

1. Should data pertaining to the existing distribution network be insufficient, the procedure exclusively depends on nocturnal illumination data, sourced from a globally validated database (Li et al., 2020), characterized by a spatial resolution of 1 km and a yearly temporal resolution spanning from 1992 to 2018. The extended temporal resolution permits the authors to identify the highest value for each pixel, mitigating errors arising from weather conditions.

2. In case distribution network data are available, the procedure allows for a more accurate estimation of the electrification status quo. Indeed, it is possible to merge data from the night-time lighting dataset (NASA) and the existing distribution network to identify communities (i.e. clusters) fed by the national grid and those fed by a local microgrid. The adopted algorithm is detailed in Fig. 3. Notably, the presented approach evaluates the existing and planned grids, processes data for

Appliance	Tier 1			Tier 2			Tier 3			Tier 4			Tier 5		
	Watts	hours /day	Min. annual consumption (kWh)	Watts	hours /day	Min. annual consumption (kWh)	Watts	hours /day	Min. annual consumption (kWh)	Watts	hours /day	Min. annual consumption (kWh)	Watts	hours /day	Min. annual consumption (kWh)
Task Lighting	1	4	1.5	2	4	2.9	2	4	2.9	2	8	5.8	2	8	20
Phone Charging	2	2	1.5	2	4	2.9	2	4	2.9	2	4	2.9	2	4	2.9
Radio	2	2	1.5	4	4	5.8	4	4	5.8	4	4	5.8	4	4	5.8
General Lighting				12	4	17.5	12	4	17.5	12	8	35	12	12	52.5
Air Circulation				20	4	29.2	40	6	87.6	40	12	175.2	40	18	262.8
Television				20	2	14.6	40	2	29.2	40	2	29.2	40	2	29.2
Food Processing							200	0.5	36.5	200	0.5	36.5	200	0.5	36.5
Washing Machine							500	1	182.5	500	1	182.5	500	1	182.5
Refrigerator										300	6	657	300	6	657
Iron										1100	0.3	120.5	1100	0.3	120.5
Air Conditioner													1500	3	1642.5

Fig. 4. Power and Functioning window of appliances (Bhatia & Angelou, 2015).

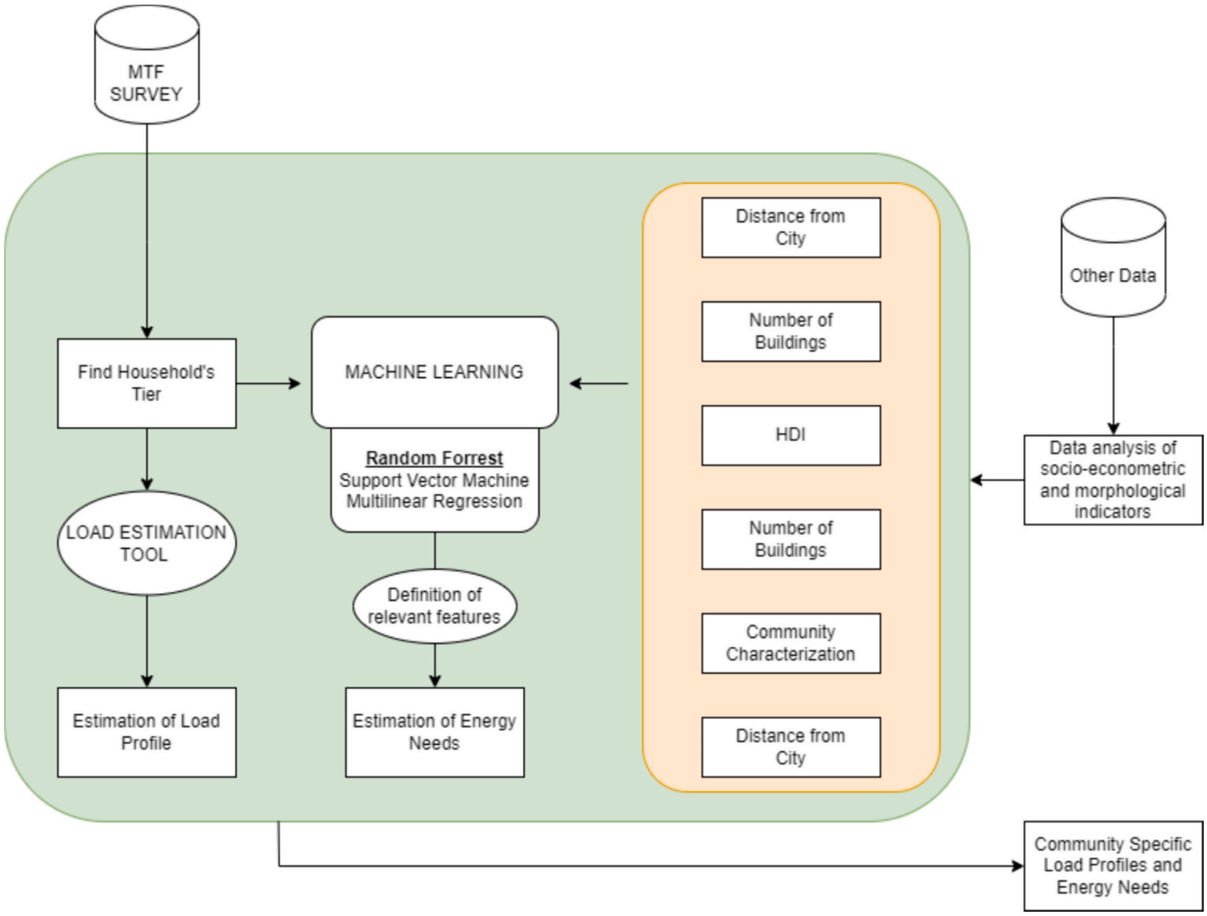


Fig. 5. General approach for the estimation of load profiles and energy needs.

each cluster, and ultimately schedules the optimization of the electrification strategy, as described in the following chapters.

MTF and energy needs estimation

Once the communities and the electrification status quo are defined, a reliable estimation of the energy needs is vital to the optimization of the energy and electrical infrastructure. In this paper, the proposed procedure is based on data acquired by the MTF, an initiative initiated in June 2015 by ESMAP. The MTF collects a comprehensive set of data at the country level, the approach recognizes that electricity access encompasses a spectrum of service levels encountered by households, enterprises, and institutions. In particular, according to the usage of energy that a household could afford related to its income, different classes are defined, from Tier 0 (no access) to 5 (the highest level of access) (SEforALL, 2016; Mullen & Wade, 2020). As shown in Fig. 4, each tier refers to the minimum required appliances, power capacity, and functioning window to calculate the energy requirements (Foster & Tre, 2000; Mullen & Wade, 2020).

Nonetheless, at the time of writing in 2024, the geographical scope of MTF data compiled by ESMAP is restricted, encompassing only 8 African countries. Within this framework, this study endeavors to employ machine learning techniques to establish correlations between these existing data and significant socio-economic parameters, thereby enabling the indirect utilization of the MTF across the globe.

Several ML techniques have been investigated: Random Forest (Liaw & Wiener, 2002; Zhang et al., 2022), Support Vector Machine (Chen et al., 2022; Winters-Hilt & Merat, 2007) and Multiple Linear Regression, to determine the most suitable option for the problem under investigation considering both accuracy and robustness in the evaluation process. In succession, ML models were trained and applied to the available dataset, followed by a stepwise simulation where features were systematically added to ascertain the optimal input, using the RMSE as the metric for comparison.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{2}$$

N = Number of samples, y_i = Actual observation as the total consumption of each community, y_i = Predicted value of total consumption.

As a result of the large amount of possible input data and characteristics, it is often required to guide the procedure to reduce the number of irrelevant attributes, with the goal of avoiding inconsistent behavior in the demand estimation, preserving accuracy and robustness (Bosisio et al., 2021). Quantitative analysis showed that the distance from the nearest city, wealth index, HDI, number of buildings, and population were posited as the most pivotal parameters for delineating the energy demand of a community. Consequently, these features were integrated as features of the dataset, over which different algorithms were utilized to select the one that minimizes the RMSE. The entire procedure is shown in Fig. 5. In the investigated case study, as detailed in 4.1, RF proved to be the most effective approach in maximizing the estimation's accuracy. The exploration and comparison of the ML algorithms falls beyond the scope of this paper.

Optimal electrification planning: GIS for rural ELEctrification (GISELE)

The final stage of the proposed approach is the optimization of the electrical infrastructure's design within each unelectrified community, and then for the entire area of interest in a holistic manner. The optimization process for devising the most cost-effective electrification plan is executed using the tool GISELE, an open-source framework in Python developed by the Energy4Growing team, which facilitates extensive large-scale electrification analysis. GISELE is a comprehensive and consolidated electrification software, with the latest edition containing significant improvements over previous releases (Corigliano et al., 2020). An exhaustive in-depth explanation of the tool and its stages is available in (Dimovski et al., 2023), whereas the software itself is accessible under a free APACHE 2.0 license. It is a cluster-oriented procedure that integrates both top-down and bottom-up methodologies to generate a realistic outcome for a least-cost electrification strategy. Unlike various tools and procedures proposed in the literature, it puts emphasis on the structure and deployment of the distribution grid by establishing a grid layout of the area that effectively accounts for diverse costs linked to morphological constraints. This is done by determining a penalty factor (*Pf*) for each point of the grid, in alignment with factors contributing to the complexity of implementing grid distribution lines, including road distance, land cover, and elevation (Monteiro et al., 2015), and their impact on the different costs related to line deployment, such as poles, conductors, permits and maintenance, with the various coefficients available in (Dimovski et al., 2023).

$$Pf = 1 + \sum_{i}^{categories} penalty_i \tag{3}$$

The distinctive merit of this approach lies in its capability to conduct an extensive assessment of the electrical system, pinpoint the optimal electrification solution, and account for potential challenges in the routing of the electric lines.

With this being said, it has been well established in the literature and practice that off-grid systems can be the preferred solution to electrification in rural areas (Palit & Chaurey, 2011; Moshi et al., 2016). The proposed procedure optimizes a hybrid generation portfolio for each community and in an optimization environment chooses the economically preferred electrification solution for the area, with the other option being an expansion of the national grid. The focus of the procedure is scalable and robust grid-oriented solutions for relatively dense areas, which is why intermediate solutions such as stand-alone systems are not considered. In the forthcoming sub-sections, the main steps of the proposed procedure are expounded on in detail.

The methodology proposed within this paper aims to integrate with GISELE the procedures described previously, VANIA and the MTF approach for energy demand estimation, to acquire more detailed information regarding the communities to be electrified, as well as their energy needs. The division of the area in clusters (i.e. communities) and the assignment of load demand are fundamental to the entire procedure and can have a large impact not only on the economic evaluation of the electrification, but also on the means of electrification proposed for each community.

Microgrid generation portfolio sizing

This part of the procedure aims to calculate the optimal microgrid portfolio for each community and related costs, to be considered as an option in the final electrification planning. The adopted procedure is based on the MILP model presented in (Petrelli et al., 2021). The model allows the design of a hybrid off-grid system with various energy sources, including DG, PV, and BESS.

The availability of renewable energy resources is evaluated through the usage of global open-source atlases ("Renewable Ninja,", 2020). For each community, this model receives data regarding the potential RES production, energy demand estimated by the procedure in 3.2, as well as costs and specifications of the components obtained as user input. The optimization function is the minimization of the NPC for the duration of the microgrid's lifetime selected by the user, based on the deployment, maintenance, and replacement cost for each element. Additionally, a salvage value is considered for each element to account for the different lifespan of various components.

$$minNPC = \sum_i (IC_i + O\&M_i + RC_i - SV_i) \tag{4}$$

where: i = Different generator types; IC_i = Investment cost; $O\&M_i$ = Discounted annual operating and maintenance costs; RC_i = Replacing cost; SV_i = Salvage value.

Different constraints and variables are considered in the methodology, with the most important listed below.

- Power Balance:

$$\sum_b \left(P^{dch}_{h,b} \bullet \eta_b - \frac{P^{dch}_{h,b}}{\eta_b} \right) + P^{ren}_h + \sum_g P^{dg}_{h,g} + D^u_h = D_h \tag{5}$$

where: $P^{dch}_{h,b}$ = discharging power of BESS b at time h [kW]; η_b = BESS efficiency; P^{ren}_h = sum of renewable power injected into the system at hour h [kW]; $P^{dg}_{h,g}$ = power produced by diesel generator at time h [kW]; D^u_h = unmet demand [kW]; D_h = load demand [kW].

- Renewable production:

$$P^{ren}_h \leq \sum_p N_p \bullet P^{pv}_{h,p} + \sum_w N_w \bullet P^{wt}_{h,w} \tag{6}$$

where: P^{ren}_h = sum of renewable power injected into the system at hour h [kW]; $P^{pv}_{h,p}$ = per unit power available from PV at time h; $P^{wt}_{h,w}$ = per unit power available from WT at time h.

- Diesel Generators constraints:

$$FC_{h,g} = A \bullet U_{h,g} + \mathrm{B} \bullet P^{dg}_{h,g} \tag{7}$$

$$P^{dg}_{h,g} + R^{dg}_{h,g} \leq C_g \bullet U_{h,g} \tag{8}$$

$$P^{dg}_{h,g} \geq \underline{P_g} \bullet U_{h,g} \bullet C_g \tag{9}$$

$$U_{h,g} \leq N_g \tag{10}$$

where: $FC_{h,g}$ = fuel consumption of generator at time h [kW]; A = cost coefficient of the diesel generator [l/h]; $U_{h,g}$ = number of diesel generators of type g active at hour h; B = cost coefficient of the diesel generator [l/h/kW]; $P^{dg}_{h,g}$ = power produced by the diesel generator at time h [kW]; $R^{dg}_{h,g}$ = reserve to be provided by diesel generator of type g; C_g = capacity of diesel generator; $\underline{P_g}$ = minimum power of the diesel generator; N_g = number of diesel generators.

More details on the formulation of the optimization with detailed information regarding the variables, parameters, and constraints can be found in (Petrelli et al., 2021).

Electric grid design within communities

This stage of the procedure is tasked with the routing of the electric lines within each community, as well as citing the necessary distribution transformers. The procedure for the design of the electric grid draws inspiration from Luke's algorithm (Lukes, 1978) and the analysis here is purely topological, based on graph-theory and following the least-cost logic of equipment deployment. It starts with the creation of a graph (G) in which the edges (*E*) are obtained by connecting the final demand points using the Delaunay triangulation, with the weights representing the Euclidean distance. Subsequently, a tree structure is obtained by using the Prim's algorithm for obtaining the minimum spanning tree (*MST*), which effectively minimizes the deployment of lines while assuring a connected network-like structure. Moreover, the algorithm has

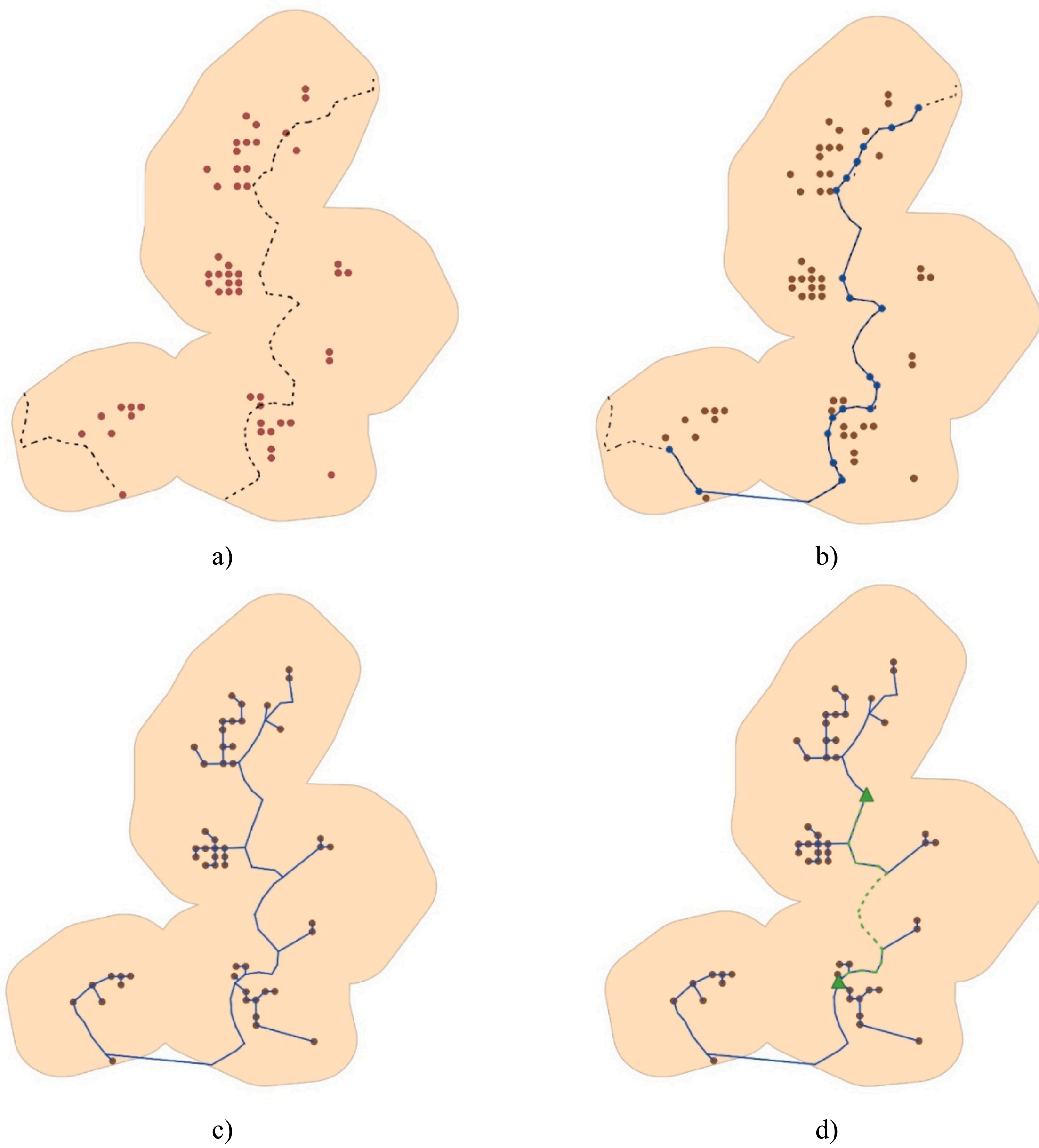


Fig. 6. Internal grid routing procedure: a) Identification of each building (red) and of the main roads (dotted black) b) candidate LV main feeder c) MST solution d) MV/LV transformers deployment and final solution.

an intermediate step which allows the consideration of roads by adding them to the graph (G) as discounted edges, in order to design a realistic structure of the distribution grid in semi-urban and urban areas by discounting the roads in a 2-step procedure detailed in Dimovski et al. (2023). This first step concludes the obtained MST as the LV lines, should the given community be selected for electrification via an off-grid system.

However, distribution is performed on the MV level, making it necessary to deploy distribution transformers and potentially, MV lines within the community. The procedure to split each community into n sub-clusters relies on agglomerative clustering, following the logic that network planners and DSOs generally adopt a parameter for the maximum length of LV lines (LV_{MAX}) when planning LV networks. This is made possible by clustering by using a complete distance matrix following the distances on the MST, rather than considering the Euclidean distances. Then, each edge of the original MST that connects populated points belonging to different sub-clusters is cut, creating n sub-trees, representing separated LV networks. Finally, the secondary substation of each LV network is placed in its weighted centroid, with a parameter that allows the user to introduce bias toward roads proximity, a common practice with DSOs for easiness of installation and maintenance. The power of the transformer is selected based on the total power requirement of the users connected to it by considering a logarithmic contemporary coefficient for a more accurate estimation of the peak load.

In case of multiple MV/LV substations within the community, the internal grid design is concluded with the deployment of MV lines. In this case, the MST algorithm is run by considering the distribution transformers as the final vertices of the tree. The entire procedure is illustrated in 4 steps on Fig. 6. To conclude, the output of this part of the procedure is the equipment deployment for each community individually, that following a pre-processing procedure, provides input to the final optimization that determines the means of electrification.

Integrated area optimization

Lastly, the third step entails the execution of a MILP model, aiming to

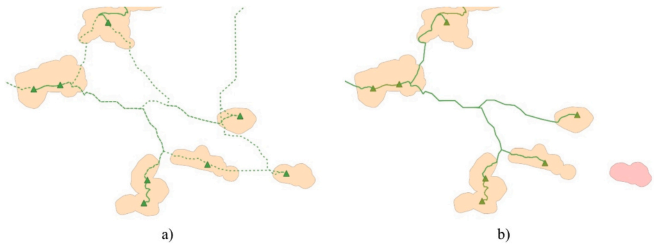


Fig. 7. Integrated area optimization procedure: a) Identification of candidate lines to connect the communities' previously determined MV grids (dotted green line); b) Final solution where the red clusters are operated as microgrids, whereas the orange ones are grid-extensions (green line).

determine the optimal electrification strategy encompassing the entire investigated region. The objective is to minimize the net present costs associated with the comprehensive electric infrastructure deployment, encompassing the expansion of the national grid, the generation portfolio of off-grid systems, and the local distribution grid within communities. Essentially, the optimization model takes on the task of determining which clusters are to be connected to the national grid, subsequently formulating the design for new distribution lines. Similarly, the model identifies clusters that are to be autonomously operated (i.e., powered by the generation portfolio developed in step 1).

The objective function (OF) subject to minimization is the following:

$$min \sum_{(i,j)\in Ad} NPC_{Ad,ij}x_{ij} + \sum_{(i)\in Mg} NPC_{Mg,i}y_i + \sum_{i\in S} NPC_{S,i}z_i + \sum_{(i)\in C}(1-y_i)D_{C,i}coe + \sum_{(i)\in C}(1-y_i)NPC_{g,i} \tag{11}$$

where: $NPC_{Ad,ij}$ = net present cost of candidate connection *(i,j)*; $x_{ij} = 1$ if connection (i; j) exists, 0 otherwise; $NPC_{Mg,i}$ = net present cost of microgrid in community c; $y_i = 1$ if microgrid is installed in node (i), 0 otherwise; $NPC_{S,i}$ = net present cost of substation s; $z_i = 1$ if substation in node (i) is used/built, 0 otherwise; $D_{C,i}$ = Energy demand of community c during microgrid's lifetime; NPC_{gli} = net present cost of the MV equipment within the communities.

The set of candidate connections (A_d) is obtained by using the Delaunay triangulation between the various communities and connection points to the existing grid, obtaining a realistic subset of the potential links. The cost and length of each line is not subject to a point-to-point connection, but its rather obtained by executing the Dijkstra algorithm on a graph-representation of the weighted grid of points explained in 3.3. An example of the preset and candidate links that are an input to the optimization procedure, as well as its output is illustrated on Fig. 7.

Looking at the objective function it can clearly be seen how the final optimization considers different costs for the internal grids and energy used based on the means of electrification selected, adding the costs of MV equipment and energy consumed only in case they would be required in the final solution. In the case of a microgrid, in the cost evaluation there is expenditure for the generation portfolio rather than energy purchased. It should be noted that due to the different lifetimes of microgrid components and conductors, a salvage value is adopted in the NPC calculation in order to put the costs on the same timeframe.

Case study & numerical simulations

The proposed framework was then applied on a real-life study case in the Federal Democratic Republic of Ethiopia. This country is characterized by 85 % of people living in rural areas (*ETHIOPIA Data Portal*, n. d.) and a share of 51.1 % electricity access in 2020, a percentage that can reach 39.4 % of the population in rural areas ("World Bank - Data,", 2020; Douglas et al., 2016). The analysis presented in this chapter

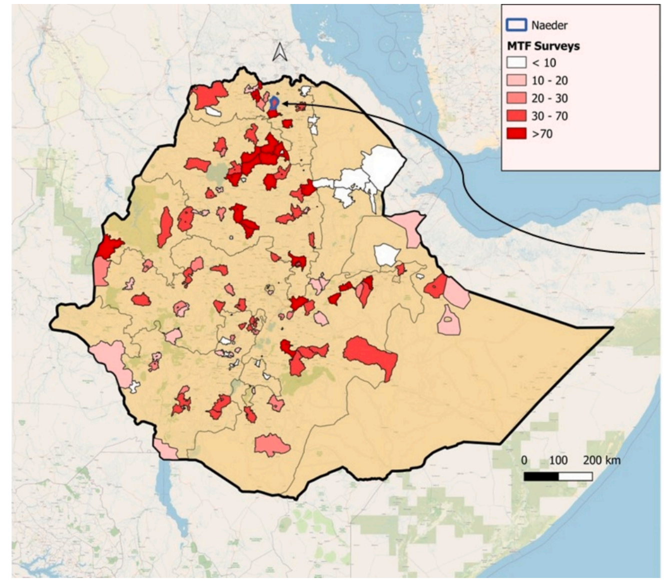


Fig. 8. Spatial distribution of MTF surveys in Ethiopia.

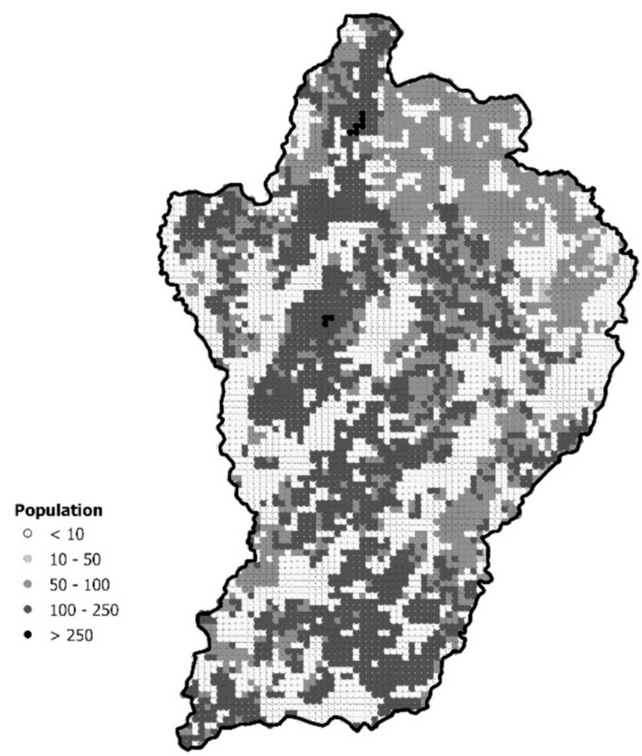


Fig. 9. Population distribution in Naeder.

consists of two parts: a) Leveraging the clustering, acquisition of socio-economic attributes and MTF surveys to provide an estimation for the energy demand of various communities. Owing to the abundance of data availability combined with low electrification rates, the Tigray region in the north of the country was selected as the most appropriate and interesting candidate for the energy demand assessment. b) Using the analysis of the electricity requirements, provide a comprehensive electrification planning for an area of interest. For this segment of the analysis, the authors chose the Naeder province in Tigray to offer a more concise and concentrated case study. It spans 500 km^2 situated near the national grid with very low electricity access in general and without an

Table 4
Summary of outputs from VANIA[1] and MTF[2] for a subset of communities in Tigray.

Community	VANIA					MTF					
	Distance from city[1] [km]	Wealth Index[1]	HDI[1]	Buildings count[1]	Population[1]	Tier 1[2]	Tier 2[2]	Tier 3[2]	Tier 4[2]	Tier 5[2]	Daily Consumption [Wh/HH]
Degehabur	190.79	−0.52	0.4	231	603	0	0.08	0.5	0.17	0.25	3133
Kebribeyah	162.99	−0.47	0.4	284	766	0	0.04	0.48	0.27	0.21	3116
Zequala	441.72	−0.45	0.5	194	343	0	0.73	0.27	0	0	416
Janamora	1179.49	−0.45	0.5	243	956	0.5	0.5	0	0	0	106
Asayita	308.73	−0.44	0.4	875	3693	0.02	0.27	0.58	0.13	0	1062
Bulen	768.94	−0.41	0.5	413	1218	0.17	0.67	0.17	0	0	302
Abaala	89.67	−0.4	0.4	131	898	0.08	0.46	0.29	0	0.17	1751
Ambasel	327.93	−0.4	0.5	358	1118	0	0.33	0.5	0.17	0	1133

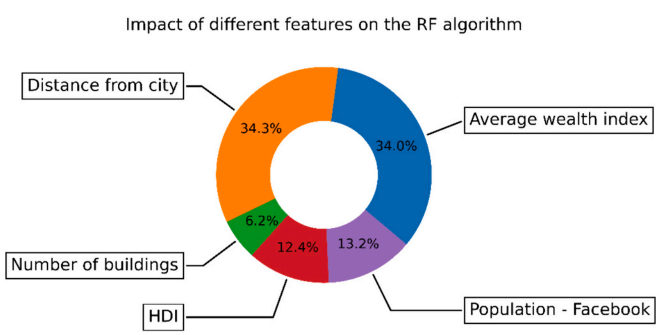


Fig. 10. Importance of features in RF algorithm.

expanded distribution network. Concentrating on this particular province will not only provide a detailed electrification strategy for a predominantly unelectrified area, but also enable a more profound understanding of the operational procedures involved. Moreover, the output of VANIA for the entire Ethiopia is available upon request.

Communities' identification and energy needs estimation

Implementing the procedure within the Tigray region involved adopting the three-step approach outlined in the preceding chapter. The Ethiopia MTF surveys acquired responses from 4317 individuals representing their respective household that consisted of 4.5 members on average, fairly close to the official national average of 4.6 members, distributed across 225 communities. The spatial distribution of MTF surveys across the country is illustrated on Fig. 8, revealing a notable concentration in the northern part encompassing Tigray and Naeder.

Having defined the boundaries of the communities for which MTF surveys were available, the data was utilized to allocate the share of households in each tier. Notably, the MTF surveys provide insights into the distribution of households across different tiers, that are used as proxy for assessing the energy demand. Moreover, Fig. 9 represents the spatial distribution of population in Naeder, which is another parameter that has a significant impact on the energy demand. Utilizing the knowledge of the tier distribution and population estimates of each community, its total energy consumption can be estimated. It is noteworthy that a significant portion of households predominantly falls within Tier 2 and Tier 3 categories. On the other hand, the procedure detailed in 3.1 has provided key socio-economic indicators, facilitating a correlation between them. An example of this pre-processing for a selected set of communities is presented on Table 4, where the indicators are shown on one side, and on the other side, the distribution of tiers and the estimation of the average daily energy demand of a household (HH) that come as a result of the MTF surveys processing.

Proceeding to the second step, the task entails estimating the daily energy requirements for each individual community located by VANIA in the Tigray region. Within this subset, 80 % of the communities were designated as a training dataset, while the remaining 20 % constituted the testing dataset. Having displayed the highest degree of stability in terms of performance and lowest error, the RF was utilized as described in 3.2.

Interestingly, as reported in Fig. 10, the features carrying the greatest weight for the RF were identified as the distance from the nearest city and the average wealth index. These features distinctly emerged as the most influential factors in the classification of energy demand.

Then, the trained RF model was employed to compute the anticipated energy consumption, subsequently juxtaposed against the actual energy requirements. An example of the comparison between real and estimated energy demand for 20 communities used as a validation set is illustrated in Fig. 11. The obtained results are subject to an RMSE value of 207 Wh and a MAE of 138 Wh, which is a reasonably accurate estimate compared to the actual demand and more than fitting for the

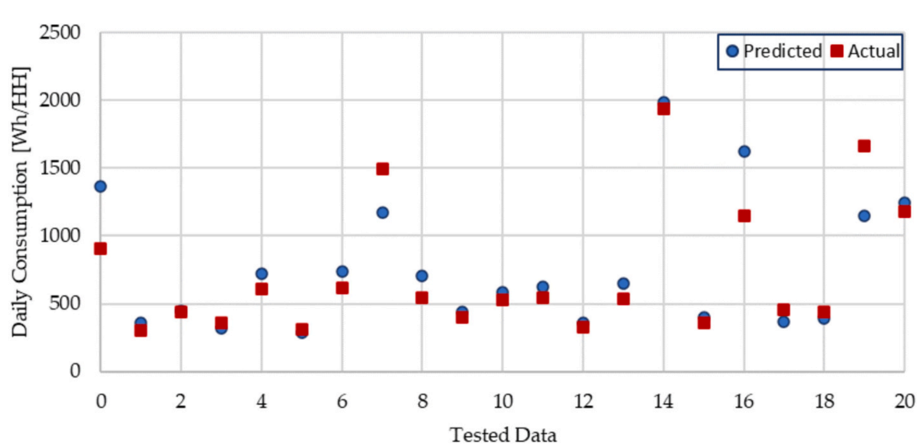


Fig. 11. Comparison of predicted and actual consumption for a validation set.

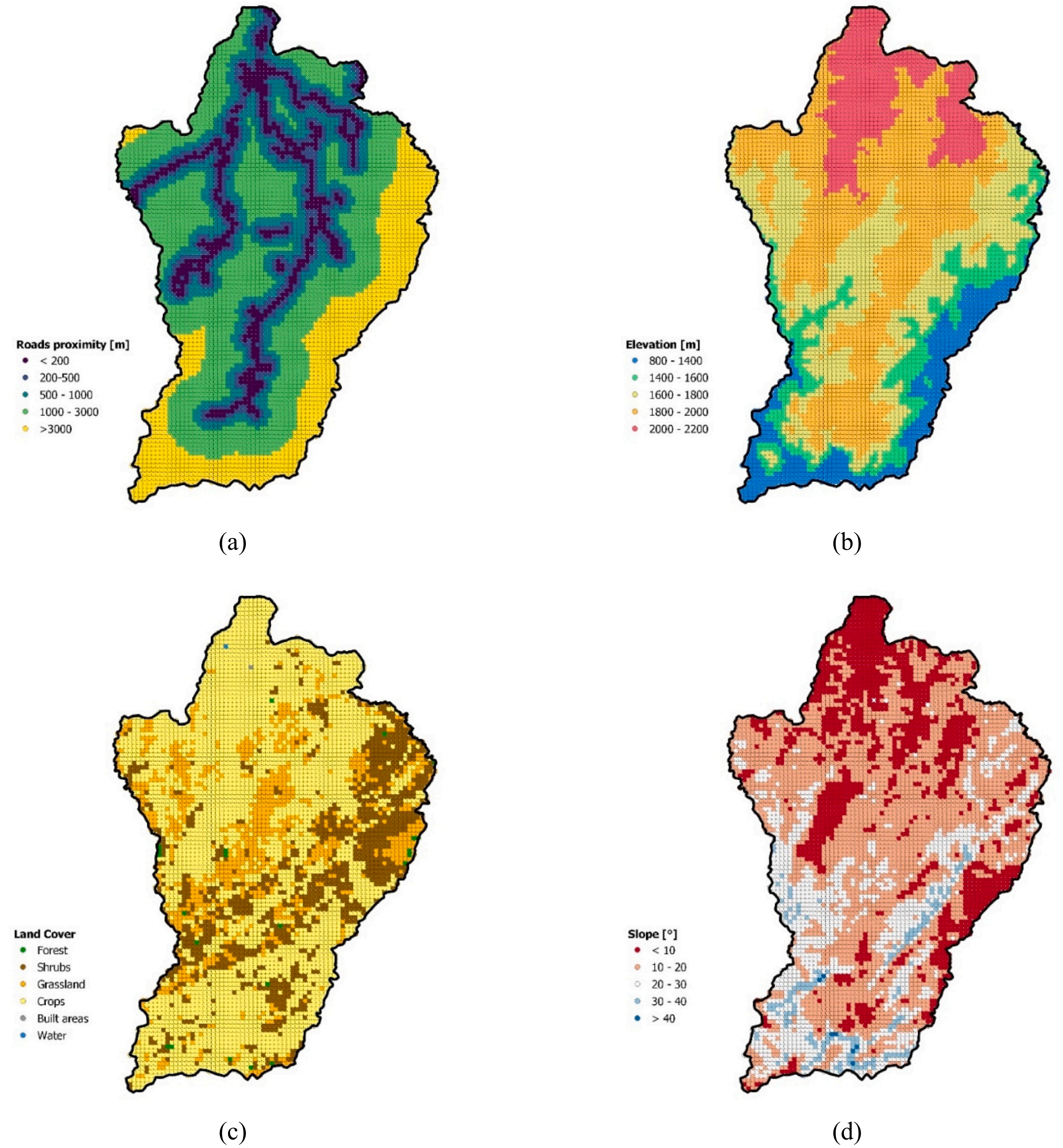


Fig. 12. Grid of point parametrization based on a) road distance, b) elevation, c) land cover, d) slope.

subsequent electrification planning.

Finally, the trained RF algorithm was adopted to estimate the daily energy demand for each community within the Tigray province of Ethiopia. To prevent overloading the paper, the detailed outcome of the energy demand assessment are omitted here, but are available in Pezham and Ahmadi (2022). The procedure provided an estimate of 1315 Wh average electricity demand per household, ranging between 330 and 2700 Wh based on their socio-economic characterization. The average household energy demand for the communities within Naeder was estimated at 1420 Wh, slightly higher than Tigray's average.

These estimations exhibit a significant range of variation and it's clear how considering just a unique energy demand may lead to a sub-optimal electrification strategy and misguide stakeholders on the required resources. To provide a clear parallel, communities with a lower household demand and low demographic density generally tend to lean toward off-grid solutions to avoid capital investments in over-designed infrastructure.

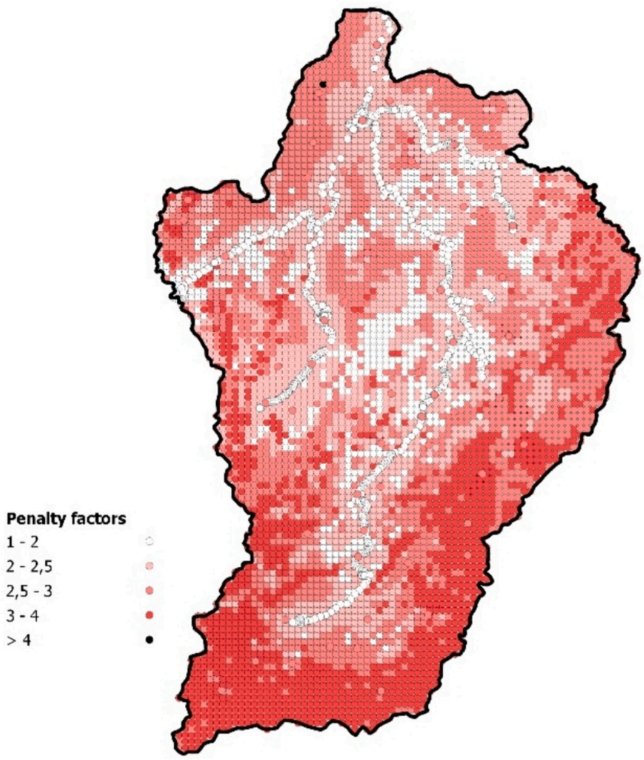


Fig. 13. Cost surface representation of Naeder.

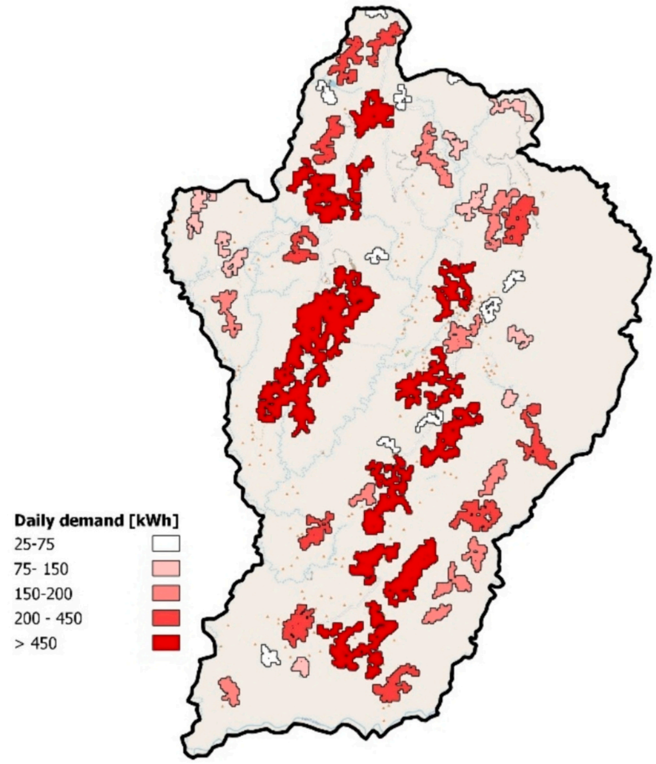


Fig. 14. Daily energy demand of the communities in Naeder.

Electrification planning

In the concluding phase of the electrification planning, the design of both the internal electric grid within each community and the potential interconnection with the national grid is executed. The population in Naeder was clustered into 50 communities, providing electrification guidelines for approximately 52,150 inhabitants. This task commences with the representation of the Naeder province through its principal spatial attributes, as illustrated in Fig. 12. These attributes encompass distance from roads (a), elevation effects (b), land cover (c), and slope (d), all of which significantly influence the feasibility of deploying electric distribution lines. These spatial characteristics were directly employed to formulate the weighted grid of points, as detailed in 3.3. The obtained cost surface can be visualized on Fig. 13, whereas the energy estimation in the first year of electrification for the various communities in Naeder is displayed on Fig. 14. The grid routing algorithm subsequently utilizes the cost surface to determine the optimal and cost-effective deployment of distribution lines.

In order to identify the optimal electrification strategy, the simulation relies upon the following assumptions: annual demand growth of 2 %, fuel cost of 0.75 € per liter, nominal voltage of 15 kV for distribution lines, grid expected lifetime set to 20 years, microgrid expected lifetime set to 10 years and a maximum allowed length of low voltage lines equal to 2 km. The electricity tariff in Ethiopia is heavily subsidized and as such is not a good representative of the actual costs of energy, which is why it was substituted for a more realistic value of 100 €/MWh. Following guidelines from stakeholders close to electrical utilities, the base cost of MV lines is set to 10,000 €/km, whereas the cost for LV lines is equal to 8000 €/km. In terms of MV/LV substations, the ones selected within the case study are 50 and 100 kVA, with costs of 1500 and 2100€, respectively. The complete list of parameters and related costs utilized in the optimizations is available in the annexures in (Dimovski et al., 2023).

As delineated in preceding chapters, the presented framework optimizes the microgrid generation portfolio for each cluster, tailored to fulfill the specific local energy demands. The load profiles used to optimize the generation portfolio were obtained with the tool RAMP (Lombardi et al., 2019) starting from the energy needs of each community. Given this procedure, the average household is estimated to consume a peak power of 165 W, which is in line with the estimations of the MTF for a user that is between Tiers 2 and 3. The result obtained an equal distribution, 50 % each, between PV + BESS and DG + PV + BESS microgrids, with a preference for the latter in case of communities with higher energy demand. The generation portfolio obtained and related costs were subsequently compared with the expenses associated with the expansion of the distribution grid adopting the integrated optimization presented in 3.3.3.

Fig. 15 shows Naeder's optimal electrification strategy suggested by the proposed procedure. Communities shaded in light yellow correspond to communities slated for connection to the national grid, facilitated by the green MV distribution lines. Meanwhile, blue shapes denote communities recommended for electrification via local stand-alone microgrids. A higher resolution focus is provided to illustrate the level of detail of the solution proposed. Within the scope of the study, the procedure selected 11 out of 50 communities for electrification via stand-alone microgrids and local generation, whereas the remaining communities were connected to the national grid. Considering Figs. 13 and 14, it becomes evident that the communities proposed for off-grid electrification are the ones in relatively difficult terrain, with a lower energy demand. In terms of electrical infrastructure required for extending the national grid, the overall electrification will require the deployment of 54 MV/LV substations to supply an estimated peak load of 686 kW, 515.8 km of LV lines and 139.35 km of MV lines.

Expecting an accurate prediction of the future would be extremely impractical, which is why policymakers are typically charged with developing scenarios to address the uncertainty of input parameters and their impact on the outcomes. Sensitivity analysis offers a practical way to deal with parameter uncertainty in situations where stochastic optimization would be computationally infeasible, as exemplified by this methodology. With this in mind, the analysis was deepened with a

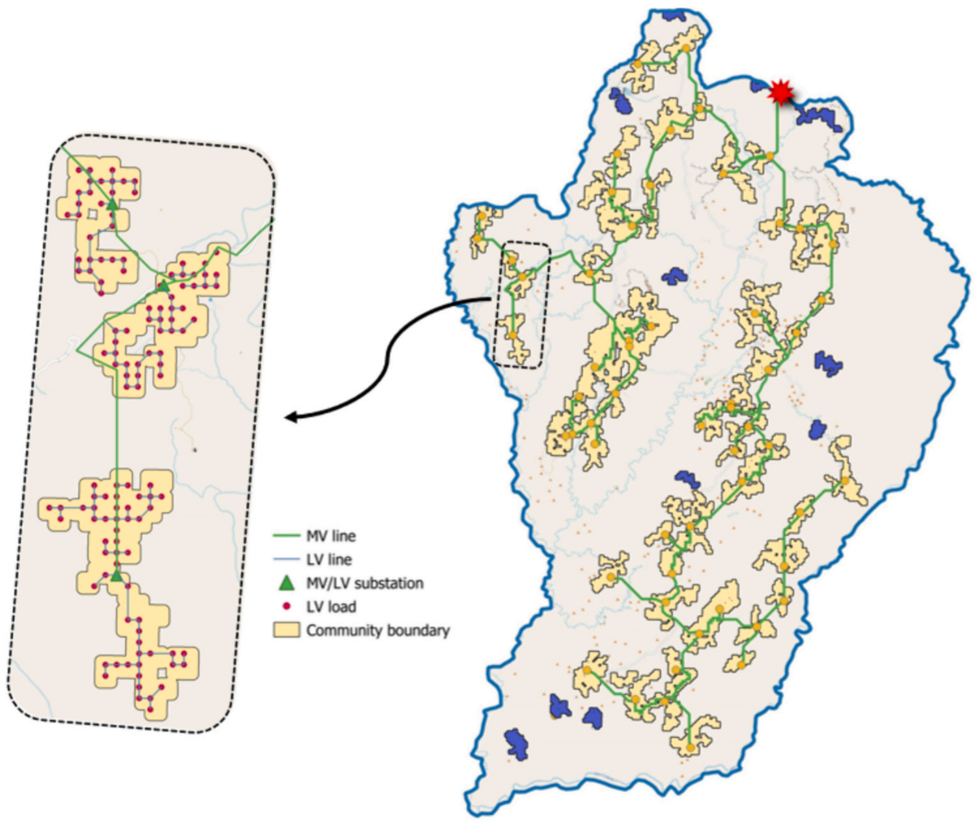
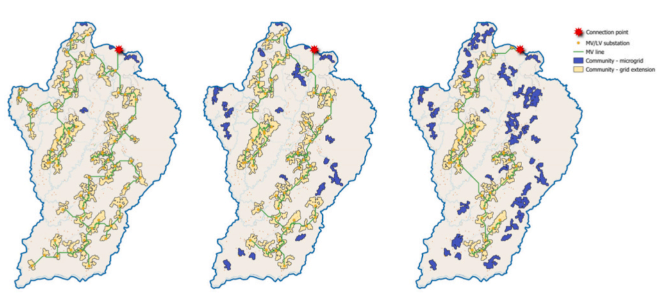


Fig. 15. Optimal electrification strategy for the area under investigation.

Table 5
Outcome of sensitivity analysis on the cost of energy.

CoE [€/MWh]	Number of microgrids	Microgrid CAPEX [k€]	MV Network Length [km]	MV Network CAPEX[a] [M€]	Peak power[b] [kW]	Population electrified via extension
80	3	81	155.3	2.91	719	51,708
90	4	129	153	2.84	715	51,440
100	5	176.5	150.45	2.79	711	51,190
110	11	489.2	139.35	2.48	686	49,479
120	11	489.2	139.35	2.48	686	49,479
130	20	1180	118.35	1.96	621	44,880
140	37	3792	78.75	1.15	455	33,050
145	50	6422	41	0.44	0	0

[a] Costs are inclusive of both the electric lines and MV/LV substations.
[b] Power requested from the national grid in the first year of electrification.



(a) CoE = 90 €/MWh (b) CoE = 130 €/MWh (c) CoE = 140 €/MWh

Fig. 16. Sensitivity of CoE on the optimal electrification strategy.

sensitivity analysis conducted on the cost of electricity purchased from the national grid, investigating prices ranging between 80 €/MWh and 145 €/MWh. In recent years the cost of electricity has been extremely volatile and it's paramount to analyze its effect on the optimal electrification solution.

The complete results from this sensitivity analysis are presented in Table 5, whereas the georeferenced solution for three different prices of electricity is displayed on Fig. 16. As anticipated, the conducted simulations illustrate that with an increase in the cost of electricity supplied by the national grid, a larger number of dispersed communities could be more economically served by local stand-alone microgrids. The analysis has shown that 3 of the 50 communities have a demand so low that given the surrounding terrain, a MV extension would not be economically justified in any case. Interestingly, fluctuations between 80 and 100 €/MWh show minor changes in the electrification strategy, with an increased preference toward off-grid solutions if the price reaches 110 €/MWh. Moreover, variations between 110 and 120 €/MWh do not have an impact on the final solution which suggests that the next communities in line for off-grid systems show a significant jump in investment cost for the generation portfolios required. Cost of electricity upwards from 120 €/MWh result in a drastic preference toward off-grid systems, with the price of 145 €/MWh being the break point above which it is not cost-effective to extend the national grid at all, but rather electrify the entire province with microgrids.

The final solution must depend on the adequacy of the underlying system to host the additional request, which makes it important to note that the optimal strategies between 80 and 130 €/MWh lead to a difference of 98 kW of peak load requested from the system, which is a rather insignificant impact.

Conclusion

This paper addressed the challenge of optimizing electrification planning in rural areas, a task that confronts key difficulties such as data reliability and the accurate assessment of energy requirements and load demand profiles. The central aim of this study was to assess the optimal electrification design for rural regions, addressing these challenges through a three-step procedure. In the initial phase, a novel approach named VANIA is introduced to identify communities within an area of interest and characterize them with key socio-economic indicators. The second step employs an integration of the MTF and a ML algorithm to estimate the energy requirements of each identified community establishing non-linear correlations between the socio-economic indicators and the energy requirements obtained from the MTF surveys. The case study analyzed demonstrated that the most significant correlations with the communities' energy requirements are associated with indicators such as distance from the nearest city and the average wealth status. The final step leverages an enhanced version of the tool GISELE to ascertain the optimal electrification strategy. Moreover, having understood that the cost of electricity is extremely volatile in post-COVID times, the flexibility of the methodology was demonstrated by conducting sensitivity analysis on the price of electricity absorbed from the national grid. In fact, this confirmed the hypothesis of the sensitivity of the final solution to variations of the input parameters. The analysis showed how the valuation of the cost of energy supplied from the national grid may impact the optimal electrification strategy. Indeed, variations between 80 and 145 €/MWh showed a significant influence, starting from a strong preference toward grid extension, to a complete electrification using isolated microgrids. Overall, the comprehensive optimization tools proposed in this paper hold the potential to streamline decision-making processes for defining the most suitable long-lasting electrification strategy in rural areas, benefiting both local and global stakeholders. The main challenge to the widespread adoption of this framework is the lack of accurate and accessible geo-referenced data, which is something that has shown improvement over time.

CRediT authorship contribution statement

Aleksandar Dimovski: Writing – review & editing, Software, Methodology, Conceptualization. **Zahra Pezham:** Writing – original draft, Investigation, Conceptualization. **Mohammad Ahmadi:** Writing – original draft, Investigation, Conceptualization. **Lorenzo Maria Filippo Albertini:** Writing – review & editing, Validation, Formal analysis. **Darlain Irenee Edeme:** Writing – original draft, Software, Methodology, Data curation. **Marco Merlo:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Amatya, R., Barbar, M. F., Brusnahan, M., Ciller Cutillas, P., Cuadra Garcìa, F., Dueñas Martínez, P., … Vergara Ramírez, C. R. (2019). *Computer-aided electrification planning in developing countries: The Reference Electrification Model (REM)*.

Avila, N., Carvallo, J. P., Shaw, B., & Kammen, D. M. (2017). *The energy challenge in sub-Saharan Africa: A guide for advocates and policy makers*. OXFAM: Research Backgrounder, Washington.

Bhatia, M., & Angelou, N. (2015). *BEYOND CONNECTIONS - Energy Access Redefined*. Washington, DC: ESMAP.

Blechinger, P., Cader, C., & Bertheau, P. (2019). Least-cost Electrification Modeling and Planning—A Case Study for five Nigerian Federal States. *Proceedings of the IEEE, 107*, 1923–1940. https://doi.org/10.1109/JPROC.2019.2924644

Bosisio, A., Moncecchi, M., Morotti, A., & Merlo, M. (2021). Machine Learning and GIS Approach for electrical load Assessment to increase distribution Networks Resilience. *Energies, 14*, 4133. https://doi.org/10.3390/en14144133

Cader, C., Blechinger, P., & Bertheau, P. (2016). Electrification Planning with Focus on Hybrid Mini-grids – A Comprehensive Modelling Approach for the Global south. *Energy Procedia, 99*, 269–276. https://doi.org/10.1016/j.egypro.2016.10.116

Caquilpan, V., Saez, D., Hernandez, R., Llanos, J., Roje, T., & Nunez, A. (2017). Load estimation based on self-organizing maps and Bayesian networks for microgrids design in rural zones. In *2017 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America)* (pp. 1–6). IEEE. https://doi.org/10.1109/ISGT-LA.2017.8126709.

Chen, X., Zheng, H., Wang, H., & Yan, T. (2022). Can machine learning algorithms perform better than multiple linear regression in predicting nitrogen excretion from lactating dairy cows. *Scientific Reports, 12*, 12478. https://doi.org/10.1038/s41598-022-16490-y

Ciller, P., Ellman, D., Vergara, C., Gonzalez-Garcia, A., Lee, S. J., Drouin, C., Brusnahan, M., Borofsky, Y., Mateo, C., Amatya, R., Palacios, R., Stoner, R., de Cuadra, F., & Perez-Arriaga, I. (2019). Optimal Electrification Planning Incorporating On- and Off-Grid Technologies: The Reference Electrification Model (REM). *Proceedings of the IEEE, 107*, 1872–1905. https://doi.org/10.1109/JPROC.2019.2922543

Cinelli, M., Burgherr, P., Kadziński, M., & Słowiński, R. (2022). Proper and improper uses of MCDA methods in energy systems analysis. *Decision Support Systems, 163*, Article 113848. https://doi.org/10.1016/j.dss.2022.113848

Corigliano, S., Carnovali, T., Edeme, D., & Merlo, M. (2020). Holistic geospatial data-based procedure for electric network design and least-cost energy strategy. *Energy for Sustainable Development, 58*, 1–15. https://doi.org/10.1016/j.esd.2020.06.008

Correa, S., Shah, Z., & Taneja, J. (2021). *This Little Light of Mine, in: Proceedings of the Twelfth ACM International Conference on Future Energy Systems* (pp. 254–258). New York, NY, USA: ACM. https://doi.org/10.1145/3447555.3464871

Dimovski, A., Corigliano, S., Edeme, D., & Merlo, M. (2023). Holistic MILP-based approach for rural electrification planning. *Energy Strategy Reviews, 49*, Article 101171. https://doi.org/10.1016/j.esr.2023.101171

Douglas, B., Golumbeanu, R., & Diaw, I. (2016). *Beyond Electricity Access: Output-based Aid and Rural Electrification in Ethiopia*. World Bank.

Energy access investment, agricultural profitability, and rural development: Time for an integrated approach. *Environmental Research: Infrastructure and Sustainability, 1*(3). (2021).

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd, 96*, 226–231.

ETHIOPIA Data Portal, n.d. Ethiopia Data At-a-Glance. URL https://ethiopia.opendataforafrica.org/.

Farret, F. A., & Godoy, S. M. (2005). Micropower system modeling with HOMER. In *Integration of Alternative sources of Energy*. John Wiley & Sons.

Foster, V., & Tre, J. P. (2000). *Measuring the Impact of Energy Interventions on the Poor—An Illustration from Guatemala*. Infrastructure for Development: Private Solutions and the Poor, London, UK.

Garfias Royo, M., Diep, L., Mulligan, J., Mukanga, P., & Parikh, P. (2022). Linking the UN Sustainable Development Goals and African agenda 2063: Understanding overlaps and gaps between the global goals and continental priorities for Africa. *World Development Sustainability, 1*, Article 100010. https://doi.org/10.1016/j.wds.2022.100010

Gershenson, D., Rohrer, B., & Lerner, A. (2019). *EnergyData: Medium-voltage distribution (predictive)* [WWW Document]. Facebook - META. URL https://energydata.info/dataset/medium-voltage-distribution-predictive. (Accessed 13 November 2022).

GISele. (2022). E4Growing. https://github.com/Energy4Growing/gisele_v02.

Google. (2022). Open buildings dataset [WWW Document] https://sites.research.google/open-buildings/. (Accessed 15 November 2022).

HOMER Energy LLC, n.d. HOMER pro - Microgrid Software for Designing Optimized Hybrid Microgrids [WWW Document]. URL http://www.homerenergy.com/ (accessed 11.16.22).

IEA. (2022a). *SDG7: Data and Projections*. Paris: International Energy Agency.

IEA. (2022b). *World Energy Investment 2022*. Paris: International Energy Agency.

IED, n.d. GEOSIM - Geospatial Rural Electrification Planning [WWW Document]. Innovation Energie Developpement. URL https://www.ied-sa.com/en/products/planning/geosim-gb.html#references (accessed 11.14.22).

Isihak, S., Akpan, U., & Bhattacharyya, S. (2022). Evolution of GIS-based rural electrification planning models and an application of OnSSET in Nigeria. *Renewable and Sustainable Energy Transition, 2*, Article 100019. https://doi.org/10.1016/j.rset.2022.100019

Khaitan, S. K., & Gupta, A. (2013). *High performance computing in power and energy systems*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-32683-7

Khavari, B., Korkovelos, A., Sahlberg, A., Howells, M., & Fuso Nerini, F. (2021). Population cluster data to assess the urban-rural split and electrification in Sub-Saharan Africa. *Scientific Data, 8*, 117. https://doi.org/10.1038/s41597-021-00897-9

Li, X., Zhou, Y., Zhao, M., & Zhao, X. (2020). A harmonized global night-time light dataset 1992–2018. *Scientific Data, 7*, 168. https://doi.org/10.1038/s41597-020-0510-y

Liaw, A., & Wiener, M. (2002). *Classification and Regression by randomForest. R NEWS*, 2/3.

Loken, E. (2007). Use of multicriteria decision analysis methods for energy planning problems. *Renewable and Sustainable Energy Reviews, 11*, 1584–1595. https://doi.org/10.1016/j.rser.2005.11.005

Lombardi, F., Balderrama, S., Quoilin, S., & 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

Lukes, J. A. (1978). Efficient algorithm for the partitioning of trees. *IBM Journal of Research and Development, 18*(3), 217–224.

Mandelli, S., Barbieri, J., Mereu, R., & Colombo, E. (2016). Off-grid systems for rural electrification in developing countries: Definitions, classification and a comprehensive literature review. *Renewable and Sustainable Energy Reviews, 58*, 1621–1646. https://doi.org/10.1016/j.rser.2015.12.338

Mateo Domingo, C., Gomez San Roman, T., Sanchez-Miralles, A., Peco Gonzalez, J. P., & Candela Martinez, A. (2011). A Reference network model for large-scale distribution planning with automatic street map generation. *IEEE Transactions on Power Systems, 26*, 190–197. https://doi.org/10.1109/TPWRS.2010.2052077

Mentis, D., Howells, M., Rogner, H., Korkovelos, A., Arderne, C., Zepeda, E., … Scholtz, E. (2017). Lighting the World: The first application of an open source, spatial electrification tool (OnSSET) on Sub-Saharan Africa. *Environmental Research Letters, 12*, Article 085003. https://doi.org/10.1088/1748-9326/aa7b29

Meta. (2022). Data for good [WWW Document]. URL https://dataforgood.facebook.com/. (Accessed 14 November 2023).

Monteiro, C., Ramirez-Rosado, I. J., Miranda, V., Zorzano-Santamaria, P. J., Garcia-Garrido, E., & Fernandez-Jimenez, L. A. (2015). GIS spatial analysis applied to electric line routing optimization. *IEEE Transactions on Power Delivery, 20*, 934–942. https://doi.org/10.1109/TPWRD.2004.839724

Moshi, G. G., Bovo, C., Berizzi, A., & Taccari, L. (2016). Optimization of integrated design and operation of microgrids under uncertainty. In *2016 Power Systems Computation Conference (PSCC)* (pp. 1–7). IEEE. https://doi.org/10.1109/PSCC.2016.7540870.

Mullen, C., & Wade, N. (2020). *Appliance Data and Multi-Tier Framework for household electrical load modelling*. Newcastle: Newcastle University.

Mwakitalima, I. J., & King'ondu, C. K. (2015). Electricity demand evaluation for rural electrification: A case study of Kikwe Village in Tanzania. *International Journal of Engineering Research & Technology (IJERT), 4*, 1025–1028.

NASA, n.d. Earth Observatory - NASA. URL https://earthobservatory.nasa.gov/features/NightLights.

Natali, C., & Carbajal, J. (2017). *Sustainable engineering lab - network planner* [WWW Document]. Columbia UNiversity - The Earth Institute. URL https://qsel.columbia.edu/network-planner/. (Accessed 14 November 2022).

Network Planner [WWW Document], n.d. URL https://network-planner.com/ (accessed 12.13.22).

Palit, D., & Chaurey, A. (2011). Off-grid rural electrification experiences from South Asia: Status and best practices. *Energy for Sustainable Development, 15*, 266–276. https://doi.org/10.1016/j.esd.2011.07.004

Petrelli, M., Fioriti, D., Berizzi, A., & Poli, D. (2021). Multi-year planning of a rural microgrid considering storage degradation. *IEEE Transactions on Power Systems, 36*, 1459–1469. https://doi.org/10.1109/TPWRS.2020.3020219

Pezham, Z., & Ahmadi, M. (2022). *MTF to Rural Electrification: A procedure to Estimate Electric load demand based on Open-Source Data (Master Thesis)*. Politecnico di Milano.

Renewable Ninja [WWW Document]. URL https://www.renewables.ninja/, (2020)–. (Accessed 24 November 2022).

Riva, F., Gardumi, F., Tognollo, A., & 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

Sarhani, M., & El Afia, A. (2015). Electric load forecasting using Hybrid Machine Learning Approach Incorporating Feature selection. In *International Conference on Big Data Cloud and Applications*. https://api.semanticscholar.org/CorpusID:10991809.

SEforALL. (2016). Multi-Tier Framework for tracking energy access - Update on the framework design and implementation of global surveys [WWW Document]. *Sustainable Energy for All.* https://www.seforall.org/events/multi-tier-framework-for-tracking-energy-access-update-on-the-framework-design-and (accessed 11.16.22).

Setiawan, A., Koprinska, I., & Agelidis, V. G. (2009). Very short-term electricity load demand forecasting using support vector regression. In *2009 International Joint Conference on Neural Networks* (pp. 2888–2894). IEEE. https://doi.org/10.1109/IJCNN.2009.5179063.

Tella, O. (2018). Agenda 2063 and its Implications for Africa's Soft Power. *Journal of Black Studies, 49*, 714–730.

The World Bank. (2013). *Toward a Sustainable Energy Future for all: Directions for the World Bank Group's Energy Sector*. Washington D.C.: World Bank Group.

The World Bank. (2022). *The World Bank in Africa: Overview*.

Uamusse, E., Mondlane, F., Santos, I., Ganhane, M., & Américo, A. (2017). *Carteira de projectos de energia renovaveis: recurso hidrico e solar*. Maputo, Mozambique: FUNAE.

Watchueng, S., Jacob, A., & Frandji, R. (2010). *Planning tools and methodologies for rural electrification*. Group of African Agencies and Structures in charge of Rural Electrification (Club-ER). https://www.wame2030.org/files/catalogue/2019/4/CLUBER_3_outilplanif_gb_bd_1.pdf.

WHO Media inquiries. (2022). *COVID-19 slows progress toward universal energy access*. Geneva, Switzerland: World Health Organisation.

Winters-Hilt, S., & Merat, S. (2007). SVM clustering. *BMC Bioinformatics, 8*, S18. https://doi.org/10.1186/1471-2105-8-S7-S18

Tatem, A. J. (Jan. 2017). WorldPop, open data for spatial demography. *Scientific Data, 4* (1), Article 170004. https://doi.org/10.1038/sdata.2017.4

World Bank - Data [WWW Document]. URL https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS?locations=ET%2C&name_desc=true, (2020)–. (Accessed 13 December 2022).

Zhang, Y., Liu, J., & Shen, W. (2022). A Review of Ensemble Learning Algorithms used in Remote Sensing applications. *Applied Sciences, 12*, 8654. https://doi.org/10.3390/app12178654