Energy Strategy Reviews 49 (2023) 101171

Contents lists available at ScienceDirect

Energy Strategy Reviews

journal homepage: www.elsevier.com/locate/esr

Holistic MILP-based approach for rural electrification planning

Aleksandar Dimovski *, Silvia Corigliano, Darlain Edeme, Marco Merlo

Department of Energy, Politecnico di Milano, Milano, Italy

ARTICLE INFO

Handling Editor: Xi Lu

Keywords:
Rural electrification
MILP
Sustainable development
Least-cost electrification
Geospatial planning

ABSTRACT

This study proposes a procedure that can help address the challenge of achieving universal energy access by facilitating the process of rural electrification. Our holistic methodology enables the analysis of electrification processes and tailored for the usage of open-source data, allowing for efficient "green-field" and "brown-field" expansion planning. To apply this methodology to a non-electrified rural area, population distribution, terrain, existing infrastructure, and accessible information from public databases are necessary inputs. This methodology adopts a cluster-oriented approach and utilises multi-parameter cost surfaces to represent terrain in order to perform routing of medium-voltage (MV) and low-voltage (LV) networks for unelectrified communities. The solution from the routing algorithm is enhanced utilising advanced features such as road tracking and pole-sharing. Following the energy demand assessment of each community, both the options of interconnection to the national grid and supplying loads with standalone microgrids were evaluated and compared to define the most cost-effective solution. This study proposes a combination of bottom-up and top-down approaches, where the solution for each community separately, in terms of the optimal microgrids' generation portfolio and grid routing, is then used as input for an integrated optimisation to obtain a realistic and cost-effective electrification plan. The overall electrification is achieved by solving a nonconvex mixed integer linear programming (MILP) optimisation constrained by electrical parameters, considering the loading of elements and voltage drops along the lines. Finally, the procedure provides an estimate of the overall materials and costs incurred, as well as a georeferenced deployment of electrical equipment. To validate the procedure and demonstrate its feasibility, the approach was applied to the electrification plan of the Butha-Buthe region in Lesotho. The proposed tool is shared on GitHub to support the adoption of a comprehensive electrification approach.

1. Introduction

Energy access is essential for various fundamental aspects of modern society, including cooking, heating, sanitation, telecommunication services, and transportation. Consequently, the United Nations (UN) has placed energy access at the centre of many of its Sustainable Development Goals (SDGs), with 17 targets that nations worldwide have been called to achieve by 2030. The 7th Goal of the SDGs, Affordable and Clean Energy [1], explicitly focuses on energy. Its subgoals include ensuring universal access to affordable and reliable energy services, substantially increasing the share of renewable energy in the global mix and improving energy efficiency [2].

Progress has been made over time, with the number of people without electricity dropping to 770 million in 2019. However, this progress has not been across the world; Southeast Asia and South America have almost reached 100% electricity access, whereas 75% of the global unelectrified population resides in sub-Saharan Africa, a share that has unfortunately risen over the years. Consequently, this region is the poorest in the world, with an average GDP per capita of approximately $1500, without a growth trend since 2014 [3]. This part of Africa requires significant planning efforts to identify the optimal path forward, and this has motivated research on least-cost electrification planning.

Nevertheless, the task of planning the development of electrical grids is a complex and multidimensional problem, owing to the various uncertainties related to socio-political developments and the lack of expertise required to operate the installed equipment. This can be observed in the low reliability of the existing national grids in this area [4]. Therefore, the scientific community and stakeholders should focus not only on deploying grids but also on education and overall development. With that being said, providing economically feasible and sustainable electrification solutions remain a priority. This is even more emphasised in rural areas, where, electrification can be even less economically justified because of the sparsity of the population and low

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

https://doi.org/10.1016/j.esr.2023.101171
Received 3 July 2022; Received in revised form 14 May 2023; Accepted 22 May 2023
Available online 29 August 2023

income levels [5].

The goal of the proposed procedure in this paper is not only to identify the optimal electrification path for rural areas but also to aid in the assessment of the overall cost for electrification, providing stakeholders with information for assessing various areas and identifying priorities. A key focus is the electric grid structure and topology, which is typically under investigated in approaches presented in the literature. The remainder of this paper is organised as follows. Section 2 presents a critical overview of the current status of literature, existing approaches, and tools for rural electrification. Section 3 describes the proposed methodology, with greater focus on topics that are expected to be the main contributions of this study. Sections 4 and 5 present a case study of Butha-Buthe in Lesotho and its results. Finally, Section 6 draws conclusions, discusses the limitations of the proposed approach, and discusses topics that need to be addressed for future improvements.

2. Literature overview

In recent years, research has actively studied solutions and provided instruments that can assist stakeholders in the complex and multivariate tasks of rural electrification [6,7]. According to Ref. [8], the complete technical design of any new electrification strategy in rural areas should comprise the following four points: (1) identification of the optimal electrification solution, including choosing between on- and off-grid solutions; (2) sizing of the generation portfolio for off-grid systems; (3) design of the electrical network; and (4) eventual upstream reinforcements of the electric network and generation portfolio. The current literature offers a wide range of tools and methodologies that can address different aspects of the rural electrification planning. These can be divided into three main classes based on their objectives and scope.

The first class includes tools with the goal of optimising energy systems from the generation side by defining the optimal mix of energy sources to be used to satisfy a certain demand. Generally, they are split into local off-grid sizing, focusing on electrical production, and wide-scale energy analysis. The most well-known commercial tool for local microgrid sizing is HOMER [9]. Other tools that can be found in the literature are Microgrids.py [10], iHOGA [11], DER-CAM [12], and the microgrid optimisation tool developed by Ref. [13]. Other interesting studies that fall within this category are [14,15] which focus not only on the economic evaluation of microgrid generation portfolios, but also the socio-environmental impact and sustainable development. Moreover, a comparison between an off- and on-grid solution is discussed; however, the comparison is made for each community without an integrated optimisation that could capture the overall costs of a grid expansion for a vast area. The goal of these tools is generally to provide a hybrid microgrid solution while focusing on the resources available in each area and addressing rural electrification particularities such as load uncertainty and long-term planning.

The second group of energy models includes frameworks that optimise energy systems from smaller communities to entire nations. They use energy models, energy flows, and balances, and consider various energy vectors, such as heat and electricity, to represent the system under analysis. The most recognized models in this group are Calliope [16], Osemosys [17,18], and Oemof [19]. While these tools are not specifically designed for rural electrification planning, their flexibility allows for easy adaption and application to various case studies. Another well-known tool for country-scale energy analyses is OnSSET [20]. It determines the means of electrification as a function of the distance to the existing grid, and the results for 56 countries are available on the ESMAP platform [21].

The second general class of tools deal mainly with the optimisation of the electrical network, hereon referred to as electrical models. These models are well suited for optimising the electric grid, which is usually the task of the local distribution system operator (DSO), by fitting them into an already established energy plan. Typically, the models represented in this group focus on the siting and sizing of primary and secondary substations and the routing of low-voltage (LV) and medium-voltage (MV) lines. Moreover, the procedures vary depending on the phase of development faced by planners. Note that there are various tools designed for industrialised countries, which may not be applicable to the problem of rural electrification. When focusing on tools that provide greenfield solutions, the main objective is to minimise the overall net present value investment. Therefore, innovative models have been proposed to address specific aspects of distribution planning in rural areas [22]. An interesting tool applicable to rural electrification is the Reference Network Model (RNM), which adopts a greedy approach to design high-, medium-, and low-voltage networks for wide areas [23]. It utilises cost-surfaces and street-tracking features to provide feasible and cost-effective solutions to the distribution grid planning problem.

Finally, the third category is represented by tools aimed at providing a comprehensive electrification plan and simultaneously optimising both the generation and distribution infrastructure. Hereafter, these tools are referred to as comprehensive models. Although traditional energy and electric models have been considered complementary, recently, some studies have merged these two approaches to obtain a more realistic solution. This type of modelling is particularly suitable for rural electrification in low-income countries because it helps in identifying cost-effective electrification approach. Because the cost is a significant factor in electrification planning, it is essential to choose between grid extension and off-grid solutions to achieve least-cost electrification. Reference Electrification Model (REM) is a comprehensive computer model that supports large-scale, as well as local electrification projects [24]. It initially applies a bottom-up greedy procedure for clustering customers and then decides, through an iterative process, which customers to connect to the grid and which to electrify using off-grid systems. In a second step, the electric grid is designed using the tool RNM. More specifically, REM provides the following outputs: (i) the optimal grouping of individual consumers into electrification clusters so that total system costs are minimised (ii) the optimal generation mix and network layout for each of the off-grid mini-grids and selecting the size of the diesel generators, photovoltaic (PV) modules, and battery energy storage system (BESS) (iii) the optimal network layout for each cluster that will be connected to the grid, designed using the RNM previously mentioned.

The GeoSim platform [25] is a modular commercial software tool based on geographical information systems (GIS). Its main innovation consists of the optimisation of energy services covering a given territory, with the goal of improving the economic and social impact of rural electrification. GEOSIM relies on geographical distance data to identify optimal solutions for rural electrification. This includes identifying areas where grid densification should be implemented, determining suitable areas for grid extension, and potential areas for off-grid systems.

REM and GEOSIM represent state-of-the-art in rural electrification planning and have been used as support tools for the redaction of national electrification plans in sub-Saharan Africa [26,27].

Other studies that can be categorised as comprehensive models are the Network Planner tool by Columbia University [28], LAPER [29], and the work presented in Ref. [30], which performs multistage planning considering grid connections, hybrid microgrids, and solar home systems.

When discussing the algorithms adopted, one approach is to formulate the electrification problem as an optimisation problem and solve it using Mixed Integer Linear Programming (MILP). This concept has been widely investigated for optimising off-grid solutions and for general electrical routing purposes. In such cases, it is extremely important to arrange the optimisation efficiently to avoid scaling issues. For example [13], applied a MILP model to optimise the generation portfolio of a microgrid while considering advanced features such as battery degradation. Similarly [31], developed a MILP-based predictive planning and dispatch algorithm for rural microgrids.

The application of MILP formulation in distribution system planning

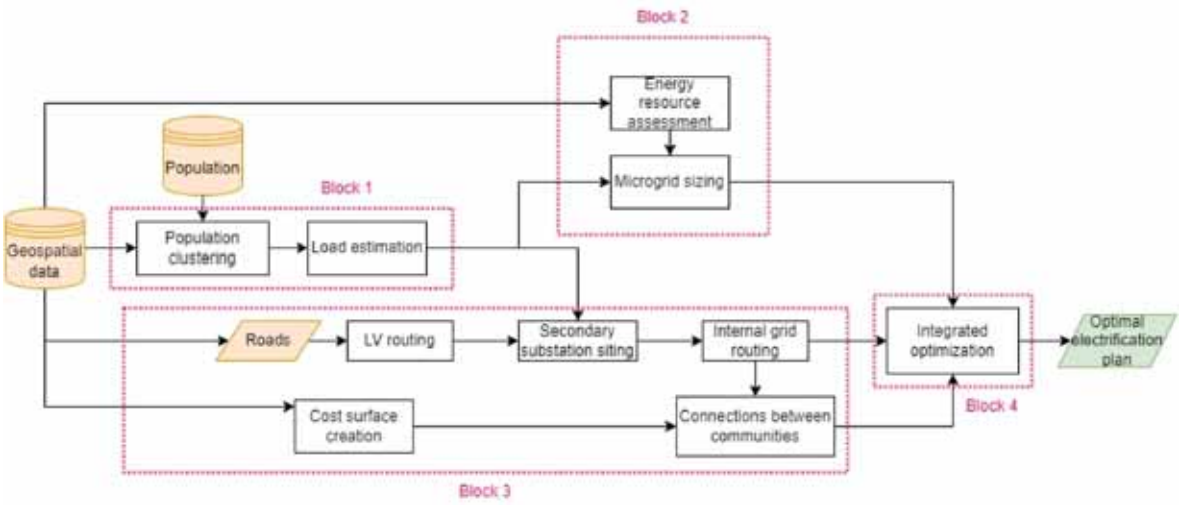

Fig. 1. Flowchart of the complete procedure.

has been extensively studied and documented in the literature for a significant period of time [32]. Moving on to recent times [33], performed a wide bibliographic overview on the optimisation techniques for distribution system planning, describing MILP as "the guarantee of optimality and the current processing capacity of computers, makes approaches based on mixed integer linear programming (MILP) models very attractive". However, the applications mostly come in terms of a very focused analysis for upgrading already developed systems [34], whereas, to the best of the authors' knowledge, in recent literature, it has not been applied for integrated optimisation analysis that would evaluate all the contemporary options. For distribution grid planning, studies have found more success in applying heuristic optimisation methods [35], such as tabu search, various genetic algorithms, and particle swarm optimisation [36]. When selecting an algorithm, there is a trade-off between scalability and accuracy. While MILP has been demonstrated to perform well for large-scale optimisation problems owing to the development of efficient solvers such as Gurobi [37], it is limited by its linear nature, which makes it hard to capture the nonlinear behaviour of power systems. Including nonlinear characteristics such as losses in the optimisation function using piecewise linearisation is possible, but it may lead to low accuracy and significantly increase the computational complexity. Therefore, it is not commonly used in recent literature as a general practice [38]. Some characteristics of the rural electrification problem follow the assumptions made in the MILP modulation. For instance, in these areas, energy consumption is typically low, and this makes it reasonable to neglect losses in the optimisation function as a simplification.

Comprehensive models are usually able to address three of four steps of the rural electrification framework: identification of the optimal electrification solution; microgrid sizing; and design of the electric network. This makes them very promising for addressing the issue of energy access, besides the large amount of data required as input to provide accurate solutions. However, some drawbacks and potential areas for improvement can be highlighted: (i) Greedy approaches are used to design the electric grid, while considering detailed data and assumptions regarding power profiles, user types, and other factors that may be more relevant in developed countries (ii) The optimal electrification solution for the communities is not based on an integrated optimisation, and considering the area as a whole, but rather on iterative approaches considering each community separately or in predefined clusters. (iii) The majority of the tools consider limited options for hybrid microgrids, mainly composed of diesel generators, PV modules and batteries. (iv) Lastly, it was observed that most of these tools are commercial and are not openly available, a condition that could hinder their usability and further development.

The goal of this paper is to propose an open-source, comprehensive analysis for large regions by formulating the electrification planning as an optimisation problem. To reduce the computational burden and enable analysis on vast territories, some case-specific and realistic approximations are made at the level of individual communities. The key contributions of this paper can be summarized as follows: First, the creation of realistic multi-parameter cost surfaces, validating the various cost parameters through a cooperation with a DSO, typically in charge of deploying lines. Second, a procedure to design the electric grid within the communities based on an adaptation of the graph-oriented Lukes algorithm, first described in Ref. [39]. Finally, the formulation of a MILP-based optimisation for the overall electrification plan for the area under consideration, which provides the optimal strategy of electrification for each community, whether with a grid extension or a microgrid solution.

3. Methodology

The procedure developed in this study aims to overcome the limitations found in the literature and provide a holistic methodology that can define a least-cost electrification plan for wide areas. This creates a synergy between various themes that can be observed in the rural electrification analysis paradigm, such as graph theory, GIS data utilisation, and optimisation. This procedure is available as an open-source tool accessible under APACHE 2.0 license [40]. As such, the default solution is provided using publicly available data; however, to provide a more accurate solution, planners can use more specific information, such as population distribution, accurate load demands, and cost estimation related to the area analysed, both in terms of grid extension and detailed costs related to deploying off-grid solutions.

Fig. 1 presents a flowchart of the entire procedure and the data used in each step. It is subdivided into four different blocks, each of which outlines one or more tasks related to rural electrification planning. To provide further detail: (1) energy demand assessment, where the task is to subdivide the population into communities starting from the boundaries of a non-electrified rural area, and determine the electrification status based on information from existing networks or nightlights, if the first is not available. In a second step, the electricity needs of each unelectrified community are estimated; (2) off-grid system sizing, where, given the boundaries and load demand of the communities the optimal sizes and generation portfolios of hybrid microgrids, capable of supplying the load to each community, are identified after estimating the RES potential of the area; (3) internal grid design, in which the LV and

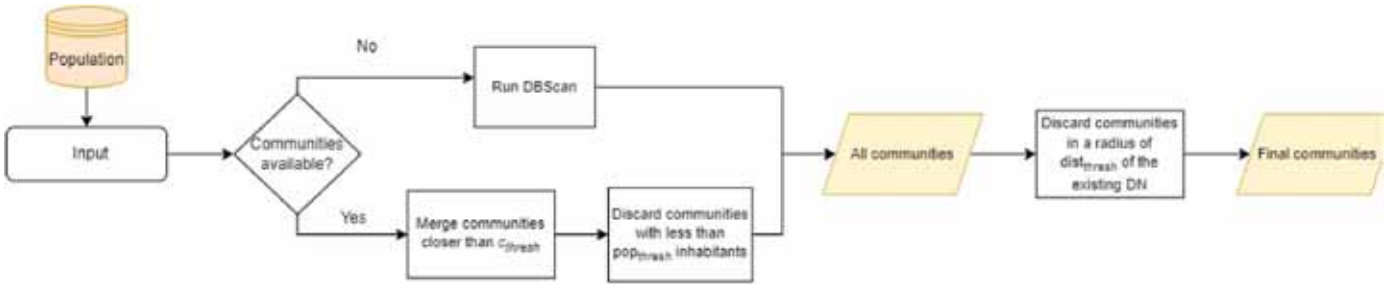

Fig. 2. Identification of communities.

potential MV distribution grids, capable of interconnecting users within communities, are designed. In this step, an electric grid is designed for each community and sites for secondary substations and routes for electric lines are identified. (4) Integrated area optimisation is the last task in this process. Here, all the outputs from the previous steps are gathered to perform a MILP-based optimisation on the entire area, design the grid interconnecting communities to the in-place national grid and identify the communities to be supplied by off-grid systems.

The subdivision into blocks allows the planner to address a problem which is otherwise too complex to define and even more complex to be solved by a one-shot single optimisation algorithm. Based on the block structure illustrated in Fig. 1, it can be inferred that the proposed approach involves simplification in such a way that the electrical grid of each community is either minimally affected or not affected at all by the internal grid structure of nearby communities. This allows the use a graph theory approach to route the electrical grid for each community individually, before performing an integrated optimisation that considers electrical constraints.

The proposed procedure requires input data such as the area to be analysed, geo-referenced information regarding the population, and possible connections to the national grid. The optimal electrification solution for each community was obtained using an electrically constrained MILP procedure, ensuring that the voltages and loading of elements were within the limits. The resulting output included a complete grid expansion project for communities that are best served by national grid expansion as well as a generation portfolio and distribution grid routing for off-grid systems.

The procedure proposed in this study was developed based on the approaches published in Refs. [41,42]. The main contributions and novel aspects of the study are as follows:

- LV grid routing and secondary substation siting can also site the infrastructure of the LV grid (substations and lines) as per the proposed procedure.
- A novel MILP-based integrated optimisation capable of managing electrical constraints is proposed, which is responsible for solving the integrated area optimisation. Previous releases of the tool were employing purely graph theory-based methods, providing a suboptimal solution and unable of managing electrical constraints).
- Improvement of cost surface creation: Penalty factors associated with line deployment have been refined to make estimates closer to the real-life costs faced by DSOs.
- Internalization of the microgrid sizing process: A MILP-based optimisation algorithm for hybrid microgrid sizing was fully integrated into the procedure, avoiding reliance on commercial external software.

3.1. Energy demand assessment

To propose an effective electrification strategy, a fundamental aspect to be addressed is the energy assessment of communities. The starting point is the identification of rural communities; that is, clusters of users grouped and electrified together, either with off-grid systems or via interconnection to the grid. Once the boundaries of the communities were evaluated, it was imperative to differentiate between electrified and unelectrified communities within the analysed area. To ensure a thorough analysis, the energy demand assessment block was divided into two modules, the first being the identification of unelectrified communities within the area of interest, which is crucial when there are already areas with partial access to electricity. The second module involves assessing the electricity demand of each of these unelectrified communities.

3.1.1. Population clustering and electrification status

To identify communities that require electrification, open-source data can provide useful information to characterise the area under study and determine the extent of interconnected energy systems. This study aims to promote the use of such open-source data and provides examples of the type of information that can be obtained, including::

- boundaries of existing villages and communities in the form of a polygon vector layer or similar; in this case, each village can be considered a separate energy system.
- Raster layer with population or building density: The population distribution is an indicator of existing communities and load centres and can be used to extrapolate the boundaries of possible energy systems.

Information on the distribution of the population to be electrified plays a key role in the overall procedure, not only for community identification, but also as a proxy for load estimation. While raster layers with population and building densities are available for the whole world from different sources of data, this is usually not the case for the precise boundaries of communities, which are often not mapped in dispersed and rural areas. In case the boundaries of communities are not well defined and geospatial data are not available, the first step in identifying communities is the clustering of populated points to find densely populated areas. The authors have identified the following population databases as the most promising in terms of promoting open-source analysis and the importance of geo-located information regarding the population: (i) A raster database with the number of buildings in pixels with a resolution of 100 m, available for the entire continent of Africa is made available by (Tatem, 2017). (ii) (Facebook, 2020) provides a raster layer obtained by a neural network locating buildings and census information about the population. The final output is the number of people residing in pixels of 30 × 30 m. (iii) [43] offers an open-source database with polygon shapes of buildings in South America and Africa. However, it is still in the validation process. Moreover, each element in the dataset is represented by a certain percentage of being accurately represented as a building.

The overall procedure for identifying unelectrified communities and the final loads to be electrified is presented in Fig. 2. If the boundaries of communities are available, a pre-screening of their distance and size is

Table 1
Description of main input parameter of RAMP procedure.

Parameter	Description
$User_j$	Name of each user class j
N_j	Number of users within each user class j
$Appliance_{i,j}$	Name of appliance j associated with class of user i
$n_{i,j}$	Number of appliances i within class j
$P_{i,j}$	Nominal power absorbed by appliance i of user j
$ft_{i,j}$	Functioning time: total time appliance i of user j is on during the day
$fc_{i,j}$	Functioning cycle: minimum time appliance i of user j is on after switch-on
$fw_{i,j}$	Functioning window: periods when each appliance i of user j can be on
$Rfc_{i,j}$	Random variation of functioning cycle
$Rfw_{i,j}$	Random variation of functioning window

performed to merge communities which are sufficiently close to be considered a unique energy system and to exclude those that are smaller than a certain threshold (pop_{thresh}). Finally, if the location of the existing distribution network is available, all communities that are within a certain distance from the grid ($dist_{tresh}$) are assumed to have been electrified.

Among the different classes of clustering algorithms available in literature, density-based clustering has been found to be the most suitable for identifying the extent of existing communities [30] for the following reasons: a) it allows the creation of clusters with nonconvex shapes that are closer to the real aspect of communities; b) it considers the presence of outliers, which are points that do not belong to any cluster; c) it is suitable for large datasets; and d) previous knowledge of the exact number of communities is not required.

Within this class, DBSCAN has been identified as the most suitable algorithm given its adaptability to large datasets, limited memory requirements and computational complexity, scalability to different problems, relatively simple parameter tuning, which allows for control of the required population density, and result interpretation. This approach was proposed and described in the first version of the procedure [42]. Different approaches for population clustering are found in the literature: REM utilises a bottom-up greedy approach to consecutively add users to a cluster while it is economically feasible to do so, while in the approach adopted by Ref. [44], users closer than a certain threshold for separation are merged into clusters. Finally, the authors of [45] divided the area under analysis into pixels, which, however, did not properly represent the real shape of communities and would not work for sparse areas.

3.1.2. Load estimation

The goal of this step is to estimate the energy and peak power requirements of each community through the creation of a yearly load profile using a bottom-up approach. Given the low availability of historical trends, utilised to infer possible load profiles for nonelectrified communities, this type of approach is preferred over top-down procedures. The load profiles were created using RAMP, a Python-based open-source tool developed at Politecnico di Milano [46]. It creates synthetic load profiles starting from classes of users, appliances, and usage habits. The main inputs required by the tool are described in Table 1, where classes of users are indicated by index j and classes of appliances by index *i*.

The load profile of each user of a specific user class is provided by the combination of the usage pattern of each appliance *ij*, computed by defining, in a stochastic manner, the time $ft_{i,j}$ that appliance j associated with user i is switched on within the day. These times must be selected from within the identified functioning windows $fw_{i,j}$. Once the appliance is on, it must remain on for at least the predefined functioning cycle ($fc_{i,j}$). The overall daily load profile results from the aggregation of the user class profiles j. Each profile j will be different from that of the others of the same user class because of the stochastic variability provided by the input parameters and the randomised selection of appliance switching-on times.

The required data for the procedure could be obtained through on-site data collection methods, such as conducting surveys among the population of the communities. Alternatively, assumptions from literature could be used if time and budget constraints prevent a local assessment. To facilitate the use of the tool by various stakeholders, the authors created a pool of standard load profiles using a combination of literature, data from the Global Survey on Energy Access program launched by the World Bank, and actual measurements. The World Bank survey campaign on energy access, launched with the support of the ESMAP, aims to collect field data following a Multi-Tier Framework (MTF). These data are now available for eight countries (Ethiopia, Kenya, Liberia, Malawi, Niger, Nigeria, Rwanda, and Zambia) and, to the authors' knowledge, are the most complete sources of data related to electricity consumption at the household and community levels [47].

The created standard profiles were related to different categories of users belonging to the three main classes of customers typically found in communities: households, business activities, and public and social facilities. More specifically, standard profiles for households belonging to each energy access tier (according to the MTF) and profiles for schools, worship, and health centres were created considering the type and number of electric appliances identified by the World Bank in their questionnaires. Typical profiles for business activities and the time of use of appliances were derived from local surveys and literature data [48].

3.2. Off-grid system sizing

This stage of the procedure is devoted to estimating the availability of renewable energy sources in the area under study and optimising the generation portfolio of RES-based hybrid microgrids which could ideally supply each community.

The availability of solar and wind resource is estimated using data from open-source global atlases [49], which, by means of satellite data, reanalysis techniques, and local measurements, provide the annual energy output of standard types of PV modules and wind turbines. The model adopted for microgrid sizing was inspired by the approach proposed by Refs. [50,51]. The choice of model allows the design of hybrid microgrids with multiple energy sources, including diesel generators (DG), photovoltaic (PV), wind turbines (WT), and battery energy storage systems (BESS). Looping through each community, the model receives the data related to the annual load demand, specific power production from renewables, and costs and technical specifications of the components as inputs and outputs the sizes of the generation units that can minimise the total net present cost (NPC) over the microgrid lifetime, as well as their hourly dispatch.

The objective function is the minimisation of the NPC:

$$\min NPC = \sum_i (IC_i + O\&M_i + RC_i - SV_i) \quad (1)$$

where index i represents the different types of generators, IC_i is the investment cost, $O\&M_i$ is the discounted annual operation and maintenance (O&M) costs RC_i is the replacement cost of each component, and SV_i is the salvage value, that is, the residual value of the components at the end of the project lifetime.

The constraints are related to energy balances performed hourly over a set of typical days, that can be represented in a simplified manner for the entire year. At each hour, the microgrid should meet the load demand with a certain percentage of loss of load, which is represented by the unmet demand D_h^u. The total amount of energy supply is given by the sum of energy fluxes from renewables (P_h^{ren}), the ones from the diesel generator (P_h^{ren}), $(\sum_g P_{h,g}^{dg})$, where g are the different types of generators that could be installed, plus the amount of energy discharged from the BESS multiplied by its efficiency $(P_{h,b}^{dch} \cdot \eta_b)$, minus the amount of energy

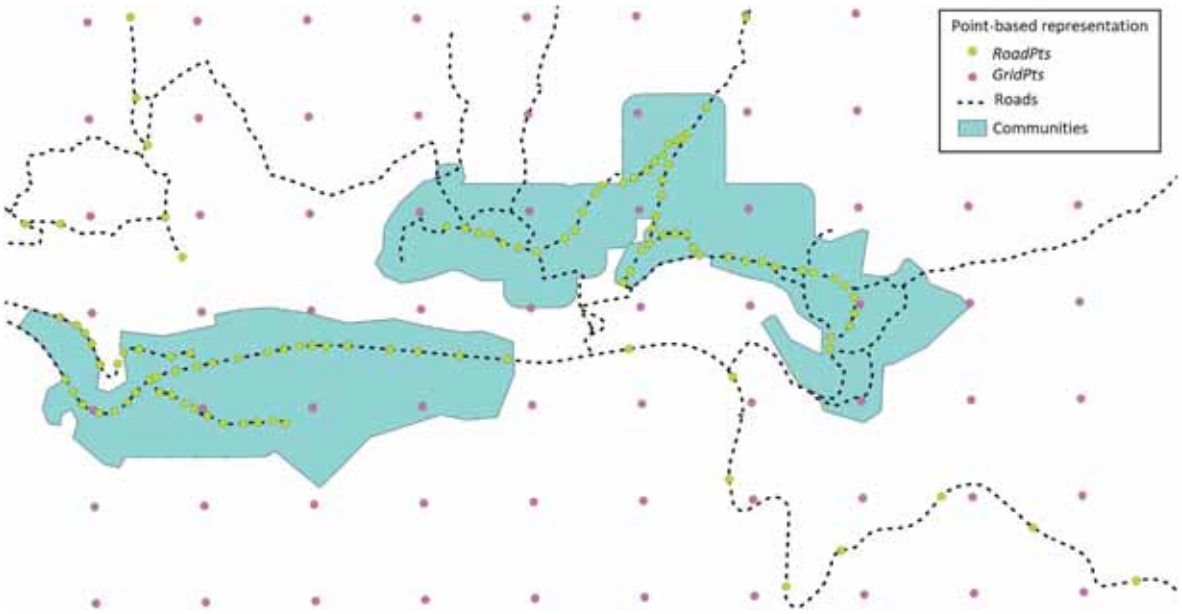

Fig. 3. Terrain representation with a grid of points and roads.

necessary to charge the BESS divided by the efficiency $\left(\frac{P^{ch}_{h,b}}{\eta_b}\right)$.

$$\sum_b \left(P^{dch}_{h,b}\cdot\eta_b - \frac{P^{ch}_{h,b}}{\eta_b}\right) + P^{ren}_h + \sum_g P^{dg}_{h,g} + \sum_{ht} P^{ht}_{h,ht} + D^u_h = D_h \quad (2)$$

The amount of energy not supplied is limited in input by setting a threshold. The minimum renewable energy penetration is also provided as an input. The amount of energy provided by each power source is limited by the installed capacity, and in the case of RES, by the availability of primary resources. The installable capacities of PV, WT, and BESS are continuous variables, while the capacity of the diesel generator is an integer variable, a multiple of a predefined unit size.

The BESS charge and discharge process is modelled as follows:

$$Q_{h,b} = Q_{h-1,b} + \left(P^{ch}_{h,b} - P^{dch}_{h,b}\right)\cdot\Delta h \quad (3)$$

The energy accumulated at hour h, ($Q_{h,b}$) is given by the sum of the energy at the previous hour, the charging energy minus the discharging energy. To avoid simultaneous charging and discharging of BESS a binary variable ($w^{dch}_{h,b}$) is used as follows:

$$P^{dch}_{h,b} \le w^{dch}_{h,b}\cdot M \quad (4)$$

$$P^{ch}_{h,b} \le \left(1 - w^{dch}_{h,b}\right)\cdot M \quad (5)$$

M is a large number, and $w^{dch}_{h,b}$ is 1 when the battery is discharging and 0 otherwise. There are additional constraints that need to be considered, such as reserve requirements to handle load and renewable energy source fluctuations, which are addressed by utilising diesel and BESS. In addition, the diesel fuel consumption is modelled as a linear function of energy production.

A complete formulation of the MILP, which contains detailed information regarding the variables, parameters, and constraints, can be found in Ref. [52]. The NPC for electrifying each community with a hybrid microgrid was then used as the input to the final block of the procedure for integrated area optimisation.

3.3. Cost surface creation

Referring back to Fig. 1, block 3 of the proposed procedure is devoted to routing the electric lines connecting users within the communities, and then creating possible least-cost links among the various communities. Routing relies on the creation of a geospatial-based cost surface that represents the cost of installing new electric lines. This depends on the characteristics of the terrain on which they are built. New medium-voltage lines should preferably be built along roads as it would provide easier access and potentially reduce costs associated with acquiring private property rights.

The cost of line deployment, which include labour and equipment installation costs, as well as operation and maintenance costs, depends on the accessibility of the area, which increases with the distance from the roads and in the presence of difficult terrain such as high slopes and forests. Finally, there are areas in which it is nearly impossible or even forbidden to install lines, such as lakes, large rivers, national parks, and protected areas. This aspect was identified as one of the limitations in some of the current literature, usually considering only forbidden areas without any reference to land cover, slope, and roads.

The novel approach proposed in this paper is designed to include all these aspects of electric grid routing by creating a cost surface, where a different deployment cost is associated with each type of terrain [24,53]. This is achieved by combining the raster and vector models (the two main approaches available in the literature to model cost surfaces) [54] representing different aspects of the physical terrain through a grid of points, obtaining a good compromise between accuracy and simplicity.

More specifically, the final cost surface provided as input to the MV grid routing procedures is given by a point vector layer composed of (1) a regular grid of points (*GridPts*) with a resolution *Res*, which is equivalent to a raster surface, because each point represents the centroid of a pixel of a raster layer, with a penalty factor (pf_{point}) associated with each point; and (2) additional points sampled along the roads, with an associated unitary cost (*RoadPt*). The utility of tracking the roads during the planning phase of distribution feeders has already been recognized in the state-of-art. For example, RNM adopts a similar approach for the creation of synthetic electric grid models [55,56]. An example of the spatial representation is presented on Fig. 3.

The penalty factor *pf* associated to each *GridPts* is the result of the combination of different characteristics gathered by sampling information from different geospatial layers (also referred to as *geospatial factors*).

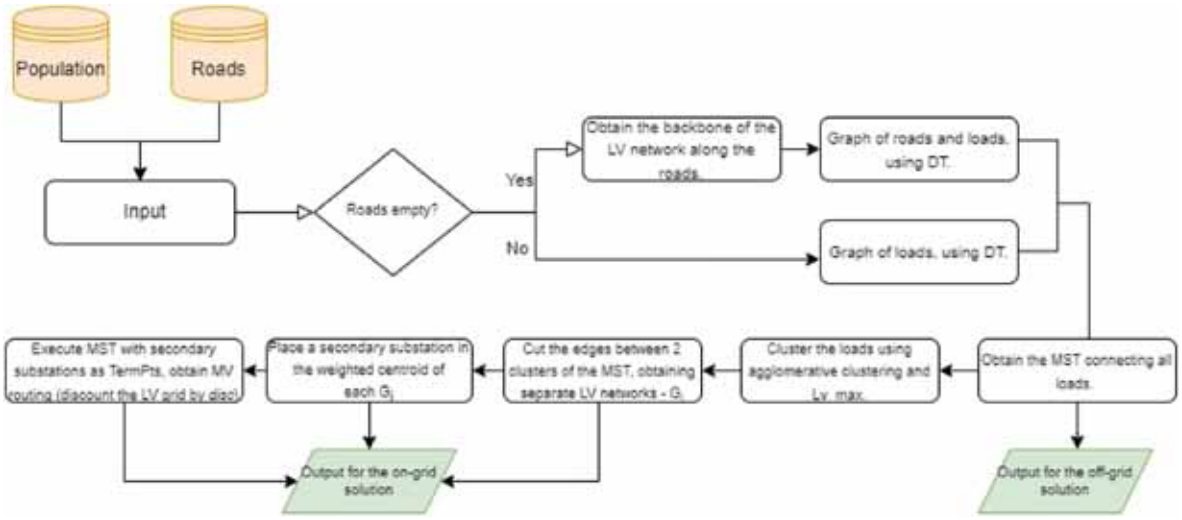

Fig. 4. Flowchart of the internal grid routing procedure.

- Distance from roads: The weight increases linearly up to a certain threshold when it reaches its maximum, and the distance is computed as the Euclidean distance of each point from a road vector layer.
- Slope: The weight increases exponentially with the slope, which is computed from the Digital Elevation Model (DEM) raster layer and sampled at each point.
- Protected Areas: To prevent the deployment of lines in forbidden areas, that are typically found in the form of polygons, the weight assigned to points within them is severely penalized.
- Water bodies: Weight is much higher than that of others to avoid crossing lakes or large rivers.
- Land Cover: Weights differ according to the type of land cover (e.g. dense forests have a higher weight than agricultural land); land cover values at each point were sampled from land cover raster layers.

The inclusion of additional *RoadPts*, which are not regular *GridPts,* allows for improvement in accuracy and refinement of the procedure. In fact, deploying equipment on roads or in close proximity has clear benefits in terms of terrain accessibility and state ownership. The sampling resolution of roads was based on whether they were within or outside the community. This concept is clearly shown in Fig. 3, wherein the roads are sampled with a higher resolution that corresponds to the distance between LV poles to promote a more detailed grid structure within the communities; values of resolution between 40 and 60 m are acceptable. Roads outside the communities were sampled with a resolution equal to *Res*.

The approach proposed for estimating the costs of line deployment splits the overall CapEx into five categories: conductor, poles, additional material, labour, and permissions. Each of the geospatial factors has a different impact on these five categories, and each category $i \in categories$ is represented by a certain percentage of the overall base cost of line deployment (inv_i). Considering that the points that represent roads have a value equal to 1, the weight assigned to each point is calculated as in (6).

$$pf_{point} = \begin{cases} \sum_{i \in cat} \left(inv_i \bullet \left(1 + \sum_{j \in factors} w_{i,j} \right) \right) \forall \text{point} \in GridPts \\ 1 \forall \text{point} \in RoadPts \end{cases} \quad (6)$$

Weight $w_{i,j}$ is the input to the procedure given in the form of a lookup table, which can be observed in ANNEXURE A. It is advisable that for better estimation of the real costs for deployment of lines, this lookup table be created specifically for the case study that is under analysis, because it considers specific parameters that can vary from country to country.

Similar approaches have been proposed in the literature; however, rather than splitting the costs into different categories, the weight is multiplied with the entire cost of the deployment of the line, [42,57]. Thich could lead to inaccuracies because each cost category is affected differently by the type of terrain. For example, costs related to man-hours of work and fuel transportation is significantly reduced in the presence of roads, while the cost for electrical lines could increase drastically when deploying underwater cables.

Finally, to consider the operational expenditures or costs related to O&M, the total NPC for the estimated lifetime of the equipment ($\overline{Y}$) can be calculated as:

$$pf_{NPC,point} = pf_{point} + \sum_{y=1}^{\overline{Y}} \frac{\left(\text{O\&M} \bullet \left(1 + \sum_{j \in factors} w_{O\&M,j} \right) \right)}{(1+r)^2} \quad (7)$$

where r is the average discount factor of the project, and $w_{O\&M,j}$ is the weight associated with the increase in difficulty of performing O&M on lines, depending on the spatial characteristic factors.

The weights assigned to each point directly affect the overall deployment and maintenance costs of the built lines. Finally, a graph $G_{area} = (V_a, E_a, w_a)$ of the entire area is created, where the vertices are composed of all the points representing the terrain, $V_a \in (GridPts \cup RoadPts)$, and the edges connecting the vertices are only the neighbouring nodes based on *Res,* following the logic in (8). Finally, (9) depicts the weight associated with connection (*i-j*) as a multiplication factor to deploy a connection between points *i* and *j*.

$$E_{i-j} = \begin{cases} 1, \forall (\text{i} - \text{j}) \in \text{dist(i, j)} < \sqrt{2} * Res \\ 0, otherwise \end{cases} \quad (8)$$

$$W_{i-j} = L_{i-j} * \frac{pf_{NPC,i} * pf_{NPC,j}}{2} \quad (9)$$

It is assumed that, within communities, the effect of the terrain is negligible for routing the LV lines; thus, the weighted graph G_{area} is used exclusively for MV grid routing.

3.4. Internal grid routing

The goal of this stage is the routing of the internal distribution grid within each community. Two different solutions are obtained as outputs, depending on whether the community is electrified with an off-grid solution or via grid extension. For the former, the presumption is that

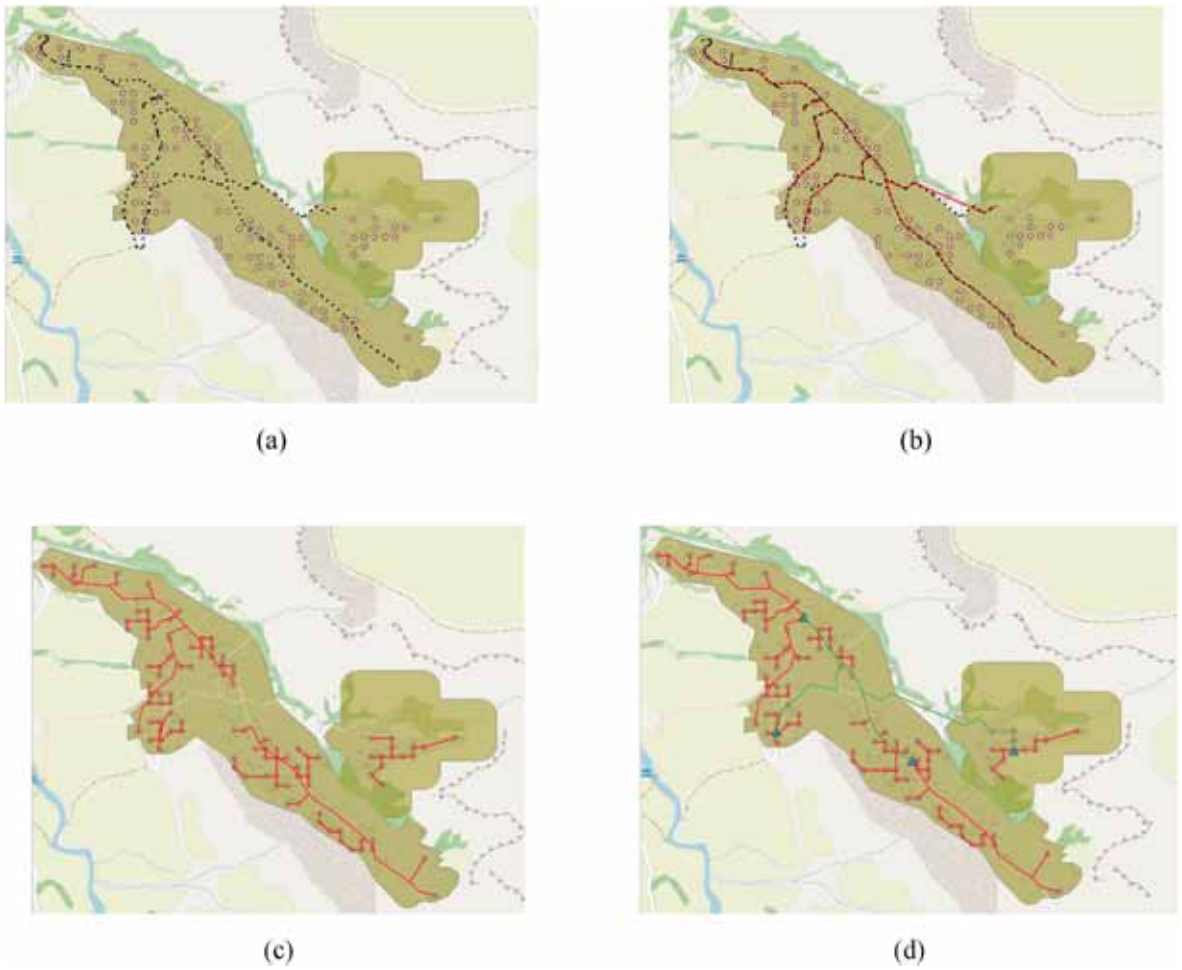

Fig. 5. Example of internal grid routingon a single community. Pink dots are the final users, black dashed lines are the roads, red lines are the LV grid, and the blue triangles are the secondary substations.

the microgrid will be deployed at the LV level; the hypothesis is supported by the limited size of the communities as well as the low energy consumption in the areas analysed. These characteristics are commonly related to areas proposed for off-grid solutions.

However, for grid extension, the output involves routing of the LV and MV grids and the siting of the secondary substations, that is the MV/LV interfaces. Looking back to the literature, the procedure for obtaining the locations of secondary substations is an improved version of Luke's algorithm [39]. The analysis performed in this block was purely based on graph theory and least-cost equipment deployment. The complete flowchart of this block, with the procedure performed for each community separately, is shown in Fig. 4. The inputs to the process are as follows: (1) the electrical load defined within Section 3.1.2 and (2) the grid of points defined in Section 3.3.

Algorithm 1

Input: $\boldsymbol{Loads}$, $\boldsymbol{RoadPts}$

Output: MST connecting all the loads.

For *com* in *communities*:

If *RoadPts* within *com*:

Create a graph from the roads: $\boldsymbol{G_{com} = Graph(V, E, w)}$; **where** $\boldsymbol{V \in RoadPts; w_{i-j} = distance_{i-j}}$

Translate each load to the closest road point: $\forall\ \boldsymbol{load \in Loads, V_j(pop) = 1 \leftrightarrow j\ is\ closest\ point\ to\ load}$

Find the MST connecting road points that are close to loads: Backbone = MST ($\boldsymbol{G_{com}}$, V(pop)=1]

Decrease the weight of the backbone in the new graph: $\mathbf{w(E) = 0, \forall\ E \in Backbone}$

Else:

Create an empty graph: $\boldsymbol{G_{com}}$ =Graph(empty)

Add loads to $\boldsymbol{G_{com}}$: $\boldsymbol{G'_{com} = Graph(V', E', w'), V' = RoadPts \cup Loads;\ E' = Delaunay(V');}$

MST connecting all the electric loads: $\boldsymbol{Tree_{com}}$ = MST ($\boldsymbol{G'_{com}}$, $\boldsymbol{Loads}$)

Algorithm 2

Input: $\mathbf{Tree_{com}}$, LV_{max}, Loads

Output: Substation siting and low voltage networks.

For *com* in *communities*:

Create a distance matrix based on $\boldsymbol{Tree_{com}}$: $\boldsymbol{dist_{i-j} = w_{i-j}(Tree_{com}),\ i,j \in Loads}$

Split the loads in clusters based on the distance threshold: cluster = Agglomerative (Loads, $\boldsymbol{LV\_max}$)

Drop each connection between 2 separate clusters: $\boldsymbol{V_{i-j} = 0 \leftrightarrow i_{cluster} \neq j_{cluster}}$, $\forall\ \mathbf{(i,j)} \in \mathbf{V(Tree_{com})}$

Obtain each separate LV network: $\boldsymbol{LV_{clus} = Tree_{com}.subtree(clus)}$, $\forall\ \mathbf{clus} \in \mathbf{clusters}$

Obtain the secondary substations ($\boldsymbol{substation_{clus}}$) in the centroid of $\boldsymbol{LV_{clus}}$

Fig. 5 illustrates an example of the entire procedure for one community with the input data presented in Fig. 5a. The proposed procedure is a bottom-up non-greedy approach in which the first step is to connect all population points in the least-cost manner. Electrical loads are represented by the associated specific load per capita. The first part, detailed in Algorithm 1, obtains the minimum spanning tree of the loads in the community. Creating a "backbone" for the LV network that tracks the roads is necessary for obtaining a realistic solution. Thus far, the procedure has relied on the assumption that roads play a crucial role, as electrical lines and substations need to be installed in close proximity to roads for ease of accessibility and deployment, especially within inhabited areas. Not considering these factors could lead to an irrational grid topology.

This is done by creating a graph $G_{com} = Graph(V,E,w)$, where $\forall V \in RoadPts$. Each populated point is translated to the closest road and the Minimum Spanning Tree (MST), which is correlated with the least-cost line deployment, is obtained by adopting the Steiner algorithm (example in Fig. 5b). Certainly, some communities in rural areas do not have roads. For these communities, the step of creating the backbone of the network along the roads was omitted.

Subsequently, the complete MST of the electrical load is obtained. This is performed with the aid of the Delaunay Triangulation (DT). The DT is executed using the population and road points. Then, each edge of the Delaunay triangle is added to graph G; the weight is equal to the length, ensuring that in the final topology, the electrical lines do not cross each other. The final graph G_{com} is used to find the overall MST of the community is obtained by discounting each edge in G_{com}, which is part of the backbone, using a discounting rate (*disc*). The discount rate must be close to 100%.

In the case of a community without roads, DT is executed using only the final loads, and the connections are added as edges to an empty graph. Finally, the MST of the community is obtained by executing the Steiner tree algorithm for graph G_{com}, using the load points as terminal nodes (example presented on Fig. 5c). Based on the assumption that microgrids are designed as LV systems, the MST represents LV grids in the case of an off-grid solution.

However, in the case of on-grid solution, it is important to note that energy distribution is performed at the MV level, which requires siting of secondary substations and routing of the potential MV network within the community. This procedure is described in Algorithm 2. Starting from the MST obtained in the previous step, the populated points are clustered using an agglomerative clustering algorithm. The main goal is to ensure that the distance between any two nodes belonging to the same cluster is not larger than a certain value, considering the distances between loads on MST - $Tree_{com}$, which was first introduced in Ref. [39]. In this application, this value is represented by the maximum allowable length of the LV network (LV_{max}), which is an index often used by DSOs when deploying LV networks. Finally, returning to the MST of the community, each edge-connecting point that belongs to a different cluster is cut to obtain separate LV networks or separate subtrees in the graph domain. The secondary substation for each LV network is placed at its weighted centroid. The peak load of the LV network is calculated based on the sum of all loads by applying a logarithmic coincidence factor ($coinc_{LV}$). Based on the peak load of the LV network, a secondary substation is selected from a list of commercially available transformers.

The final step is the routing of the MV network within the community, which is determined as the MST of G'_{area}, which is an updated version of the G_{area} defined in Section 3.3, in which the secondary substations are added as terminal nodes, with connections to the closest point on G_{area}. Moreover, an additional feature is inserted that incentivises pole sharing between the MV and LV conductors. Mainly, the edges of the previously obtained LV network are discounted by a factor of "c_{pl}". The final internal structure is shown in Fig. 5d.

Table 2
Sets of the integrated area MILP formulation.

N_c	Communities
N_s	Secondary substations
N_{ip}	Intersection points of the communities' internal grid
N_{lp}	*Load points* ($N_s \cup N_{ip}$)
N_{cp}	Points of connection to the existing grid
A_d	Decision links
A_c	Communities' internal links
A	*All links* ($A_d \cup A_c$)

3.5. Integrated area optimisation

The objective of this block of the procedure is to perform an integrated optimisation to minimise costs, considering not only the design of the electrical grid but also the means of electrification of each community. Consequently, each community can be connected to the national grid or electrified via an off-grid system.

The optimisation problem is formulated as a MILP model with the objective function of minimising the overall NPC costs for the expansion project, considering the electrical equipment, cost of microgrid components, and cost of electricity distributed through the national network:

$$\min \sum_{(i,j)\in Ad} NPC_{Ad,ij}x_{ij} + \sum_{c\in Nc} NPC_{mg,c}y_c + \sum_{s\in Ncp} NPC_{cp,s}z_s + \sum_{c\in C}(1-y_c)D_{e,c}coe + \sum_{c\in C}(1-y_c)NPC_{g,c} \tag{10}$$

To make grid extension and off-grid electrification as two

Table 3
Scalars of the integrated area MILP formulation.

coe	Wholesale cost of electricity	[€/kWh]
Vref	Voltage level for expansion	[kV]
r	Per length line resistance	[ohm/km]
x	Per length line reactance	[ohm/km]
Pa	Maximum power	[MVA]
Emin	Minimum acceptable voltage	[p.u.]
PF	Loads' power factor	[p.u.]

Table 4
Parameters of the integrated area MILP formulation.

D_c	Energy demand of community *c* during microgrid's lifetime	$c \in N_c$	MWh
P_{lp}	Peak demand of load points *p* adjusted for MV coincidence	$p \in N_{lp}$	[MVA]
P_c	Peak demand of community c	$c \in N_c$	[MVA]
$NPCmg_c$	Net present cost of microgrid for community c	$c \in N_c$	[$]
$NPCg_c$	Net present cost for the MV grid elements community c	$c \in N_c$	[$]
$NPCad_{(i,j)}$	Net present cost for decision line (i,j)	$(i, j) \in Ad$	[$]
NPC_{cp}	Net present cost of connecting to s	$s \in N_{cp}$	[$]
$L_{i,j}$	Length of connection (i,j)	$(i,j) \in A$	[km]
$W_{i,j}$	Weight of decision links (i,j)	$(i, j) \in Ad$	Const
E_s	Per unit voltage of connection point s	$s \in N_{cp}$	p.u.
P_s	Maximum power of connection point s	$s \in N_{cp}$	[MVA]

Table 5
Variables of the integrated area MILP formulation.

$x_{i,j}$	1 if connection (I,j) is built; otherwise 0	$(i,j) \in Ad$	[0,1]
y_c	1 if community c is a microgrid; otherwise 0	$c \in N_c$	[0,1]
z_s	1 if substation s is used/built; otherwise 0	$s \in N_{cp}$	[0,1]
P_a	Power flow of connection *(i,j)*	$(i,j) \in A$	[MVA]
P_s	Power provided by connection point *s*	$s \in N_{cp}$	[MVA]
E_i	Linearized voltage at node *i* $i \in$	$(N_{lp} \cup N_{cp})$	[p.u.]

comparable solutions, a salvage value was considered for computing the NPC electric lines. The salvage value is computed assuming a linear depreciation of components, a lifetime of 40 years for electric lines and a project lifetime of 10 years. The following is the complete formulation of the optimisation problem, where the parameters represent values that are correlated to a certain set, whereas the scalars are single values used in the optimisation. A detailed formulation of MILP in terms of sets, scalars, parameters, and variables is presented in Tables 2–5.

The optimisation problem is constrained by the maximum loading of the lines and substations, as well as a linearized nodal voltage formulation. The voltage drop in lines *i-j* is computed using the well-known linear approximation [52] over a radial network presented in (11):

$$\Delta E = \frac{R \cdot P + X \cdot Q}{En} \tag{11}$$

where R and X are the line resistance and reactance, respectively; P and Q are the active and reactive power flows on the lines, respectively; and *En* is the nominal voltage.

Similar to the approach for the internal grid presented in Section 3.5, the optimisation formulation is represented as a weighted graph with nodes and lines. The load nodes (N_{lp}) are represented by the secondary substations (N_s) and intersection points of the internal MV networks obtained using the procedure detailed in 3.4 (N_{ip}), as well as the possible points of connection to the existing grid (N_{cp}). The weight of each load node is represented by its power consumption (P_c), which for secondary substations is the value calculated in the previous block and adjusted with an MV coincidence factor. The weight of the intersection points is zero, as shown in (12).

$$P_{lp,n} = \begin{cases} Pn * MV\_coin, n \in Ns \\ 0, n \in Nip \end{cases} \tag{12}$$

The inclusion of intersection points increases the computational burden of optimisation; however, it is necessary to obtain a complete representation of the MV grid within the communities. Moreover, fictional points (N_c) were included in the MILP model to represent the communities and correlate them with the respective off-grid options. In this case, the weight of $P_{c,i}$ is the sum of the loads of the points within community *i*, allowing proper convergence of the load flow (13).

$$P_{c,i} = \sum_{j \in i} P_{lp,j} \ (i \in N_c) \tag{13}$$

This procedure is viable for both green and brownfield expansion projects. As such, the points of connection to the grid are generalised, that is, they can represent not only primary substations, but also points belonging to the existing MV network. As such, NPC_{cp} can assume different values depending on whether connection point *s* is an existing substation, a proposed substation, or simply a connection point in the existing MV grid.

Following the procedure explained in Section 3.5, it should be specified that the internal structure of the MV network is known and is considered as a parameter in the optimisation procedure, following the assumption that the structure of the MV network inside one cluster is not affected by the overall expansion. This assumption allows for the analysis of a detailed expansion model of the grid while also considering the option for off-grid solutions. For this reason, the set of all lines in graph (A) is split between decision lines (A_d), which are the proposed lines connecting communities and points of connection to the existing grid, and internal lines (A_c), which represent the internal structure of the communities. This bifurcation is important because the binary variable $x_{i,j}$ is assigned only to the decision lines. Moreover, the optimisation formulation allows the inclusion of two different types of lines to allow the model to build a realistic grid topology with a larger conductor on the main branch close to the point of connection. In this case, two different binary variables are required for the decision lines and one binary variable for the internal lines.

The last input required for optimisation is the set of decision lines (A_d), as other inputs have already been calculated in the previous blocks of the procedure. These decision lines represent the least-cost paths between two communities or between a community and a connection point. The cost and length of the links were calculated by adopting the cost-surface representation of the area described in Section 3.3. Considering the typical structure of electrical grids, it is not necessary to include as decision variables all possible connections between communities, because it is certain that the lines that are built will not jump over communities. Moreover, we can safely assume that if a MV line passes within or near a community, the optimal means of electrification will be grid extension (such an assumption could be different for high-voltage lines). Therefore, the optimisation procedure is simplified by drastically reducing the number of possible connections by adopting DT. The procedure for identifying the decision links (A_d) to be used as inputs to the optimisation process is detailed in Algorithm 3.

Algorithm 3

Input: Communities ($\boldsymbol{N_c}$), Secondary substations($\boldsymbol{N_s}$), Connection points ($\boldsymbol{N_{cp}}$)

Output: Decision links ($\boldsymbol{A_d}$)

Step 1: Perform DT on $\mathbf{N_s}$, obtaining acceptable links between $\boldsymbol{N_c}$ and $\boldsymbol{N_{cp}}$ -> $\boldsymbol{C_{d(i-j)}}$ $\boldsymbol{i} \in \boldsymbol{N_s}, \boldsymbol{j} \in (\boldsymbol{N_s} \cup \boldsymbol{N_{cp}})$

Step 2: For option in $\boldsymbol{Cd_{i,j}}$:

$\boldsymbol{A_{d(i-j)}}$=min (Dijkstra(n,m) $\forall \boldsymbol{n} \in \boldsymbol{i}, \boldsymbol{m} \in \boldsymbol{j}$) ->Closest links between communities considering substations

$\boldsymbol{L_{d(i-j)}}$= Length ($\boldsymbol{A_{d(i-j)}}$) ; $\boldsymbol{W_{d(i-j)}}$ = Cost ($\boldsymbol{A_{d(i-j)}}$)

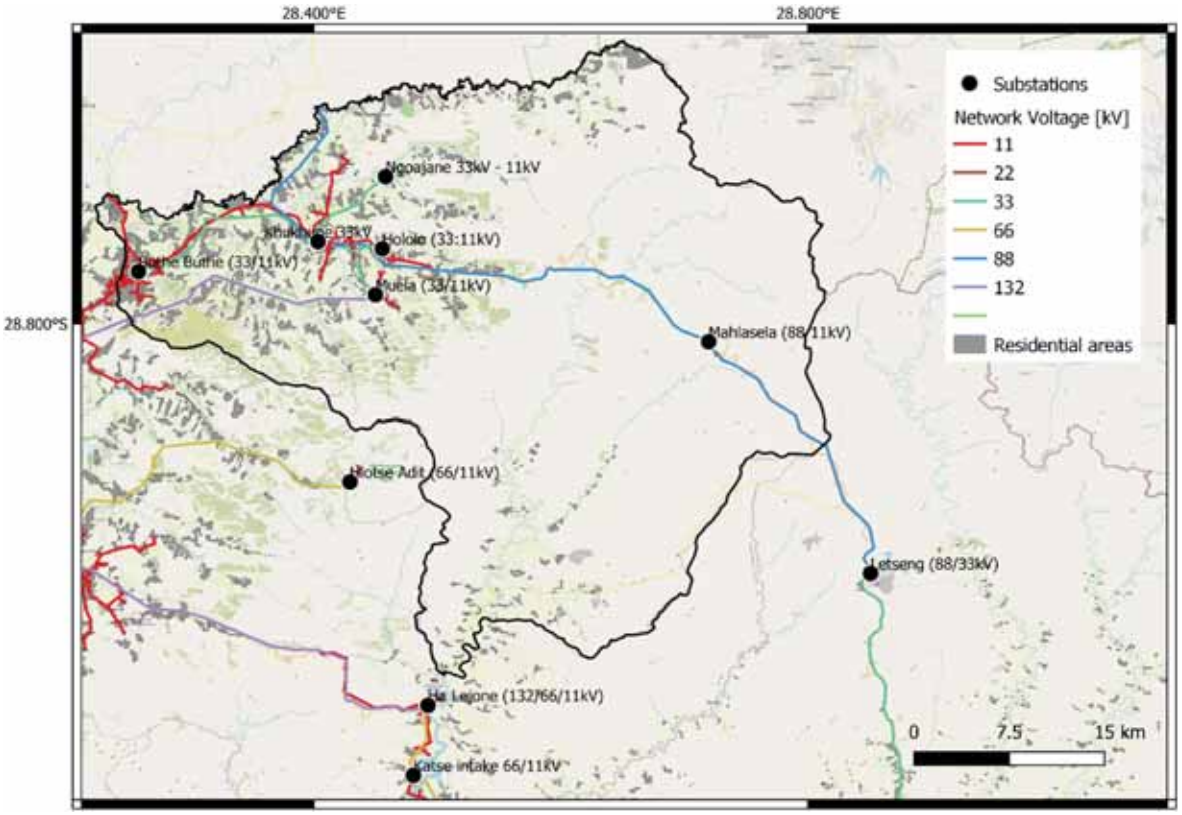

Fig. 6. Input information for case study.

The output of this procedure is a set of decision lines (A_d), together with the associated lengths (L_d) and costs (W_d). Finally, the complete parameters L and W are obtained, representing the lengths and costs of both the decision and internal lines.

Because of the low-consumption profiles, it is considered that the off-grid solutions could be designed as LV microgrids. Thus, secondary substations and internal MV grids can only be built in communities where electrification is considered an extension of national grids. For this purpose, another parameter ($NPCg_c$) was included, which represents the net present cost of the secondary substations and the MV network within community c.

Constraining the optimisation with limits for thermal loading and maximum voltage drop is conditioned by the calculation of these parameters. First, the power on each line i-j was calculated using the formulation of the First Kirchoff's Law. Because the power supplied by the points of connection is variable, two separate equations are required, as shown in (14) and (15).

$$\sum_{j:(j,c)\in A} P_{A,cj} = P_c \; ; \forall c \in Ncp \tag{14}$$

$$\sum_{j:(j,l)\in A} P_{A,jl} - \sum_{j:(j,l)\in A} P_{A,lj} = P_l \; ; \forall l \in Nip, l \neq j \tag{15}$$

in (14), the direction of the power is known, because it is clear that the power is supplied by the connection point, which is equal to the sum of the powers on the lines connected to it. However, for the remaining lines (15), the direction of the power flow is unknown. Therefore, it is important to include both the lines entering and exiting the node. Moreover, it is important to clarify that a power-flow equation for fictive microgrid nodes would be redundant, as that line is already implicitly included in (15). Further, it is crucial to constrain the maximum power of the lines according to the respective electrical parameters of the specific conductor, while also constraining lines that are not built to have a power equal to 0. To properly formulate this constraint, differentiation is required between the decision and internal links, as detailed in (16) and (17).

$$-\overline{P_A} x_{ij} \leq P_{ij} \leq \overline{P_A} x_{ij} \forall (\mathrm{i,j}) \in A_d \tag{16}$$

$$-\overline{P_A} \leq P_{ij} \leq \overline{P_A} \; ; \forall (\mathrm{i,j}) \in A_c \tag{17}$$

The maximum loading constraint is also defined for the maximum power that can be supplied by the connection points, while fixing the power at 0 if the connection point is not used (18).

$$P_s \leq \overline{P_s} Z_s \; \mathrm{s} \in \mathrm{Npc} \tag{18}$$

Similarly, the formulation of the voltage drop must be split between the decision and internal links to ensure that the equations are viable for both built and unbuilt lines. However, owing to the uncertainty of the binary variable x_{ij}, the voltage drop must be modelled as a double inequality constraint, as shown in (19) and (20). However, this uncertainty does not exist for the elements in A_c, thus one equality constraint is sufficient (21).

$$\left(E_{n,i} - E_{n,j}\right) - \left(1 - x_{ij}\right) \leq \frac{P_{ij}^{pu}}{Z_{ref}} \left(rL_{ij} + xL_{ij}\, cos\, \varphi\right) \; ; \forall (\mathrm{i,j}) \in A_d \tag{19}$$

$$\left(E_{n,i} - E_{n,j}\right) + \left(1 - x_{ij}\right) \geq \frac{P_{ij}^{pu}}{Z_{ref}} \left(rL_{ij} + xL_{ij}\, cos\, \varphi\right) \; ; \forall (\mathrm{i,j}) \in A_d \tag{20}$$

$$\left(E_{n,i} - E_{n,j}\right) = \frac{P_{ij}^{pu}}{Z_{ref}} \left(rL_{ij} + xL_{ij}\, cos\, \varphi\right) \; ; \forall (\mathrm{i,j}) \in A_c \tag{21}$$

Note that conversion to the per-unit system is performed using the logic detailed in (22).

$$V_{ref} = V_n \; [kV]; \; A_{ref} = 1 \; [MVA]; Z_{ref} = \frac{1}{V_{ref}^2}; \tag{22}$$

Table 6
Secondary substations spreadsheet.

Rated power [kVA]	Cost [$]	Rated power [kVA]	Cost [$]
25	1000	100	2100
50	1500	200	2800

Once the nodal voltages are calculated, they are constrained by the minimum allowable voltage. The voltage of the connection points should be set according to the predefined logic that the primary substations have a set point of 1 p.u., whereas the points along the MV network have a reduced voltage which is a function of the distance to their respective primary substations, unless accurate information is available. This is the input for the procedure. Furthermore, the voltage setpoint of the fictive microgrid nodes is set to 1 to allow for proper convergence of the optimisation.

$$\underline{E} \leq E_{N,i}\ ; \forall \mathrm{i} \in \mathrm{Nip} \tag{23}$$

$$E_{N,s} \leq E_s\ ; \forall \mathrm{i} \in (\mathrm{Nps} \cup Nmg) \tag{24}$$

The balance equation (25) is necessary to ensure that all communities are electrified. The final constraint was the definition of radiality in the network structure (26). It is clear that the least-cost solution is always radial. However, the authors found that setting this constraint aids the solver in reaching the optimal solution.

$$\sum_{ss \in Nip} P_{ss} - \sum_{s \in Ncp} P_s - \sum_{c \in Nc} P_c = 0 \tag{25}$$

$$\sum_{(i,j) \in A} x_{ij} = n_c \tag{26}$$

4. Case study: Butha-Buthe

The proposed approach was applied to an electrification study of the partially electrified region of Butha-Buthe in Lesotho as part of a collaboration with the local DSO (Lesotho Electricity Company (LEC)). The case study presents a brownfield electrification application, which, in this case, increases the computational burden of the optimisation due to the various options for points of connection for each community. The area is already partially electrified with a distribution network at the 33 kV and 11 kV levels, and the Lesotho Master Plan for Electrification (2018) plans to further extend it to isolated communities. Butha-Buthe covers an area of 1776 km^2 with approximately 118, 000 inhabitants. As a consequence of its heavily mountainous terrain with forests, the deployment and maintenance of electric lines are associated with high costs, making roads crucial in reducing the overall expenses. Owing to the availability of community shapes on OSM (grey shapes of residential areas) and the existing distribution network, as displayed in Fig. 6, the final unelectrified clusters were obtained using the procedure detailed in Fig. 2. In cooperation with the local DSO, connection points for the new lines were identified based on the existing distribution grid and the existing 33/11 kV substations. More details on this will be provided in Section 5 that discusses the results from the study.

In terms of secondary substations, the spreadsheet with rated powers available for electrification in Butha-Buthe and their associated costs are detailed in Table 6. The OpEx considered for each type is 2% from the initial cost, per year. Finally, the characteristics of the conductor suggested by the DSO for use in the case of grid expansion are presented in Table 7, with a nominal voltage level for grid expansion of 11 kV.

Table 8 details the specific parameters used for the simulations, such as the maximum allowable length for the LV feeders and the cost of electricity (CoE), along with the specific block of the procedure in which they are used.

The geospatial data used to create the cost surface are listed in Table 9. They are a mix of raster and vector layers which provide information on the elevation from which the slope layer is computed, average river flow rate, penalized crossing of large rivers, and land cover, with 10 different land cover types provided at high resolutions and road paths.

For the energy demand assessment, load profiles were generated starting from the MTF data and adapted to the energy and peak load values considered in the Lesotho Master Plan for Electrification (Table 10). The categories of users and the associated RAMP input parameters are reported in **ANNEXURE B**.

The parameters used in the input to the off-grid sizing procedure were mainly taken from literature data and are reported in **ANNEXURE C**, Table 15.

5. Results

Fig. 7 shows the output of the clustering procedure, adopted starting from the shapes of the residential areas (grey polygons in the figure), as described in Section 3.1.1. The values of the three input parameters are listed in Table 8. The study identified a total of 72 communities in Butha-

Table 8
Project parameters.

Block	Parameter	Value	Block	Parameter	Value
Block 1	$pop_{threshval}$	120 people	Block 3	*Res*	200 m
Block 1	$dist_{thresh}$	1000 m	Block 3	*disc*	95%
Block 1	c_{thresh}	150 m	Block 3	c_{pl}	0.75
Block 2&4	*Microgrid project lifetime*	10 years	Block 4	*Emin*	0.9 p.u.
Block 3&4	*Power Factor*	0.9	Block 4	*Grid lifetime*	40 years
Block 3	LV_{max}	1000 m	Block 4	*CoE*	150 €/km

Table 9
Sources for geo-spatial data for Butha-Buthe.

Data	Type	Source
Elevation	Raster-30 m	NASA SRTM Digital Elevation [58]
Landcover	Raster-30 m	ESA Landcover CCI [59]
Protected Areas	Vector-polygons	World Database on Protected Areas [60]
Roads	Vector-lines	[61]

Table 10
Energy needs of different categories of users (Butha-Buthe).

User	Peak Power [kW]	Energy [MWh/year]
Households	0.4	2.8
Schools	0.8	12.6
Health Centres	1.5	15.4
Churches	0.5	3.7
Other Businesses	0.8	5.9

Table 7
MV conductor.

Type	r[Ω/km]	x[Ω/km]	Imax [A]	Conductor cost [$/km]	Deployment Cost [$/km]
Al-94mm2	0.306	0.33	350	5800	4200

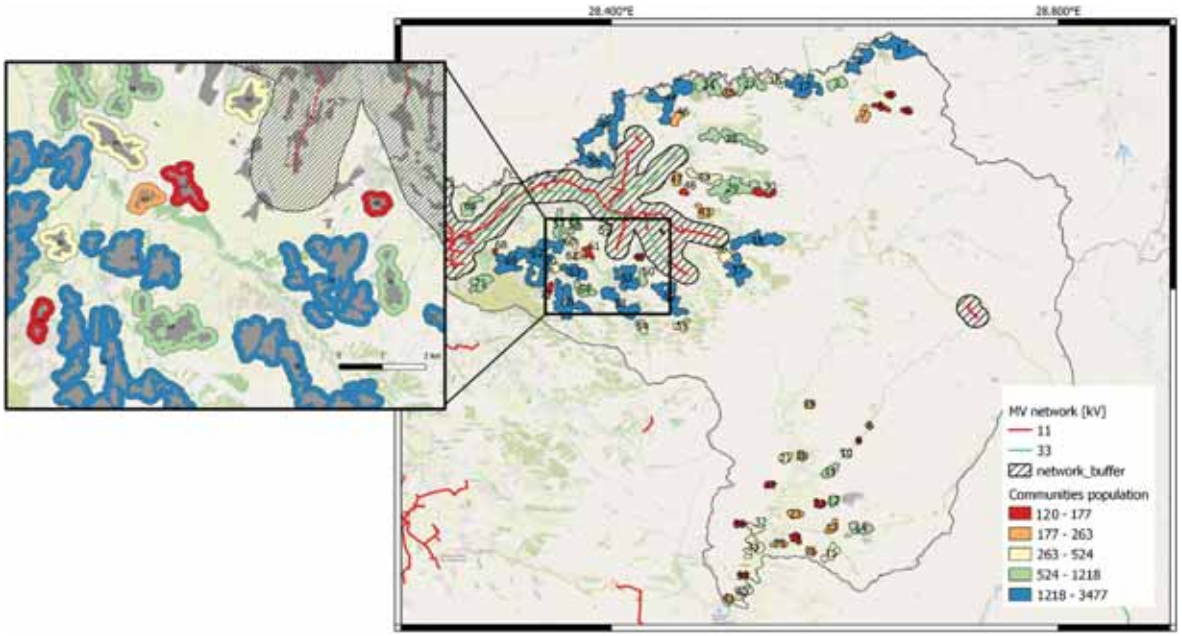

Fig. 7. Identified unelectrified communities.

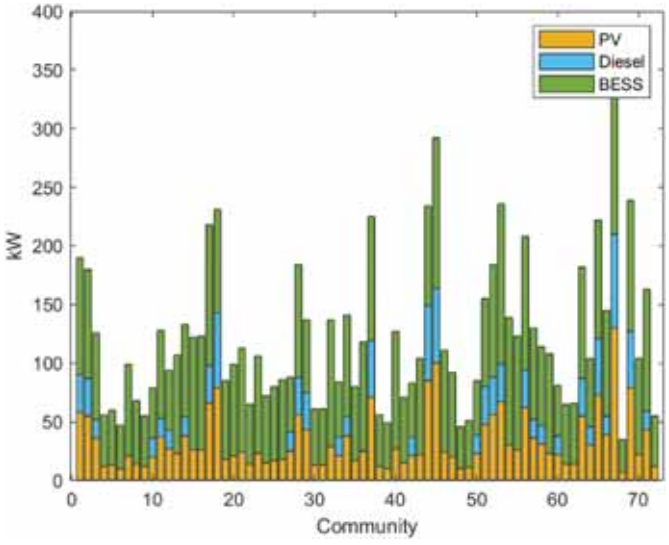

Fig. 8. Results of microgrid sizing procedure.

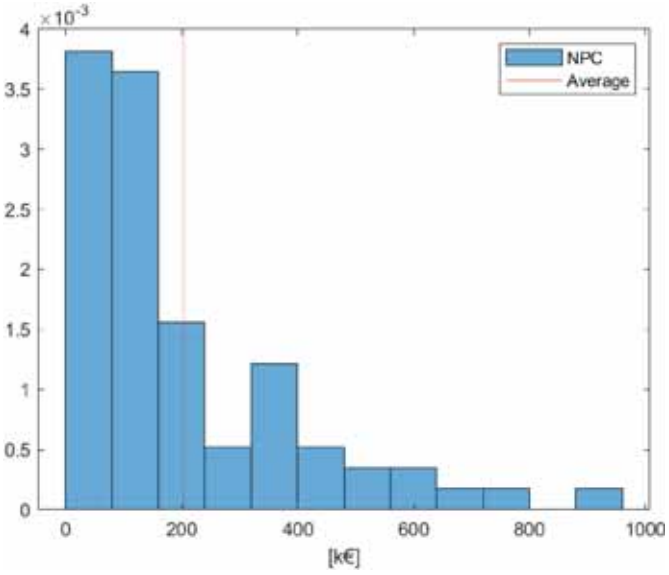

Fig. 9. Probability density function of microgrids' NPC.

Buthe that need to be electrified, with population ranging from 120 to 3477, totalling 46536. These communities represent 39.4% of the total population in Butha-Buthe. Each community is labelled with an integer representing its ID and coloured based on its population size.

Due to the unavailability of specific data on the communities' load demand composition, a generic load profile in p.u. was estimated for the case study. Then, for each community, this load profile was scaled according to the respective population size. These load profiles are useful for dimensioning the microgrid solutions for each community. The peak load associated with each household for the purpose of dimensioning the infrastructure in the case of a grid extension was instead defined starting from data provided by the DSO and set equal to 0.5 kW, which is higher than the peak power of the estimated energy profile. This is because, during the grid expansion planning phase, the DSO must consider the worst-case scenario, considering a larger load growth than forecasted. However, over-dimensioning a microgrid can drastically increase the initial investment, having a detrimental effect on its economic feasibility.

The community load profiles and availability of RES in the area, as well as the techno-economic parameters reported in ANNEXURE C, were used to run MILP optimisation for hybrid microgrid sizing. The results for each community, in terms of the sizes of installed components, are reported in Fig. 8. The solution with a diesel generator was not selected in 54% of the communities, where the 100% renewable solution was found to be more suitable. Wind technology, on the other hand, was not found to be convenient for any of the communities. The microgrids' NPC ranged from a minimum of 32 k$ to a maximum of 954 k$, with the probability distribution shown in Fig. 9.

As previously described, the solution obtained for the stakeholder is both a geospatial distribution of the equipment to be deployed and a list of the equipment and its related costs. A weighted point-based representation of the terrain, that is, the cost surface is shown in Fig. 10. Lesotho has a very challenging terrain, with numerous mountains and

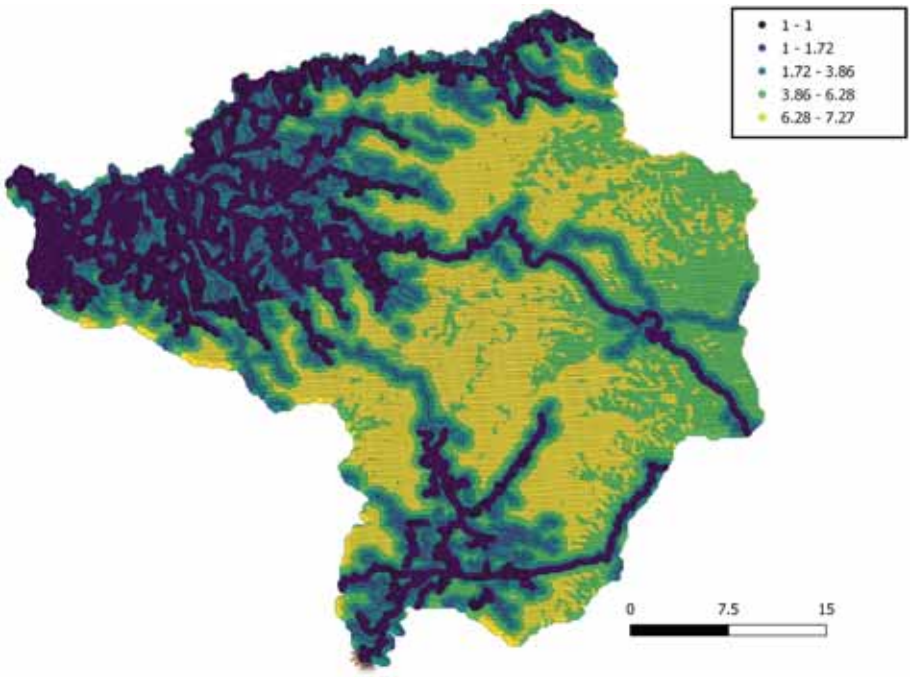

Fig. 10. Cost representation of the terrain in Butha-Buthe.

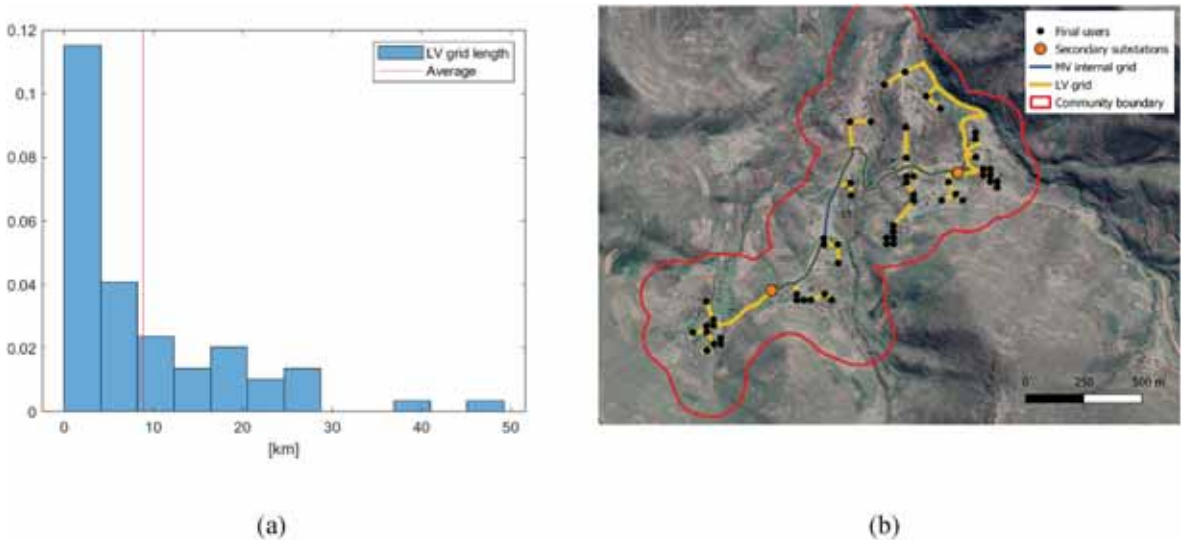

Fig. 11. LV grid routing.

dense forests, which makes deploying electrical lines difficult (represented by yellow points with weights up to 7.27). However, the road network is well-developed, which enables more cost-effective deployment and maintenance of electrical lines. The algorithm recognizes this fact, and thus the points representing the roads are darker and have values approaching one.

The internal grid routing analysis revealed that 37 communities required only one secondary substation owing to their small size, and therefore, no MV grid was needed to interconnect users within the community. The remaining communities had between 2 and 10 substations, indicating a larger size and thus, an internal MV grid. The probability density function of the length of the LV grids of each community is shown in Fig. 11a, while Fig. 11b shows the details of how the grid is routed, following existing roads. The identified length of the MV grids ranged from a minimum of 700 m to a maximum of 13 km.

The final step of the procedure is to input the location, estimated available power, and per-unit voltage of the possible connection points to the existing distribution grid. In the case study, eight connection points were identified, four of which were primary substations and four were connections to the 11 kV MV network. Due to lack of more accurate values, the voltage at the primary substations was defined as 1 p.u., whereas the connections to the existing MV network have a voltage between 0.97–0.99 p.u., depending on their proximity to the respective primary substation. The complete list of the parameters used to

Table 11
Connection points to the existing grid.

Type of connection	Power [MVA]	Voltage [p.u.]	Cost [$]
Primary substation	5	1	1250
MV connection point	2.5	0.97–0.99	750

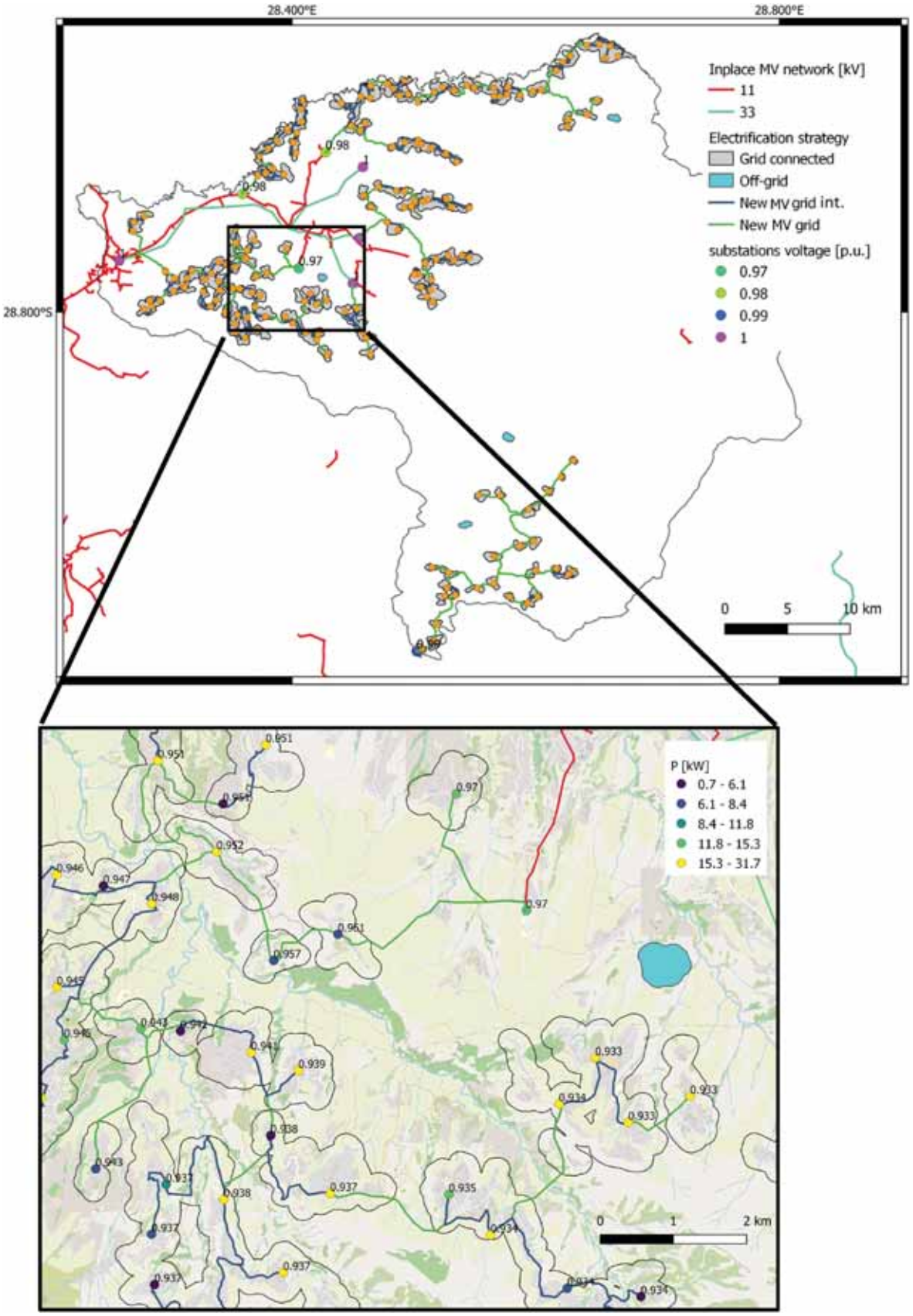

Fig. 12. Results of integrated area optimisation.

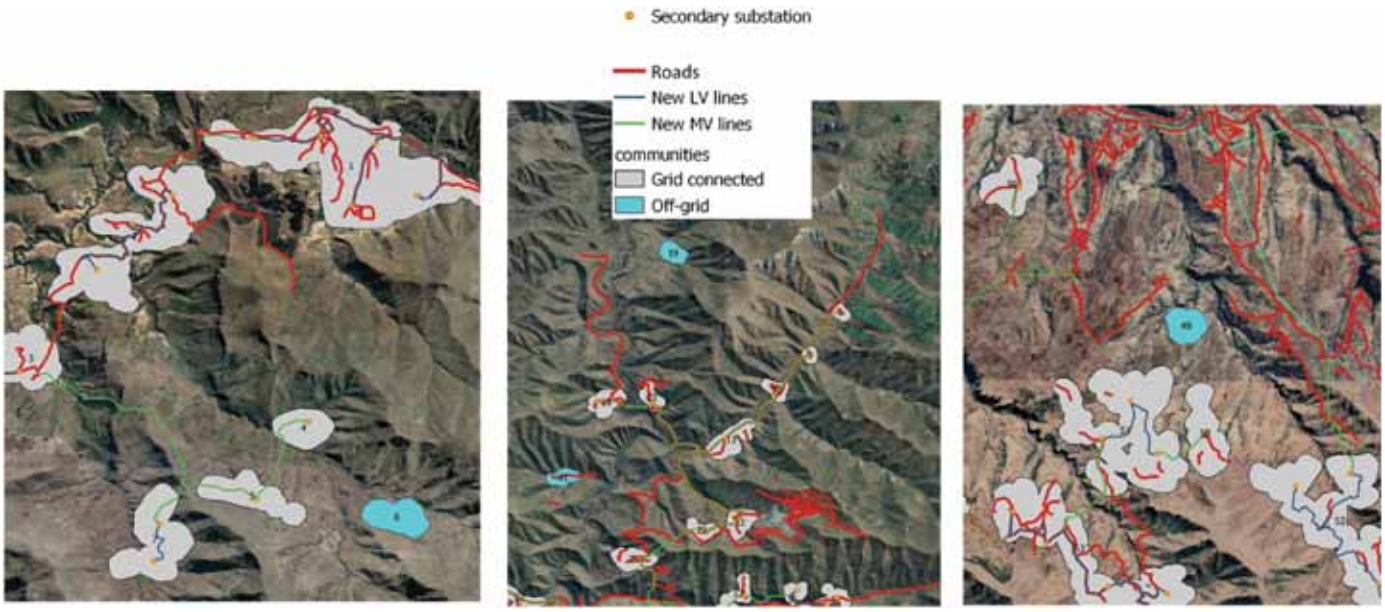

Fig. 13. Zoomed-in off-grid communities.

Table 12
Final output.

Off-grid systems	4	NPC substations [k$]	270
Population electrified	46536	NPC_{40} LV [M$]	6.385
Secondary substations	180	NPC_{40} MV [M$]	3.405
LV [km]	638.5	NPC_{10} MG [k$]	216.6
MV [km]	308.9	NPC_{40} connections[k$]	82.5

represent the points of connection to the existing grid is presented in Table 11.

Fig. 12 represents the overall electrification expansion plan for the area, with the upper image presenting the overall view and the lower showing a specific segment to appreciate the level of detail and output in terms of the steady-state voltage for each secondary substation. The eight connection points to the in-place MV grid are coloured according to the per-unit voltage associated with them. In the magnified image, the secondary substations are coloured according to the nominal power that they must supply. Four communities were identified for electrification with off-grid systems, whereas the others were connected to the existing grid either directly or through other communities. The characteristics that make the algorithm identify communities 6, 19, 31, and 49 as suitable for off-grid electrification are manyfold and help explain the importance of an integrated optimisation rather than an a priori analysis (Fig. 13). They are small communities, with three having a population below 150 inhabitants and one (community 19) having a population of 217 people; therefore, they are among the smallest in the area, as shown in Fig. 7. Communities 6, 19, and 31 were also far from the in-place grid, with distances of the first two of approximately 20 km and of the third being 12 km. Community 49, on the contrary, is closer, at a distance of just a couple of kilometres. Nevertheless, as in the other three off-grid communities, it is poorly interconnected with the rest of the area and is located in a valley without direct road connections. Finally, another factor that influences the choice is the distance between communities; in this case community 19, which is surrounded by difficult terrain, is at a distance of 5 km from the closest one. In conclusion, none of these indicators alone would be able to explain the choice of one strategy over another, and it is only the integrated optimisation that allows capturing complex interrelations among all aspects and provides a holistic solution.

Finally, a complete overview of the costs and equipment used is presented in Table 12. The total investment required to electrify 46 536 inhabitants of Butha-Buthe is close to 10 million.

6. Conclusion and future directions

The procedure proposed in this study is a holistic cost-effective tool for rural electrification planning aimed at assisting stakeholders in the decision-making process, with the provision of detailed expansion plans and cost estimates. It promotes the use of open-source data and is freely available for adoption and improvement. While the state-of-art literature propose a criterion for determining the means of electrification solely as a function of the distance to the electrification network [20], it should be noted that each case is unique and that factors such as load demand, terrain, and population sparsity play a significant role in the optimal electrification plan.

With all the novelties and advancements that this approach brings to the table, it is important to list some of the assumptions made, which are similar to the state-of-the-art. For example, the formulation of the optimisation lies within the hypothesis that there is no correlation between additional consumption and the price of electricity, which is reasonable for a mid-sized expansion project such as the case study in this study. However, when moving to a country-scale analysis, the status of the generation portfolio and the HV network should also be considered. Moreover, the reliability of the national grid and the associated energy costs were not considered, which could be addressed in future studies. Finally, according to the authors, an even more challenging task for future research in this area would be to consider a time-based approach that also takes into account the time to electrification. Including the time dimension in the electrification planning process would be a challenging but crucial step, as it would enable stakeholders to analyse the optimal electrification means for each community based not only on a snapshot cost estimation but also on the time required to deploy electrical infrastructure. Thus, the fact that distant communities would remain unelectrified for a long time if a grid extension strategy was chosen could be considered in the optimisation function as energy not supplied. In this case, the options to first deploy a microgrid and then connect it to the national grid can be analysed. However, this would require a comprehensive analysis of the national grid code and the ability of the system to foster generation units in the distribution network. Ultimately, this would also be beneficial for improving the reliability of the entire network and benefit the planning process as a

whole.

Author credit statement

Aleksandar Dimovski: Methodology, Writing – original draft Methodology, Software, Silvia Corigliano: Conceptualization, Writing – original draft, Software, Darlain Irene Edeme: Data curation, Investigation, Validation, Marco Merlo: Supervision, Writing- Reviewing and Editing, Project administration

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.

Data availability

Data will be made available on request.

Acknowledgments

Aleksandar Dimovski is funded in his research activities by ABB SpA.

ANNEXURE A.

Table 13
Coefficients used for the cost surface creation

Attribute	Value	Linear-coeff	Material-conductor	Material-poles	Material-additional	Works	Permission	O&M
Cost (fraction)			0.18	0.36	0.06	0.22	0.18	0.02
LandCover	1 Trees		0	0	0	4	0	4
LandCover	2 Shrubs		0	0	0	1	0	1
LandCover	3 Grassland		0	0	0	0	0	0
LandCover	4 Cropland		0	0	0	0	4	0
LandCover	5 Aquatic vegetation		0	2	0.1	4	0	4
LandCover	6 Sparse vegetation		0	0	0	1	0	1
LandCover	7 Bare areas		0	0	0	0	0	0
LandCover	8 Built up areas		0	0	0	0	1	0
LandCover	9 Snow and/or ice		0.2	2	0.1	2	0	2
LandCover	10 Open water		10	0	10	6	0	6
Road Distance	<100 m		0	0	0	0.1	0	0.1
Road Distance	<500 m		0	0	0	0.2	0	0.2
Road Distance		intercept	0	0	0	0.2		0.2
Road Distance		Angular coefficient	0	0	0	0.000833	0	0.000833
Slope	<2°		0	0	0	0	0	0
Slope	<5°		0	0.1	0	0.1	0	0.1
Slope	<10°		0	0.5	0	0.5	0	0.5
Slope	<20°		0	2	0	2	0	2
Slope	>20		0	10	0	10	0	10
Protected areas			0	0	0	0	100	0
Rivers	>100		0	10	2	10	0	10

ANNEXURE B.

Table 14
RAMP input parameters

Appliance	*nij*	*Pij*	*fwij*	*Rfwij*	*fcij*	*ftij*	*Rftij*
		[W]	*[h]*	*%*	*[h]*	*[h]*	*%*
Health-urban							
Internal Lights	36	20	8–12; 14–24	20	3	12	20
External Lights	15	25	16–24	20	13	13	20
Phone charger	10	5	0–24	20	0.5	5	20
Sterilizer	2	1500	6–22	20	0.5	1	20
TV	3	60	7–17	20	0.5	2	20
PC	10	50	8–12; 17–24	20	0.1	5	20
Fridge	4	250	0–24	20	0.5		20
Fridge2	2	500	0–24	20	0.5		20
Microscope	3	200	7–17	20	0.5	3	20
Centrifuge	3	200	7–17	20	0.5	5	20
Monitor	3	50	7–17	20	0.5	5	20
School							
External Lights	4	25	17–06	0	5	5	0
Internal Lights	18	20	7–17	0	5	5	0
PC	13	50	7–17	0	10	10	0
TV	3	60	7–17	0	10	10	0
Worship							
External Lights	3	25	16–21	0	7	7	
Internal Lights	8	20	16–21	0	7	7	
TV	1	20	16–21	0	2	6	30

(*continued on next page*)

Table 14 (*continued*)

Appliance	*nij*	*Pij*	*fwij*	*Rfwij*	*fcij*	*ftij*	*Rftij*
		[W]	*[h]*	*%*	*[h]*	*[h]*	*%*
PC	1	5	16–21	0	1	3	90
Merchants							
External Lights	2	25	16–22	0	6	6	0
Internal Lights	3	20	16–22	10	6	6	0
Freezer	1	300	10–18	0	8	8	0
Sound System	1	20	10–22	0	2	6	30
Phone Charger	1	5	10–22	0	1	3	90
Fan	1	80	10–22	0	3	6	10
Tailors							
External Lights	1	25	16–19	0	3	3	0
Internal Lights	3	20	16–19	0	4	6	10
Phone Charger	1	5	10–19	0	1	3	90
Radio	1	5	10–19	0	3	6	25
Fan	1	60	10–19	0	3	6	30
Barbers							
External Lights	1	20	16–19	0	7	7	0
Internal Lights	3	15	16–19	0	4	6	10
Radio	1	5	10–19	0	1	5	25
Phone Charger	1	5	10–19	0	1	3	90
Households Tier 3							
Internal Lights	8	5	18–03	20	0.5	5	20
Phone charger	4	5	0–10; 13–16; 18-24	20	0.5	3	20
Radio	1	5	6–10; 17–24	20	0.5	5	20
External Lights	1	7	18–7	20	1	12	20
TV	1	90	8–15; 17–24	20	0.15	4	20
PC	2	60	8–24	20	0.15	4	20
Fan	NN 2	60	424	20	0.15	6	20
Fridge	−1	200	0–24	20	0.5		20
Food processor	2	350	18–20	20	0.15	0.5	20
Water pump	NNN 2	60	1201	20	0.15	3	20
Rice cooker	2	60	12–-15; 20–01	20	0.15	1	20
Households Tier 4							
Internal Lights	16	7	18–03	20	0.5	5	20
Phone charger	4	5	0–10; 13–16; 18-24	20	0.5	3	20
Radio	1	5	6–10:17–24	20	0.5		20
External Lights	2	7	18–7	20	1	12	20
TV	1	150	8–15; 17–24	20	0.15	4	20
PC	2	60	824	20	0.15	446	20
Fan	2	60	424	20	0.15	6	20
Fridge	2	300	0–24	20	0.5		20
Food processor	2	350	18–20	20	0.5	1	20
Iron	1	1000	620	20	0.5	1	20
Ari dryer	1	1000	17–24	20	0.15	0.5	20
Toaster	1	1000	6–9; 18–21	20	0.15	0.5	20
Microwave	1	700	6–9; 11–14; 18-21	20	0.15	1	20

ANNEXURE C.

Table 15
Input parameters of microgrid optimisation

Symbol	Parameter	Value	U.M.
F	Cost of fuel	0.75	[$/1]
y	project lifetime	10	[years]
Na	number of typical days	12	[day/year]
ENS	maximum energy not supplied	5	[%]
RES	minimum energy produced by RES	0	[%]
Yd	load forecast error	0.1	[0–1]
Ypu	PV forecast error	0.1	[0-1
	WT forecast error	0.25	[0–1]
ir	Interest Rate	0.06	[0–1]
	PV modules		
Cp	Unitary capacity	1	[kW]
CC_2	Capital cost	1400	[$/unit]
Mp ylife	O &M yearly cost Lifetime	1020	[$/unit/y][years]
	Wind turbine		
Cu	Unitary capacity	10	[kW]
CCW	Capital cost	27000	[$/unit]
M_2	O &M yearly cost	540	[$/unit/y]
Ylife	Lifetime	20	[years]
Diesel generator			
Symbol	Parameter	Value	U.M.

(*continued on next page*)

Table 15 (*continued*)

Symbol	Parameter	Value	U.M.
Cg	Unitary capacity	16	[kW]
CCg	Capital cost	11000	[$/unit]
M_1 Hlife	O &M hourly cost Lifetime	0.215000	[$/unit/h][hours]
A	cost coefficient	0.4672	[1/h]
B	cost coefficient	0.3	[1/h/kW]
Pg	Min Power of DG	0.3	[0–1]
BESS			
Cb	Unitary capacity	1	[kWh]
CCb	Capital cost	400	[$/unit]
M_2	O &M yearly cost	10	[$/unit/y]
Hlife	Lifetime	3000	[kWh]
ylife	Lifetime	15	[years]
nb	efficiency	0.975	H
PQb	maximum BESS power-to-energy ratio	1	[kW/kwh]
DOD	maximum BESS depth of discharge	0.9	[0–1]

References

[1] UNDP, United Nations Development Programme, Sustainable Development Goals-Goal 7: Affordable and Clean Energy, 2020.

[2] B. Pillot, M. Muselli, P. Poggi, J.B. Dias, On the impact of the global energy policy framework on the development and sustainability of renewable power systems in Sub-Saharan Africa: the case of solar PV. http://arxiv.org/abs/1704.01480, 2017.

[3] Our World in Data, Energy intensity vs. gdp per capita, 2016, https://ourworldindata.org/grapher/energy-use-per-capita-vs-gdp-per-capita, 2020.

[4] K. Aidoo, R.C. Briggs, Underpowered: rolling blackouts in Africa disproportionately hurt the poor, Afr. Stud. Rev. 62 (3) (2019) 112–131, https://doi.org/10.1017/asr.2018.78.

[5] M. Torero, The impact of rural electrification: challenges and ways forward, Rev. Écon. Du. Développement 23 (HS) (2016) 49–75, https://doi.org/10.3917/edd.hs03.0049.

[6] F. Kemausuor, M.D. Sedzro, I. Osei, Decentralised energy systems in Africa: coordination and integration of off-grid and grid power systems—review of planning tools to identify renewable energy deployment options for rural electrification in Africa, in: Current Sustainable/Renewable Energy Reports, vol. 5, Springer Nature, 2018, pp. 214–223, https://doi.org/10.1007/s40518-018-0118-4, 4.

[7] J. Morrissey, Achieving universal electricity access at the lowest cost: a comparison of least-cost electrification models, Energy for Sustainable Development 53 (2019) 81–96, https://doi.org/10.13140/RG.2.2.23778.68807.

[8] P. Ciller, S. Lumbreras, Electricity for all: the contribution of large-scale planning tools to the energy-access problem, in: Renewable and Sustainable Energy Reviews, vol. 120, Elsevier Ltd, 2020, https://doi.org/10.1016/j.rser.2019.109624.

[9] HOMER Energy LLC, HOMER pro - microgrid software for designing optimized hybrid microgrids, n.d. http://www.homerenergy.com/.

[10] S.L. Balderrama, V. Lemort, S. Balderrama, W. Canedo, M. Fernandez, S. Quoilin, Techno-economic Optimization of Isolate Micro-grids Including PV and Li-Ion Batteries in the Bolivian Context, Proceedings of ECOS 2016-the 29th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems, 2016. https://www.researchgate.net/publication/304525451.

[11] R. Dufo-López, J.M. Lujano-Rojas, J.L. Bernal-Agustín, Comparison of different lead–acid battery lifetime prediction models for use in simulation of stand-alone photovoltaic systems, Appl. Energy 115 (2014) 242–253, https://doi.org/10.1016/j.apenergy.2013.11.021.

[12] S. Mashayekh, M. Stadler, G. Cardoso, M. Heleno, A mixed integer linear programming approach for optimal DER portfolio, sizing, and placement in multi-energy microgrids, Appl. Energy 187 (2017) 154–168, https://doi.org/10.1016/j.apenergy.2016.11.020.

[13] M. Petrelli, D. Fioriti, A. Berizzi, D. Poli, Multi-year planning of a rural microgrid considering storage degradation, IEEE Trans. Power Syst. 36 (2) (2021) 1459–1469, https://doi.org/10.1109/TPWRS.2020.3020219.

[14] S. Szabó, K. Bódis, T. Huld, M. Moner-Girona, Sustainable energy planning: leapfrogging the energy poverty gap in Africa, Renew. Sustain. Energy Rev. 28 (2013) 500–509, https://doi.org/10.1016/j.rser.2013.08.044.

[15] M. Moner-Girona, M. Solano-Peralta, M. Lazopoulou, E.K. Ackom, X. Vallve, S. Szabó, Electrification of Sub-Saharan Africa through PV/hybrid mini-grids: reducing the gap between current business models and on-site experience, Renew. Sustain. Energy Rev. 91 (2018) 1148–1161, https://doi.org/10.1016/j.rser.2018.04.018.

[16] S. Pfenninger, B. Pickering, Calliope: a multi-scale energy system modelling framework, J. Open Source Softw. 3 (29) (2018) 825, https://doi.org/10.21105/joss.00825.

[17] M. Howells, H. Rogner, N. Strachan, C. Heaps, H. Huntington, S. Kypreos, A. Hughes, S. Silveira, J. DeCarolis, M. Bazillian, A. Roehrl, OSeMOSYS: the open source energy modeling system, Energy Pol. 39 (10) (2011) 5850–5870, https://doi.org/10.1016/j.enpol.2011.06.033.

[18] T. Niet, A. Shivakumar, F. Gardumi, W. Usher, E. Williams, M. Howells, Developing a community of practice around an open source energy modelling tool, Energy Strategy Rev. 35 (2021), 100650, https://doi.org/10.1016/j.esr.2021.100650.

[19] S. Hilpert, C. Kaldemeyer, U. Krien, S. Günther, C. Wingenbach, G. Plessmann, The Open Energy Modelling Framework (oemof) - a new approach to facilitate open science in energy system modelling, Energy Strategy Rev. 22 (2018) 16–25, https://doi.org/10.1016/j.esr.2018.07.001.

[20] A. Korkovelos, B. Khavari, A. Sahlberg, M. Howells, C. Arderne, The role of open access data in geospatial electrification planning and the achievement of SDG7. An OnSSET-based case study for Malawi, Energies 12 (7) (2019) 1395, https://doi.org/10.3390/en12071395.

[21] ESMAP and WBG, Senegal's SE4ALL. Rural Electrification. Action Agenda and Investment Prospectus 2018, 2018.

[22] A. Bosisio, E. Amaldi, A. Berizzi, C. Bovo, S. Fratti, A MILP Approach to Plan an Electric Urban Distribution Network with an H-Shaped Layout, 2015 IEEE Eindhoven PowerTech, 2015, pp. 1–5, https://doi.org/10.1109/PTC.2015.7232652.

[23] C.M. Domingo, T. Gomez San Roman, A. Sanchez-Miralles, J.P. Peco Gonzalez, A. Candela Martinez, A reference network model for large-scale distribution planning with automatic street map generation, IEEE Trans. Power Syst. 26 (1) (2011) 190–197, https://doi.org/10.1109/TPWRS.2010.2052077.

[24] P. Ciller, D. Ellman, C. Vergara, A. Gonzalez-Garcia, S.J. Lee, C. Drouin, M. Brusnahan, Y. Borofsky, C. Mateo, R. Amatya, R. Palacios, R. Stoner, F. de Cuadra, I. Perez-Arriaga, Optimal electrification planning incorporating on- and off-grid technologies: the reference electrification model (REM), Proc. IEEE 107 (9) (2019) 1872–1905, https://doi.org/10.1109/JPROC.2019.2922543.

[25] IED, Geosim [online], https://www.ied-sa.com/en/products/planning/geosim-gb.html, 2018.

[26] Rwanda Energy Group, TheNational Electrification Plan: Report on Definition of Technologies (On-Grid and Off-Grid) at Village Level, 2019.

[27] IED, Cambodia-rural electrification plan, 2011 Technical report. https://energypedia.info/wiki/File:Rural_Electrification_Plan_Final_Report_-_Cambodia.pdf.

[28] F. Kemausuor, E. Adkins, I. Adu-Poku, A. Brew-Hammond, V. Modi, Electrification planning using Network Planner tool: the case of Ghana, Energy for Sustainable Development 19 (2014) 92–101, https://doi.org/10.1016/j.esd.2013.12.009.

[29] R. Fronius, S. Levy, S. Gamier, Rural electrification in developing countries using laper software, in: Sixteenth European Photovoltaic Solar Energy Conference, 2000, 1st ed.

[30] P. Blechinger, C. Cader, P. Bertheau, Least-cost electrification modeling and planning—a case study for five Nigerian federal states, Proc. IEEE 107 (9) (2019) 1923–1940, https://doi.org/10.1109/JPROC.2019.2924644.

[31] L. Moretti, M. Astolfi, C. Vergara, E. Macchi, J.I. Pérez-Arriaga, G. Manzolini, A design and dispatch optimization algorithm based on mixed integer linear programming for rural electrification, Appl. Energy 233 (234) (2019) 1104–1121, https://doi.org/10.1016/j.apenergy.2018.09.194.

[32] T. Gönen, B.L. Foote, Distribution-system planning using mixed-integer programming, IEE Proceedings C Generation, Transmission and Distribution 128 (2) (1981) 70, https://doi.org/10.1049/ip-c.1981.0010.

[33] M. Resener, S. Haffner, L.A. Pereira, P.M. Pardalos, Optimization techniques applied to planning of electric power distribution systems: a bibliographic survey, Energy Systems 9 (3) (2018) 473–509, https://doi.org/10.1007/s12667-018-0276-x.

[34] M. Jooshaki, A. Abbaspour, M. Fotuhi-Firuzabad, G. Muñoz-Delgado, J. Contreras, M. Lehtonen, J.M. Arroyo, An enhanced MILP model for multistage reliability-constrained distribution network expansion planning, IEEE Trans. Power Syst. 37 (1) (2022) 118–131, https://doi.org/10.1109/TPWRS.2021.3098065.

[35] M. Sedghi, A. Ahmadian, M. Aliakbar-Golkar, Assessment of optimization algorithms capability in distribution network planning: review, comparison and modification techniques, Renew. Sustain. Energy Rev. 66 (2016) 415–434, https://doi.org/10.1016/j.rser.2016.08.027.

[36] S. Ganguly, N.C. Sahoo, D. Das, Mono- and multi-objective planning of electrical distribution networks using particle swarm optimization, Applied Soft Computing Journal 11 (2) (2011) 2391–2405, https://doi.org/10.1016/j.asoc.2010.09.002.
[37] Gurobi. https://www.gurobi.com/, 2022.
[38] B. Turkay, Distribution system planning usign mixed integer programming, ELEKTRIK 6 (1) (1998) 37–48.
[39] J.A. Lukes, Efficient algorithm for the partitioning of trees, IBM J. Res. Dev. 18 (3) (1978) 217–224.
[40] GISele, E4Growing. https://github.com/Energy4Growing/gisele_v02, 2022.
[41] S. Corigliano, T. Carnovali, D. Edeme, M. Merlo, Holistic geospatial data-based procedure for electric network design and least-cost energy strategy, Energy for Sustainable Development 58 (2020) 1–15, https://doi.org/10.1016/j.esd.2020.06.008.
[42] G.T.F. Vinicius, C. Silvia, D. Aleksandar, B. Massimo, M. Marco, Rural electrification planning based on graph theory and geospatial data: a realistic topology oriented approach, Sustainable Energy, Grids and Networks 28 (2021), 100525, https://doi.org/10.1016/j.segan.2021.100525.
[43] Microsoft, Buildings' footprints. https://www.microsoft.com/en-us/maps/building-footprints, 2021.
[44] B. Khavari, A. Korkovelos, A. Sahlberg, M. Howells, F. Fuso Nerini, Population cluster data to assess the urban-rural split and electrification in Sub-Saharan Africa, Sci. Data 8 (1) (2021), https://doi.org/10.1038/s41597-021-00897-9.
[45] M. Yosef, M.M. Sayed, H.K.M. Youssef, Allocation and sizing of distribution transformers and feeders for optimal planning of MV/LV distribution networks using optimal integrated biogeography based optimization method, Elec. Power Syst. Res. 128 (2015) 100–112, https://doi.org/10.1016/j.epsr.2015.06.022.
[46] F. Lombardi, S. Balderrama, S. Quoilin, E. Colombo, Generating high-resolution multi-energy load profiles for remote areas with an open-source stochastic model, Energy 177 (2019) 433–444, https://doi.org/10.1016/j.energy.2019.04.097.
[47] G. Padam, D. Rysankova, E. Portale, B.B. Koo, S. Keller, G. Fleurentin, Ethiopia - beyond Connections: Energy Access Diagnostic Report Based on the Multi-Tier Framework, 2018.
[48] G. Falchetta, N. Stevanato, M. Moner-Girona, D. Mazzoni, E. Colombo, M. Hafner, The M-LED platform: advancing electricity demand assessment for communities living in energy poverty, Environ. Res. Lett. 16 (7) (2021), 074038, https://doi.org/10.1088/1748-9326/ac0cab.
[49] Renewables Ninja, n.d. https://www.renewables.ninja/. (Accessed 5 November 2021).
[50] S. Pfenninger, I. Staffell, Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data, Energy 114 (2016) 1251–1265, https://doi.org/10.1016/j.energy.2016.08.060.
[51] M.M. Rienecker, M.J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M. G. Bosilovich, S.D. Schubert, L. Takacs, G.-K. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. da Silva, W. Gu, J. Joiner, R.D. Koster, R. Lucchesi, J. Woollen, MERRA: NASA's modern-era retrospective analysis for research and applications, J. Clim. 24 (14) (2011) 3624–3648, https://doi.org/10.1175/JCLI-D-11-00015.1.
[52] S. Corigliano, Geospatial based Methodology for Rural Electrification Planning [PhD], Politecnico di, Milano, 2022.
[53] T.R. Etherington, Least-cost modelling and landscape ecology: concepts, applications, and opportunities, Current Landscape Ecology Reports 1 (2017) 40–53.
[54] J. van Bemmelen, W. Quak, M. van Hekken, P. van Oosterom, Vector vs. Raster-based algorithms for cross country movement planning, Proceedings Auto-Carto 11 (1993) 304–317.
[55] H. Li, et al., Building highly detailed synthetic electric grid data sets for combined transmission and distribution systems, IEEE Open Access Journal of Power and Energy 7 (2020) 478–488, https://doi.org/10.1109/OAJPE.2020.3029278.
[56] C. Mateo, et al., Building large-scale U.S. Synthetic electric distribution system models, IEEE Trans. Smart Grid 11 (6) (2020) 5301–5313, https://doi.org/10.1109/TSG.2020.3001495.
[57] C. Monteiro, I.J. Ramirez-Rosado, V. Miranda, P.J. Zorzano-Santamaria, E. Garcia-Garrido, L.A. Fernandez-Jimenez, GIS spatial analysis applied to electric line routing optimization, IEEE Trans. Power Deliv. 20 (2) (2005) 934–942, https://doi.org/10.1109/TPWRD.2004.839724.
[58] T.G. Farr, P.A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank, D. Alsdorf, The shuttle radar topography mission, Rev. Geophys. 45 (2) (2007) RG2004, https://doi.org/10.1029/2005RG000183.
[59] W. Li, N. MacBean, P. Ciais, P. Defourny, C. Lamarche, S. Bontemps, R. A. Houghton, S. Peng, Gross and net land cover changes in the main plant functional types derived from the annual ESA CCI land cover maps (1992–2015), Earth Syst. Sci. Data 10 (1) (2018) 219–234, https://doi.org/10.5194/essd-10-219-2018.
[60] UNEP-WCMC and IUCN, Protected planet: the world database on protected areas (WDPA). www.protectedplanet.net, 2020.
[61] OSM. (n.d.). OpenStreetMap Contributors. OpenStreetMap. Retrieved January 10, 2022, from openstreetmaps.org.