Engineering Applications of Artificial Intelligence 137 (2024) 109249

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

A geospatial clustering algorithm and its integration into a techno-economic rural electrification planning model

Mirelys Torres-Pérez [a], Javier Domínguez [b,*], Luis Arribas [b], Julio Amador [c], Pedro Ciller [d], Andrés González-García [d,e]

[a] Department of Informatics, University of Las Tunas, Las Tunas, 75100, Cuba
[b] Renewable Energies Division, CIEMAT, Av. Complutense, 40, Madrid, 28040, Spain
[c] Department of Electrical Engineering, Polytechnic University of Madrid, Madrid, 28012, Spain
[d] Instituto de Investigación Tecnológica, Universidad Pontificia Comillas (IIT-Comillas), 28015, Madrid, Spain
[e] Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA

ARTICLE INFO

Keywords:
Constrained clustering
Density-based clustering
Graph-based clustering
Rural electrification
Geospatial analysis
Techno-economic software tool

ABSTRACT

Rural electrification planning is a complex process requiring careful consideration of various factors to ensure efficient and cost-effective solutions. Existing clustering methods in academic literature often fall short in this context, as they typically do not account for geographical barriers, restricted areas, and key electrical and geospatial metrics simultaneously. This can result in clusters that do not meet the energy needs of the study region, potentially causing inefficient energy distribution and increased costs. This study presents a novel clustering algorithm, RElect_MGEC (Rural Electrification Microgrid and Grid Extension Clustering), specifically designed for techno-economic planning in rural areas. The RElect_MGEC algorithm combines density-based and graph clustering methods to group households while considering constraints imposed by geographic barriers, electricity power, and distance from the generation center. The algorithm was implemented within the IntiGIS (Geographic Information System for Rural Electrification) model and evaluated using a real-world dataset of 10,995 unelectrified households in rural Yoro, Honduras. The evaluation involved comparisons with established clustering algorithms, focusing on metrics such as the number of valid clusters, Levelized Cost of Electricity (LCOE), and execution time. The results demonstrate the algorithm's effectiveness in scenarios with equal and varying demands, highlighting its robustness, flexibility, and ability to achieve cost savings within shorter timeframes. Additionally, this approach enables the assessment of distribution infrastructures, such as microgrids and grid extensions, ensuring an effective power generation and distribution. The integration of the RElect_MGEC algorithm into IntiGIS results in an enhanced model that enables a comprehensive and informed decision-making process for rural electrification planning.

1. Introduction

Artificial Intelligence (AI) holds significant potential in advancing Sustainable Development Goals (SDGs) across various domains. AI technologies can contribute to poverty alleviation (Hall et al., 2022; Lopez-Vargas et al., 2022; Jejeniwa et al., 2024), enhance quality education (Kabudi, 2022; Lin et al., 2023), and improve clean water and sanitation efforts (Mehmood et al., 2020).

For example, AI-driven predictive models and gene expression programming have been shown to optimize decision-making processes in construction, as evidenced by Nawaz et al.'s (2024b) work on predicting soil cohesion and friction angles. Similarly, AI-enabled multivariate formulations help estimate the frictional strength of fiber-reinforced soils, demonstrating the power of advanced modeling techniques in resource allocation (Nawaz et al., 2024a).

Furthermore, AI technologies can enhance climate action, support biodiversity monitoring, aid in conservation efforts, and contribute to sustainable urban development, thereby playing a critical role in achieving specific SDG targets (Vinuesa et al., 2020). Geospatial analysis, such as the interpolation of geotechnical data and spatial mapping of soil parameters conducted by Hassan et al. (2022, 2023), emphasizes the importance of accurate data representation in planning efforts.

The seventh Sustainable Development Goal (SDG 7) aims to "Ensure

* Corresponding author.
E-mail address: javier.dominguez@ciemat.es (J. Domínguez).

https://doi.org/10.1016/j.engappai.2024.109249
Received 23 August 2023; Received in revised form 22 August 2024; Accepted 30 August 2024
Available online 8 September 2024

Abbreviations and acronyms:

RElect_MGEC	Rural Electrification Microgrid and Grid Extension Clustering
RElect_BUC	Rural Electrification Bottom-up Clustering
MGE	Microgrid and Grid Extension
LCOE	Levelized Cost of Electricity
LCEM	Least-Cost Electrification Models
GIS	Geographic Information System
CIEMAT	Center for Energy, Environmental, Environmental and Technological Research
IntiGIS®	Geographic Information System for Rural Electrification official project of CIEMAT
QGIS	Quantum Geographic Information System
GUI	Graphical User Interface
LV	Low Voltage
MST	Minimum Spanning Tree
DT	Delaunay Triangulation
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
HCA	Hierarchical Clustering Agglomerative
cons	consumers
NC	Number of Clusters
NCV	Number of Valid Clusters
Exp	Experiment
PV	Photovoltaic
GTI	Global Irradiation at optimum Tilt
RLI-GEPT	Reiner Lemoine Institut geospatial electrification planning tool
IEA	International Energy Agency
REM	Reference Electrification Model
OnSSET	Open Source Spatial Electrification Tool
RE2NAF	Renewable Energies for Rural Electrification of Africa
GEOSIM	Geospatial planning for rural electrification
LECGIS	Levelized Electricity Cost Geographic Information System
GDAL/OGR	“GDAL” Geospatial Data Abstraction Library and “OGR” Simple Features Library
Map_{Cr}	Map with Consumer Clustering
UN	United Nations
UNDP	United Nations Development Program
AI	Artificial Intelligence
SDGs	Sustainable Development Goals
SDG 7	The seventh Sustainable Development Goal
ESCAP	Economic and Social Commission for Asia and the Pacific

access to affordable, reliable, sustainable, and modern energy for all by 2030” (UN, 2024). Access to energy is essential for economic and social development, playing a crucial role in eradicating poverty, enhancing quality of life, and fostering the development of rural areas. Moreover, SDG 7 has been found to positively influence and strengthen 16 other SDGs, highlighting its central role in achieving broader sustainable development objectives (ESCAP, 2016).

The COVID-19 crisis abruptly halted several years of consistent progress and exacerbated households’ already limited energy purchasing power in developing countries (IEA, 2022). The pandemic caused significant disruptions in supply chains, delayed infrastructure projects, and placed considerable strain on government budgets, leading to setbacks in energy access initiatives, particularly in developing regions (UN, 2020; IEA, 2022).

Furthermore, many households that had previously gained access to electricity faced challenges in maintaining service due to worsening economic conditions. In India, lockdown measures disrupted solar energy projects, resulting in substantial delays in rural electrification efforts (IEA, 2021). The global economic downturn further diverted resources away from renewable energy investments, impeding the advancement of SDG 7 (Min and Perucci, 2020).

The 2024 edition of “Tracking SDG 7: The Energy Progress Report” states that achieving universal access to clean and affordable energy by 2030 remains a significant challenge, especially in developing regions where population growth could offset progress (IEA, 2024). According to (UNDP, 2024), nearly 733 million people worldwide lack access to electricity, with the majority living in rural areas.

Rural electrification is a critical issue that requires effective planning to ensure efficient resource utilization. However, this planning is hindered by several factors, including the high cost of extending the national grid to remote areas, the low population density in dispersed settlements, and limited financial resources. To address these challenges, decision-makers rely on energy planning tools and models, such as Least-Cost Electrification Models (LCEMs).

A significant limitation of most existing LCEMs is their tendency to group consumers based on predefined natural boundaries of communities or raster cells. This approach fails to accurately identify the distribution infrastructure required to connect individual consumers to the power source, such as main grid extensions or microgrid generation sites (Morrissey, 2019; Ciller and Lumbreras, 2020).

Additionally, when a model operates at the community or raster cell level, it will assign a single mode of electrification (grid extension, microgrid, or a combination of individual isolated systems) to the entire community or cell. However, the optimal solution may require a mix of different modes of electrification (for example, a combination of microgrids for several consumers within the community or cell, along with some isolated systems for more dispersed consumers).

In the realm of artificial intelligence, clustering falls under the field of unsupervised learning. Clustering algorithms have emerged as a highly effective tool for analyzing complex datasets and uncovering underlying patterns (Rodriguez et al., 2019). These algorithms have been applied across various domains, such as developing recommendation systems and analyzing social media networks (Oyelade et al., 2019).

Clustering algorithms can address the limitations of LCEMs by identifying the actual patterns of individual consumers within rural settlements. This approach offers a more granular understanding of consumer distribution across a region, enabling the identification of the most suitable electrification solutions—isolated systems, microgrids, or grid extensions—for each cluster of consumers.

This enhanced granularity in modeling enables more accurate identification of electrification needs, ensuring that the selected infrastructure is optimized for the actual spatial patterns and demands of the population. By leveraging the power of clustering algorithms, planners can move beyond traditional, often oversimplified approaches, gaining a nuanced understanding that supports efficient resource allocation, reduces costs, and improves the overall effectiveness of electrification initiatives.

Furthermore, as part of machine learning, clustering contributes to the broader goals of AI-driven decision-making by offering a scalable, data-driven approach to problem-solving. In rural electrification, clustering can support the efficient allocation of resources, reduce costs, and improve the effectiveness of electrification initiatives. By doing so, clustering algorithms directly contribute to the achievement of SDG 7.

Building on the background, we pose the research question: How can consumer clustering be incorporated into LCEMs to improve the evaluation of electrification alternatives and enhance decision-making accuracy in energy planning?

To address this question, the objectives of the study are.

- Design a geospatial clustering algorithm for techno-economic rural electrification planning that balances computational efficiency with solution accuracy.
- Implement and integrate the proposed clustering method into a techno-economic rural electrification planning model.
- Evaluate the effectiveness of the developed approach through experiments on a real-world dataset, comparing it with established clustering algorithms.

The rest of the paper is organized as follows. Section 2 analyzes the state-of-the-art clustering approaches and their challenges in large-scale techno-economic and geospatial planning for rural electrification. Section 3 details the proposed geospatial clustering algorithm and the evolution of the IntiGIS model, with the integration of the clustering algorithm being the key advancement. Section 4 describes the application of the proposed method and its comparison with different clustering variants in two rural scenarios in Yoro, Honduras, summarizing the main results. Finally, Sections 5 and 6 present the conclusions, future directions and study limitations. In addition, the work includes four appendices. The first of these contains the three complementary algorithms (MaxLongCGen, centre_Graph and node_MaxLongCGen). The second provides the Descriptive Statistics of the Clusters in Tables 10 and 11. The remaining two appendices present a Statistical Summary of Electrification Solutions for two scenarios. The first one includes Tables 12-22 and the second one Tables 23-33. By incorporating clustering algorithms into rural electrification planning, this study aims to improve the accuracy and efficiency of the planning process, leading to more effective and sustainable electrification solutions.

2. Related works: clustering challenges in large-scale techno-economic and geospatial planning for rural electrification

From a techno-economic perspective, rural electrification planning involves determining the combination of stand-alone systems, microgrids, and extensions of the electrical grid, along with their specific designs, to supply energy to a predefined set of consumers (Ciller and Lumbreras, 2020). To calculate the cost of a distribution infrastructure, whether a microgrid or a grid extension, it is necessary to solve both the generation sizing problem and the network design problem. However, a fundamental part of addressing these challenges is initially defining the number and location of consumers the distribution infrastructure will serve, which constitutes a significant clustering problem. Therefore, in this context, clustering is intrinsically linked with the challenges of generation sizing and network design (Ciller and Lumbreras, 2020).

The electrical distribution infrastructure for microgrids and grid extensions consists of a generation center usually located near the loads' geometrical center and power lines connecting the center to consumers. Within the framework of this research, the term "cluster" refers to this distribution infrastructure with LV lines.

The literature presents a vast array of clustering algorithms that employ diverse approaches and techniques, such as partitional (Swarndeep Saket and Pandya, 2016; Ikotun et al., 2022), hierarchical (Ran et al., 2023), density-based (Bhattacharjee and Mitra, 2021), and graph-based methods (Schaeffer, 2007; Aggarwal and Wang, 2010; Nascimento and De Carvalho, 2011). Each has strengths and weaknesses (Golalipour et al., 2021; Ezugwu et al., 2022). These algorithms find applications across a broad spectrum of domains, from network design, and transport analysis to biology, among others (Ghosal et al., 2020; Ezugwu et al., 2022; Chaudhry et al., 2023; Lenssen et al., 2023). The choice of a specific method hinges on factors like the data type and structure, clustering objectives, and available computational resources.

In rural electrification in low-income countries, it is crucial to account for certain characteristics. Scattered settlements and a low population density characterize rural regions. The frequent lack of reliable data and historical trends complicate accurate estimation and forecasting. Given these conditions, applying graph learning and similar advanced methods becomes challenging. These methods (Li et al., 2022, 2023; Sun et al., 2023a, 2023b) often require large volumes of high-quality data for effective implementation. Consequently, the scarcity of data in rural low-income areas limits their applicability to electrification planning.

On the other hand, methods discussed in (Xu and Tian, 2015; Ezugwu et al., 2021), such as those based on swarm intelligence, genetic algorithms, and models with neural networks and decision trees, are effective at characterizing each grouping. Nevertheless, they do not scale efficiently when applied to large data sets.

While many effective clustering methods are documented in academic literature, they may not entirely address the specific requirements of rural electrification planning. Often, these methods do not consider geographical barriers or restricted areas and key electrical and geospatial metrics. This can result in a grouping that does not align with the specific energy needs of the study region. The metrics in question include the maximum longitude from the generation/transformation center, serving as a proxy for voltage drops, and the power associated with each distribution infrastructure, which should comply with certain maximum and minimum thresholds.

Geographical barriers could encompass critical wildlife habitats, private properties, or other areas with stringent restrictions that distribution lines cannot cross or are unsuitable for situating energy production systems. By accounting for these factors, clustering algorithms can efficiently identify the most suitable electrification solutions, mitigating adverse environmental impacts and fostering sustainable development.

This research focuses on methods that address the clustering problem in large-scale techno-economic and geospatial rural electrification planning. The term "large-scale" is interpreted to encompass a region as large as an entire country, rather than continental scales. Small-scale methods and tools (Raj & Bhattacharyya, 2016, 2018; Shaikh et al., 2020, 2022a, 2022b, 2022c, 2023; Akbas et al., 2022; Ammari et al., 2022) are not included because the nature of the problem is different, and in most cases, there is no need to group consumers because an entire community or village will be electrified as a single system. In addition, the computational resources needed to solve a small-scale problem are reduced, which allows the use of classic optimization techniques or computationally intensive procedures that would fail in a large-scale problem.

In the present investigation, the aim is to ensure that the algorithms employed can effectively manage data sets of realistic sizes on standard personal computers, while also aiming to provide a feasible solution[1] within a reasonable time frame. The objective is to find a solution that is not only accurate and reliable from an energetic point of view, but also practical in terms of computational resources and time efficiency. Achieving a balance between computational efficiency and the accuracy of the solution is a crucial aspect of regional electrification planning.

2.1. Exploring clustering in models for large-scale techno-economic and geospatial planning of rural electrification

The literature review provides a comprehensive overview of various models focusing on techno-economic and geospatial planning for large-scale rural electrification. These models are often referred to as Least-Cost Electrification Models (LCEM). For a more in-depth understanding of LCEMs, refer to the works of (Morrissey, 2019; Ciller and Lumbreras, 2020). Among these models are IntiGIS I and II (Pinedo-Pascua, 2010; Romero Otero, 2016), the Reference Electrification Model (REM) (Ciller, 2021), Gisele (Vinicius et al., 2021; Corigliano, 2022), OnSSET detailed (Sahlberg, 2023), ONSSET light (Korkovelos, 2020), Renewable Energies for Rural Electrification of Africa (RE2NAF) (Szabó et al., 2013;

[1] Feasible solution refers to one that satisfies all the specified constraints and requirements of a given problem.

Moner-Girona et al., 2016), Mahapatra and Dasappa (2012), Van Ruijven et al. (2012), Dagnachew et al. (2017), Sahai (2013), RLI-GEPT (Bertheau et al., 2017; Blechinger et al., 2019), Abdul-Salam and Phimister (2016b, 2016a), Zeyringer et al. (2015), Network Planner (Kemausuor et al., 2014), Deichmann et al. (2011), Levin and Thomas (2012), Geospatial planning for rural electrification (GEOSIM) (Innovation Énergie Développement, 2021) and Banks et al. (2000).

These models evaluate rural electrification alternatives (individual systems, microgrids, extensions of the electrical grid) using the Levelized Cost of Electricity (LCOE) as a key metric. The alternative with the lowest LCOE is recommended, with the assignment of technologies being influenced by the granularity level of the model. The general formula of the LCOE is calculated using equation (1), as referenced in (Amador, 2000).

$$LCOE_{i,j} = \frac{CTA_{i,j}}{E_{i,j}} (monetary\ unit / kWh) \tag{1}$$

Where:

$LCOE_{i,j}$: is the LCOE corresponding to alternative j for node i.

$CTA_{i,j}$: total annualized cost (in monetary units) of alternative j for node i.

$E_{i,j}$: annual electrical energy produced (kWh) by alternative j for node i.

monetay unit: the type of currency used, such as USD, EUR, and CUP, among others.

A common practice in LCEMs is to group consumers based on predefined natural boundaries of communities or raster cells (Morrissey, 2019). This simplification reduces the model's computational complexity and eliminates the need for a clustering algorithm. However, it introduces challenges in estimating the cost of distribution infrastructure and can lead to inaccuracies in network design. Such inaccuracies can manifest as oversized or undersized networks, potentially incurring additional costs and causing delays in electrification projects. Furthermore, overlooking the specific energy demands of each consumer can result in inaccurate LCOE estimates.

These issues can compromise the model's capacity to evaluate different electrification alternatives effectively, potentially leading to unnecessary investments in low-demand areas and insufficient investment in high-demand areas. Therefore, addressing these limitations is crucial to enhance the effectiveness of LCEMs in planning rural electrification.

As Morrissey (2019) elucidates, the challenge of clustering in the context of LCEMs involves determining whether the demand density is enough to justify establishing a more extensive energy delivery system, such as a microgrid or an extension of the existing electrical grid. Consequently, the algorithm must identify which homes (or consumers) are close to each other to justify their inclusion in a cluster, and which are too distant or isolated, making clustering impractical.

Among the models mentioned above, only OnSSET detailed and REM include clustering algorithms that work at the consumer level. OnSSET detailed incorporates the DBSCAN algorithm, but it assumes that all houses have the same demand level (Sahlberg, 2023). REM, for instance, utilizes two algorithms: an exhaustive clustering algorithm and a top-down algorithm (Ciller, 2021). The "exhaustive" algorithm employs an agglomerative hierarchical approach to thoroughly explore the solution space and cluster customers based on cost considerations. Conversely, the top-down clustering algorithm, developed in collaboration with (Oladeji, 2018), calculates a power grid extension that connects all consumers and then assesses the cost-effectiveness of disconnecting certain elements and utilizing off-grid alternatives.

The clustering solutions in the REM model can be characterized as ad-hoc strategies. The decision-making algorithms, which determine whether to join or keep clusters separate, rely on cost comparisons from a representative set of generation designs. These strategies depend on the cost of the electrification systems being evaluated. However, it is important to note that the REM model encompasses a limited array of generation technologies. Therefore, applying REM's clustering solutions to other LCEMs remains uncertain. In light of these observations, exploring more flexible and adaptable clustering solutions that can cater to a wider range of LCEMs would be beneficial.

Concerning the tools, both models lack a GUI for inputs, making them difficult for non-experts to use. OnSSET requires coding skills in Python and REM relies on the commercial software MATLAB. The Gisele model, which includes DBSCAN and Hierarchical Cluster Analysis (HCA) algorithms, operates at a raster level and does not consider the Low Voltage (LV) distribution lines of individual households for simplification (Corigliano, 2022).

2.2. Application of clustering algorithms in rural electrification

In addition to their use in LCEM, clustering algorithms have also been used in some studies to group rural consumers into clusters for better planning and decision-making of rural electrification. However, literature on applying clustering techniques in rural electrification planning is scarce, with only a few references available. For instance (Leonard, 2022), employed the DBSCAN algorithm to cluster houses into "electricity communities" that could benefit from grid extension or an autonomous local grid system.

Certain studies focus solely on specific types of systems, such as microgrid formation (Cheong et al., 2017) stand-alone systems (Fletcher et al., 2017), grid extensions (Parreno Jr and Del Mundo, 2015), or a combination of microgrids and stand-alone systems (Rosenberg et al., 2022). As a result, they only address a partial aspect of the rural electrification problem. Moreover, most of these studies were conducted at a local level, leading to uncertainties regarding the adaptability of these algorithms to large-scale rural electrification challenges. Also (Parreno Jr and Del Mundo, 2015) present a specific heuristic for the problem where it is applied, which is not suitable for the models mentioned above.

3. Proposed method

This section presents a novel clustering algorithm specifically designed for techno-economic rural electrification planning. It also describes the implementation and integration of the proposed clustering method into the IntiGIS model, emphasizing the enhancements made to support more precise and efficient electrification planning.

3.1. The proposed geospatial clustering algorithm

This section describes the RElect_MGEC algorithm (Rural Electrification Microgrid and Grid Extension Clustering). Next, fundamental definitions are provided in section 3.1.1, followed by a step-by-step description of the proposed algorithm in section 3.1.2.

3.1.1. Definitions and notations

A tree data structure represents the electrical infrastructure associated with the clusters. Next, we define and describe a set of characteristics of the tree data structure used by the proposed algorithm.

Given a tree T (see Fig. 1), which has a set of nodes V representing consumers, and edges E corresponding to the power lines connecting the nodes:

- Each node V is associated with an ID (identifier) and a power value. Fig. 1a shows an example with nodes represented in orange, each labeled with its respective ID. Fig. 1b displays each node labeled with its power value.
- The edges E refer to segments of power distribution lines that supply energy to the consumers. In Fig. 1a, the edges are depicted as solid black lines and labeled with their lengths.

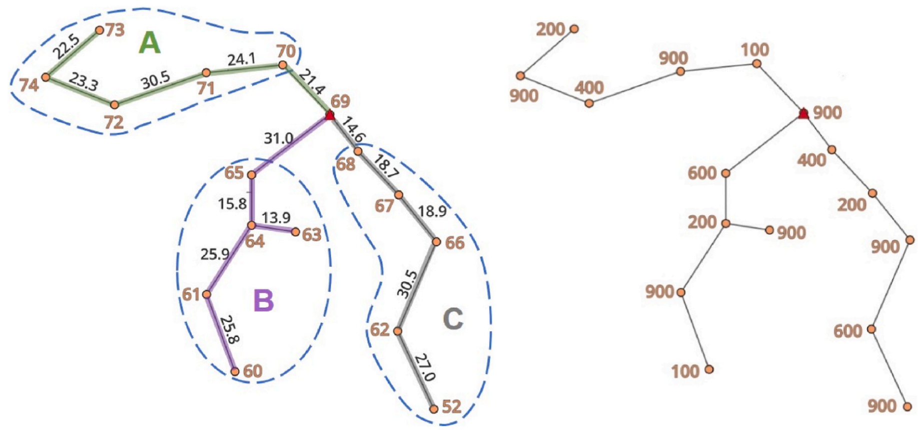

Fig. 1. Example data structure of a T tree representing a cluster.

- L_{LV}: Total length of the edges in the tree. It also represents the total length of Low Voltage (LV) lines in the distribution infrastructure.
- Branch: outgoing path from the central node (root) that traverses one of the subtrees[2] (children) of the central node. The distance of a branch is assumed to be the length of this path. The subtrees are outlined in blue in Fig. 1a, where three branches (A, B, and C) highlighted in green, purple, and gray, respectively, are observed. In this context, the number of branches of a tree corresponds to the degree of the central node.
- MaxLongC: length of the longest branch of the tree. Branch A in Fig. 1a.
- Center: node (highlighted in red in Fig. 1) representing the generation center of a microgrid or the transformation center of a main electrical grid extension. The consumer with the smallest MaxLongC is selected as the central node.
- P_{agr}: aggregated power of the tree measured in Watts (W). It is obtained from equation (2), by summing the power values of the nodes (consumers) in the tree (cluster C_i).

$$P_{agr,C_i} = \sum_{i}^{n} P_{h_i} \tag{2}$$

Where, P_{h_i} represents the contracted power or energy (W) for a consumer h_i.

The Minimum Spanning Tree (MST) was chosen to connect the set of nodes, whose effectiveness has been demonstrated in electrical network planning (Liao et al., 2020). Furthermore, several regional planning tools estimate the network cost by applying methods based on MST calculation (Levin and Thomas, 2012; Abdul-Salam and Phimister, 2016b; World Bank Group, 2016; Blechinger et al., 2019; Ciller et al., 2021; Sahlberg, 2023). In this context, the MST represents the connections between consumers so that all are connected directly or indirectly through other consumers.

3.1.2. The RElect_MGEC clustering algorithm

The clustering algorithm proposed in this study is Rural Electrification Microgrid and Grid Extension Clustering (RElect_MGEC). This algorithm involves three phases: exploratory clustering, evaluation of potential clusters, and generation of the results, as visualized in Fig. 2. The outcomes of each phase are illustrated in Fig. 4.

The inputs for RElect_MGEC consist of a map of consumers without electricity and their respective power consumption and a map of sensitive areas or barriers (optional). As well as the parameters: MaxLongC, PotMax, MinCons, and Eps defined below. This algorithm adopts the approach of constrained clustering, a semi-supervised clustering method, to group data while incorporating domain knowledge in the form of constraints (Qin et al., 2019). With this approach, we include the following parameters for cluster formation.

- MinCons: parameter that acts as a constraint to ensure that a cluster has a minimum number of consumers to be considered valid.
- Eps: search radius in which the MinCons must be found.
- MaxLongC: this parameter corresponds to the property of the same name defined previously. It is calculated using **Algorithm A.1.** MaxLongC sets a limit on the length of the network to mitigate energy losses through distribution. The algorithm utilizes auxiliary functions such as "center_Graph" (**Algorithm A.2**) and "node_MaxLongCGen" (**Algorithm A.3**) to calculate the center of the graph and the MaxLongC, respectively.
- PotMax: maximum power of the cluster (W). The value restricts the P_{agr} of a cluster, ensuring that the generated electricity meets the demand without exceeding the capacity of the infrastructure.

In this context, the algorithm's objective is to ascertain the number of clusters within a specified study area that satisfy the constraint conditions of PotMax, MaxLongC, and MinCons. Additionally, the algorithm can incorporate geographical barriers as optional constraint conditions, enhancing its applicability.

In Phase 1, the exploratory clustering phase, consumers with power consumption greater than or equal to the PotMax are added to a list of isolated consumers. Then, consumers with power consumption less than PotMax are selected from the map, and DBSCAN clustering (Schubert et al., 2017) is applied to identify high-density areas. Noise points that do not meet the minimum number of consumers required to form a cluster are also added to the list of isolated consumers.

In Phase 2, the potential clusters resulting from the DBSCAN runs are evaluated. For each potential cluster, the consumers' Delaunay Triangulation (DT) is computed (refer to Fig. 4C). If a sensitive area or barrier map is provided, edges that intersect with these areas or are longer than the MaxLongC are removed. A weighted graph is constructed. Each edge of the weighted graph is assigned a weight derived from equation (3), where the weights represents the efficiency of the connection between two consumers. The nodes, representing the consumers, are assigned weights equivalent to their power.

[2] In a tree data structure, each child of a node forms a subtree, so a subtree is any tree generated from a specific section of another tree.

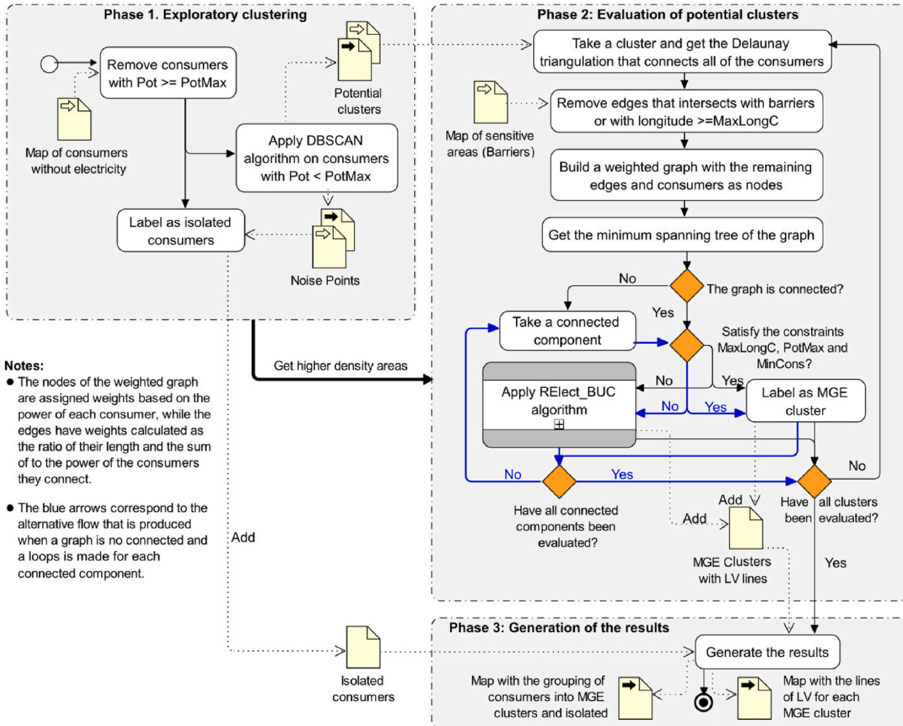

Fig. 2. Flowchart of the RElect_MGEC clustering algorithm.

$$w = \frac{L}{P_1 + P_2} \tag{3}$$

Where:

L: represent the longitude of the edge.

$P_1 y\, P_2$: represent the powers of the nodes (consumers) that the edge connects.

This algorithm is built on the principle that a distribution infrastructure is more efficient by connecting consumers with higher power and closer proximity. This approach minimizes the costs associated with the length of the connecting lines and reduces energy losses.

The weighted graph is constructed to form a Minimum Spanning Tree[3] MST). The MST is designed to connect a set of consumers in a way that minimizes the cost/benefit ratio (Length/Power), thereby enhancing the overall efficiency of the distribution system. The MST is obtained from the weighted graph using the Kruskal algorithm. The Kruskal algorithm is chosen for its proven efficiency in solving real-world problems that involve sparse and potentially disconnected graphs, a common characteristic of rural settlements.

If the graph is connected and satisfies the PotMax, MaxLongC, and MinCons constraints, it is added to the list of MGE clusters. Otherwise, the same evaluation is applied for each connected component, and if it satisfies the requirements, it is also added to the list of clusters. In cases where the restrictions are not met, the Rural Electrification Bottom-up Clustering (RElect_BUC) algorithm is used to continue partitioning consumers. Finally, consumers not included in the list of MGE clusters are labeled as isolated.

RElect_BUC, as shown in Fig. 3, utilizes graph theory to execute agglomerative clustering to ensure that the resulting clusters meet the established metrics. First, all graph edges are sorted by weight, as indicated in equation (3), from smallest to largest. A new graph T′ is created where all edges are initially deactivated. In this context, the edges correspond to clustering decisions. An edge can be activated,

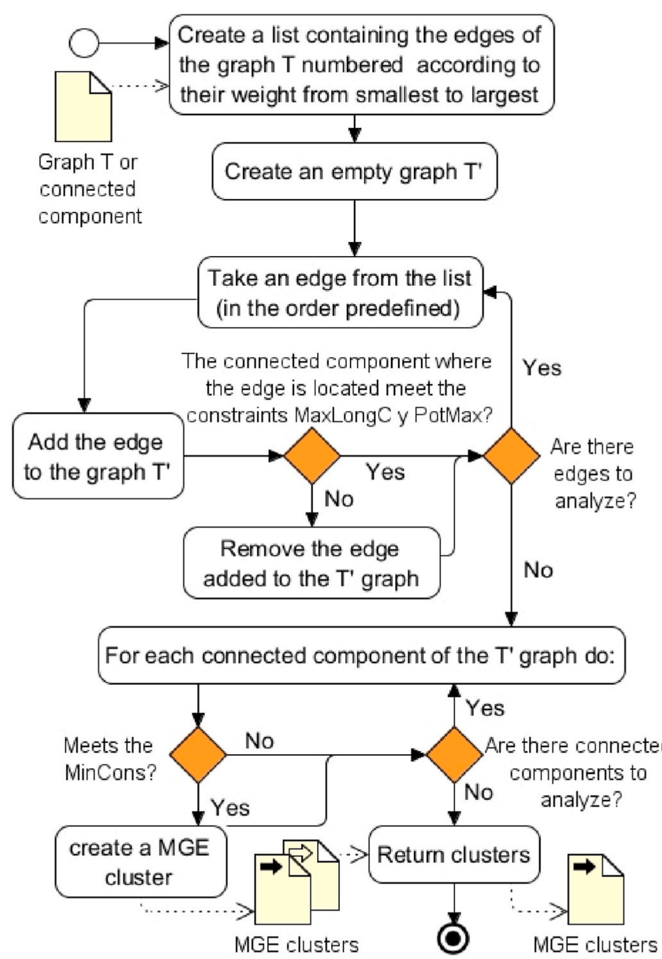

Fig. 3. Flowchart of the RElect_MGEC clustering algorithm.

[3] Generated with Kruskal's algorithm using the "minimum_spanning_tree" function available in NetworkX library.

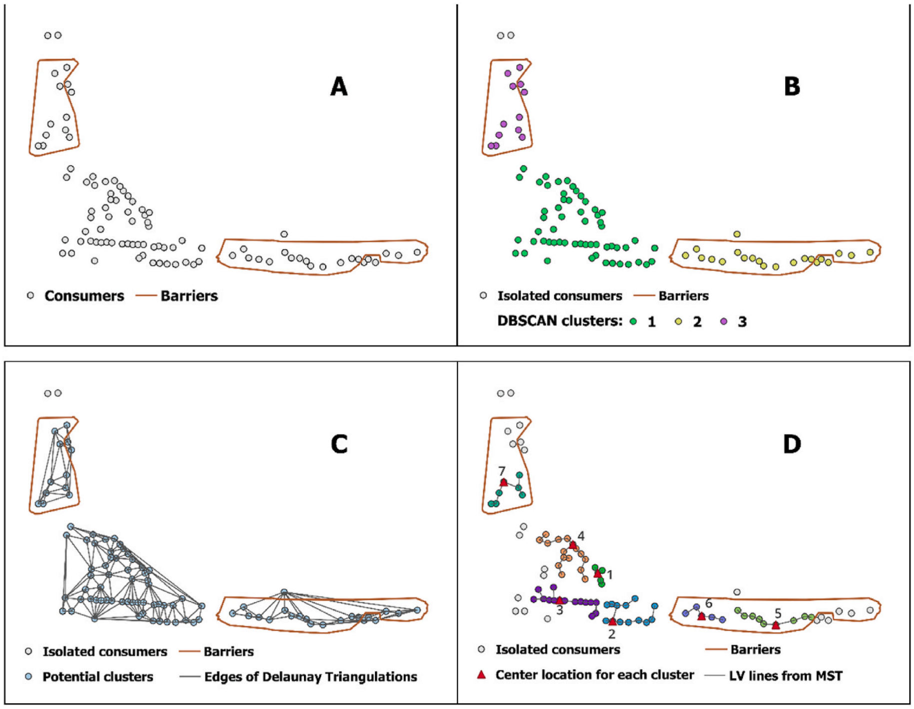

Fig. 4. Example of distribution infrastructure using the RElect_MGEC clustering algorithm. A) Location of consumers. B) Results after phase 1: exploratory clustering. C) Delaunay Triangulation obtained for each potential cluster. D) Final results include 7 MGE clusters, locations of generation/transformation centers, and LV distribution lines that connect consumers within each cluster through the MST.

merging the consumers at both ends into a single cluster, or remain deactivated, keeping them in separate clusters. This decision hinges on whether they fulfill the defined constraints for maximum length and maximum power. Finally, once all the edges have been considered, the series of interconnected consumers that meet the constraints of MaxLongC, PotMax, and MinCons form microgrids and grid extension clusters.

Phase 3 involves generating the results. The clusters of consumers obtained are saved as a map in.shp format. Each MGE cluster is associated with attributes such as ID, Size, Total Power, MaxLongC, and Total LV lines. Additionally, the low voltage lines required to connect the consumers of each cluster to the generation center are identified and saved in a separate map. The attributes for each line segment include ID, length, and the ID of consumers it connects. Both maps (Fig. 4D) are returned as output of the algorithm in.shp format for ease of use and accessibility.

Finally, Fig. 5 displays the interface of the "Clustering of Consumers" functionality, which includes input fields for the parameter values. This feature is part of the upgrades made to the LECGIS plugin described in (Torres-Pérez et al., 2021), which was also updated to work with version 3.x of QGIS. It was implemented using PyQGIS (QGIS Development Team, 2023), Geospatial Data Abstraction Software Library (GDAL/OGR) (Rouault et al., 2023), and the NetworkX library (Hagberg et al., 2008; NetworkX Developers, 2020).

3.2. IntiGIS model

The IntiGIS model facilitates the assessment of various technological options for electrifying rural areas that lack access to this service. Developed by CIEMAT in Spain, IntiGIS has been successfully applied in several countries, demostrating its flexibility and adaptability to diverse scenarios.

Fig. 6 outlines the progression of IntiGIS, divided into three distinct phases (the dates provided are estimates). The initial phase is linked to SolarGIS (Mahmmud et al., 1996; Monteiro et al., 1998; Vandenbergh et al., 1999) and its successor SolarGIS II (Amador, 2000; Amador and Domínguez, 2005). The second phase saw the development of IntiGIS I, as discussed in (Domínguez Bravo et al., 2008; Pinedo-Pascua, 2010; Pons et al., 2013; Martínez Sarmiento et al., 2014) and IntiGIS II, highlighted in (Page Arias, 2015; Romero Otero, 2016) emerged.

The IntiGIS I and II versions, by operating at the raster cell level, exhibit the same deficiency as LCEMs that group consumers based on predefined natural boundaries of communities or raster cells. Consequently, this model is a suitable candidate for evaluating the suitability of the RElect_MGEC algorithm.

The third and current phase began with the research conducted by (Torres-Pérez et al., 2019, 2021), which led to significant enhancements to the model. For a detailed account of this progression refer to (Torres-Pérez et al., 2021). Unlike its predecessors, the tool's latest version was developed using the QGIS free software environment (QGIS Development Team, 2022). This version features the ability to calculate and compare the LCOE of seven different electrification alternatives: stand-alone (powered by photovoltaic, wind, or diesel), microgrid (powered by diesel, wind-diesel, or photovoltaic-diesel), and grid extension. It also allows it to operate at the consumer level, assigning varying demand and power values and grouping them into microgrid clusters and grid extensions (Torres-Pérez et al., 2024).

The updated version of the model is based on three components: C1, responsible for territorial ordering analysis; C2, which handles geospatial clustering; and C3, which conducts a technical-economic analysis. Fig. 7 shows an overview of the new version focusing on component 2, geospatial clustering, and the interrelation of this with the C1 and C2.

C1 can be employed to conduct an analysis focused on territorial planning, aiming to identify sensitive areas (non-viable) for placing energy production systems, using María Rodríguez's methodology as presented in (Torres-Pérez et al., 2022). In this regard, houses (consumers) located in areas where land use restrictions apply would be initially excluded from implementing centralized systems (microgrids and grid extensions). As a result, it is possible to deselect consumers from the map of unelectrified consumers that intersect with these areas,

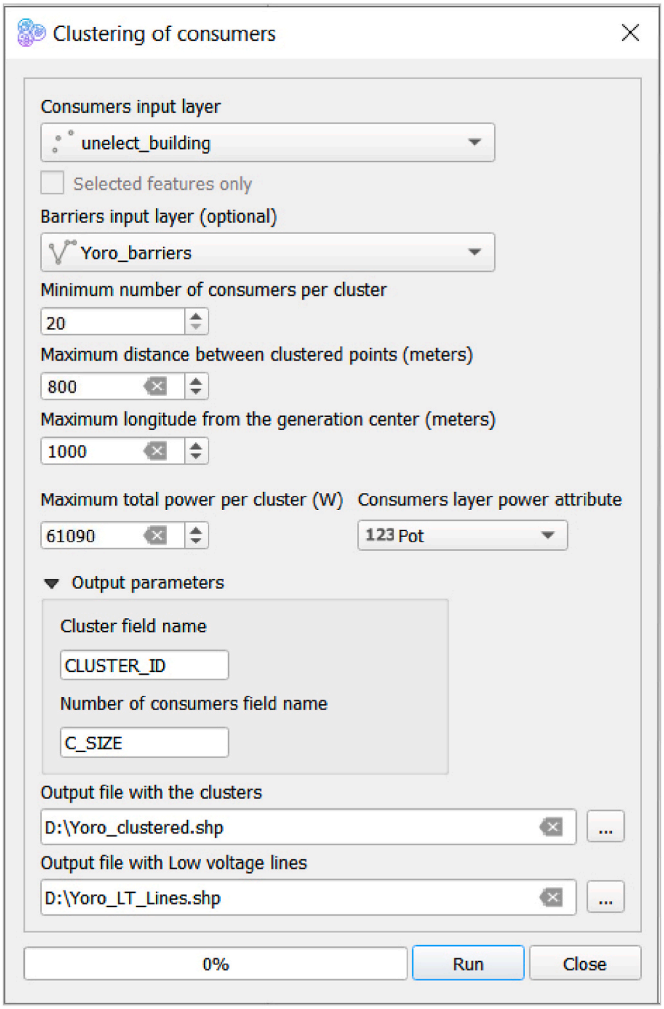

Fig. 5. Interface of functionality "Clustering of consumers".

which serves as input for component 2. Furthermore, within component 3, stand-alone alternatives would only be evaluated for these households.

The diagram starts with component 1, which can also be used to determine barriers (obstacles) that the layout of low voltage distribution lines cannot cross. Some potential obstacles may include critical wildlife habitats, private properties, or other areas with strict constraints. A way to generate these barriers is by using the functionality to determine sensitive areas of the ExamZonas plugin, as described in (Torres-Pérez et al., 2019). Or another way, for example, is simply using the municipal limits or other administrative borders as barriers. The barriers can be utilized as input for Component 2 to support the clustering of consumers, providing the user with various options to shape the formation of clusters.

Component 2 handles the grouping of consumers into microgrid clusters, grid extensions, and stand-alone systems. This process involves executing the RElect_MGEC algorithm, as outlined in section 3.1. The clustering outcome produced by Component 2 serves as input for Component 3. This component conducts a technical and economic evaluation of electrification alternatives utilizing the "Techno-Economic Analysis" feature of the LECGIS plugin, as detailed in (Torres-Pérez et al., 2021).

4. Application and results discussion

In this section, we analyze and contrast the results and performance of the new RElect_MGEC algorithm with well-recognized algorithms in the scientific literature applied in rural electrification.

Table 1 summarizes the characteristics of the experimental design. Seven clustering variants (from V1 to V7) were employed (as indicated in Table 3). Eleven experiments (E1 to E11) were carried out in each scenario, combining the variants with different input parameters. Each variant was implemented within the model's C2. The outcome of each clustering variant (C2) serves as input for component 3 (C3), enabling us to evaluate how the clustering approach affects the quality of alternative assessments.

The study evaluated four electrification alternatives: stand-alone photovoltaic, microgrids (diesel and photovoltaic-diesel), and grid connection. The techno-economic parameters used to define these alternatives were based on the reference values established in the original case study conducted by (Quevedo Saldias, 2022). However, the solar radiation values were obtained from the Global Irradiation at optimum Tilt (GTI) map (Solargis, 2019).

To compare the different experiments, we considered the following metrics: MinCons or Number of Consumers, MaxLongC, and PotMax. A cluster is considered valid if it satisfies these metrics. In this context, the number of valid clusters is a variable that ensures compliance with MinCons, MaxLongC, and PotMax. Finally, it was decided to measure three key variables for each experiment: the number of valid clusters (NCV), execution time, and the Annualized Total Cost (ATC). The ATC is calculated using equation (5).

It should be noted that the time for C3 includes the calculations of LCOE for the study area and the generation of the final PDF report for the study area. However, it does not encompass the calculations of certain parameters required for the calculations, such as the distance to the existing electrical grid for each consumer and the global radiation. The community/cluster-type PDF report was also not calculated for these

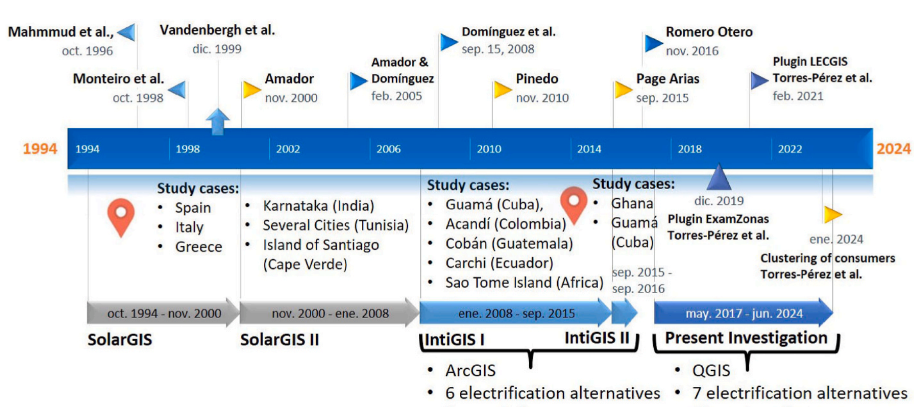

Fig. 6. Progression of the IntiGIS model.

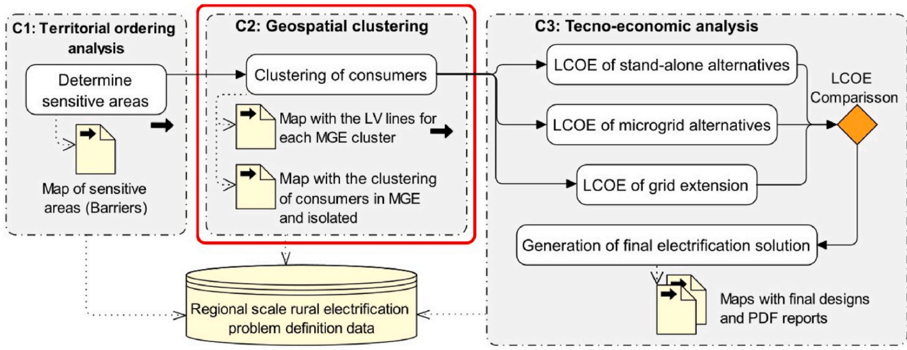

Fig. 7. General diagram of the new version of the IntiGIS model with a focus on the Geospatial Clustering Component (C2).

Table 1
Characteristics of the experimental design.

Inputs	Analysis and Recommendations			Outputs
Consumers	C2: Geospatial Clustering		C3: Techno-economic Analysis	Electrification Solutions
	Methods	Experiments		
10,995	V1	E1	Electrification	S1
		E2	Alternatives:	S2
		E3	1 Grid extension.	S3
	V2	E4	2 Photovoltaic	S4
	V3	E5	(stand-alone).	S5
	V4	E6	3 Diesel	S6
		E7	(microgrid).	S7
	V5	E8	4 Photovoltaic-	S8
	V6	E9	Diesel	S9
		E10	(microgrid).	S10
	V7	E11		S11

Table 2
Demands levels in scenario 2.

Level	Demand (Wh/day)	Power (W)	Frequency (%)	Nº of cons.
1.	725	100	30	3299
2.	1000	200	50	5497
3.	3400	800	15	1649
4.	8200	2000	5	550

experiments.

We executed the experiments in a computer with Intel(R) Core (TM) i7-8750H CPU @ 2.20 GHz, NVIDIA GeForce GTX 1050 GPU with 4 GB of DRAM and 8 GB RAM.

4.1. Description of the data and test scenarios

To evaluate the effectiveness of the proposed approach, we used a sample dataset corresponding to the case study conducted by (Quevedo Saldias, 2022). The dataset includes 10,995 households without electricity in Sulaco, Victoria, and Yorito municipalities, located in the southwest region of the Yoro department in Honduras (visualized in Fig. 8). The study area spans over 1,155,686 km^2.

Two scenarios were considered for the experiments. In Scenario 1, all consumers in the dataset were assigned a uniform daily demand of 725 Wh/day and a power of 100 W. Scenario 2 utilized the same consumer locations as Scenario 1, but introduced random variations in demand levels following the distribution in Table 2.

Table 3
Clustering variants for experiments.

Variant/ Algorithm	Description	Objective
V1/ RElect_MGEC	Implements RElect_MGEC.	Test the RElect_MGEC algorithm.
V2/RElect_BUC	Directly applies RElect_BUC without using the DBSCAN algorithm in Phase 1. Starts by calculating the DT that connects all consumers in the area.	Test this variant and compare it with V1, that use DBSCAN in Phase 1.
V3/DBSCAN	A variation of the original RElect_MGEC, where RElect_BUC is not used for post-processing.	Test DBSCAN algorithm and compare this variant with V1, that use RElect_BUC for post-processing.
V4/K-means + RElect_BUC	Similar to the original RElect_MGEC, but replaces the DBSCAN algorithm with K-means in Phase 1.	Test the combination of K-means and RElect_BUC.
V5/K-means	Similar to V4, but does not use RElect_BUC for post-processing.	To test the K-means algorithm (Ahmed et al., 2020).
V6/HCA + RElect_BUC	Similar to the original RElect_MGEC, but replaces the DBSCAN algorithm with a single-link Hierarchical Clustering Algorithm (HCA) in Phase 1. Uses the number of clusters as a stopping rule and Euclidean distance to calculate the linear distance between two points.	To test the combination of HCA and RElect_BUC
V7/HCA	Similar to V6, but does not use RElect_BUC for post-processing.	To test the Hierarchical Clustering Algorithm (HCA)

4.1.1. Algorithms for comparison and established parameters

The selection of algorithms for comparison was based on a review of the state-of-the-art clustering techniques applied in the context of rural electrification, as discussed in section 2. The DBSCAN, K-means[4] and HCA[5] algorithms were chosen because they are well-established methods commonly used in this field.

To demonstrate the effectiveness of our proposal compared to other clustering algorithms, we designed seven variants, as detailed in Table 3. Each variant was integrated into the C2 component of the model (refer to Fig. 2 for a visual representation of the RElect_MGEC algorithm). For instance, variant V4 involves substituting the DBSCAN algorithm with K-

[4] We use the DBSCAN and K-means implementations available in QGIS algorithms.

[5] We use the HCA implementations available in the QGIS plugin Jenkner, J. (2020). *Cluster Points* Retrieved from https://jjenkner.com/ClusterPoints/(28 June 2023).

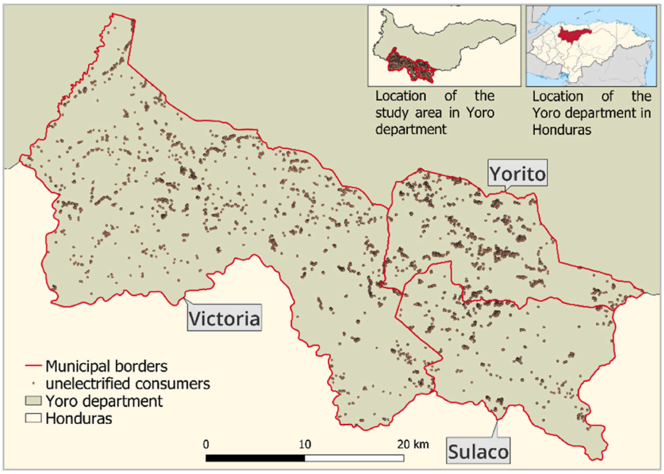

Fig. 8. Location of consumers in the study area.

means during Phase 1 of the diagram in Fig. 2.

The Lukes graph partitioning algorithm (Lukes, 1974) was initially implemented but later discarded due to its long execution time. For variants 4 to 7, when the algorithm in question (K-means or HCA) is applied in the first phase, the clusters obtained that meet the MinCons requirement will be considered as potential and evaluated in Phase 2. Meanwhile, those not meeting the MinCons requirement will be marked as noise and added to the list of isolated consumers.

The input parameters were established as follows: MinCons = 20, Eps = 800 m, MaxLongC = 1000 m, and PotMax = 61090 W, based on the results of the REM Model in the case study (Quevedo Saldias, 2022) and expert recommendations. The consumer map used for the experiment included 10,995 households, depicted in Fig. 8, with municipal borders as the map of barriers. Fig. 5 displays the parameters for the experiment 1 (E1).

Multiple executions of the experiments that implement the V1 and V3 variants that utilize DBSCAN in the initial phase were conducted to determine the value of the Eps parameter in both scenarios. The goal was to identify the value that yielded the highest number of valid clusters. Additionally, the "k-nearest neighbor distance" method (refer to Fig. 9) was employed, as suggested in (Schubert et al., 2017) to establish an appropriate Eps value for the DBSCAN algorithm. Moreover, the E3 experiment evaluated the V1 variant using the optimal Eps value obtained for E5 (V3).

In scenario 1, the NC parameter for E6 (V4) and E9 (V6) was determined by dividing the total power load by PotMax. Each household in scenario 1 has a power of 100 W, resulting in a total power of 1,099,500 W for the study area. By dividing this total power by PotMax

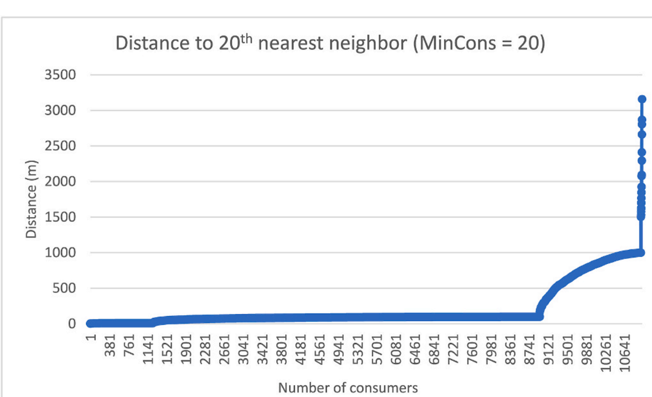

Fig. 9. Sorted 20th nearest neighbor distance plot for the dataset.

(61,090 W), a minimum of 18 clusters was obtained based on peak power considerations.

In scenario 2, the total power of the study area increased to 3,848,500 W. Dividing this total by PotMax would result in a minimum of 63 clusters for E6 and E9. However, if the clusters obtained in the first stage do not meet certain constraints, such as the MaxLongC distance constraints, further subdivision is performed using the RElect-BUC algorithm.

To determine the value of the NC parameter for experiments E8 (V5) and E11 (V7), an iterative process was carried out until the value that maximized the NCV for each experiment was found. Consequently, experiments E7 and E10 were performed to evaluate V4 and V6 variants, with the NC values found optimal for E8 (V5) and E11 (V7), respectively.

4.2. Execution of the experiments and analysis of the metrics

A total of 11 experiments were conducted for each scenario to demonstrate the algorithm's behavior under different parameter variations. The experiments were carried out by implementing our model and the clustering variants in an environment that utilizes QGIS libraries and modules in Jupyter Notebook (William, 2019).

4.2.1. Application of component 2: geospatial clustering

Tables 10 and 11 provide descriptive statistics of the clusters obtained after applying each experiment in Component C2 of the model. In Scenario 1, Experiment E2 (V1) yielded the highest NCV with 194, followed by E4 (V2) with 193, and E1 (V1) and E9 (V6) with 192. In Scenario 2, E1 (V1) and E2 (V1) produced the highest NCV with 197, followed by E4 (V2) with 195 and E9 (V6) with 192 (see Table 12).

It was observed that among the variants utilizing the same algorithm in the first phase, those that incorporated the RElect_BUC algorithm during post-processing generated a higher NCV than those that did not.

The input parameters specified for experiments E2 (V1), E6 (V4), and E9 (V6) enable the creation of denser and larger clusters in the first phase, which are then further processed by the RElect_BUC algorithm in the second phase. This demonstrates the effectiveness of the RElect_BUC algorithm in partitioning large clusters and converting them into valid clusters.

It is worth noting that experiments E3 (V1), E7 (V4), and E10 (V6), which were executed with the most suitable parameters for DBSCAN, K-means, and HCA, respectively, resulted in a smaller number of valid clusters compared to their preceding experiments that use the same variants. This can be attributed to the generation of smaller clusters in the first phase.

4.2.2. Application of component 3: Techno-Economic Analysis

After executing Component 2 with each clustering experiment (E1 to E11), 11 Map_{Cr} outputs were generated for each scenario. These Map_{Cr} outputs were subsequently used as inputs for Component 3 to evaluate four electrification alternatives (refer to Table 1).

Appendix C and Appendix D provide descriptive statistics for the electrification solutions (S1 to S11) obtained in each scenario. The weighted average LCOE for each alternative j was calculated using equation (4), which utilizes the energy generated by each system i as a weighting factor. This approach ensures that the average LCOE of a system accurately reflects the demand it meets. The ATC is derived using equation (5), formulated by rearranging the general LCOE equation (1).

$$Weighted\_Avg\_\_LCOE_j \frac{= \sum_{i=1}^{n}(\text{LCOE}_i \ * \ \text{Energy}_i)}{\sum_{i=1}^{n} \text{Energy}_i} \tag{4}$$

$$ATC_j = \sum_{i=1}^{n}(\text{LCOE}_i \ * \ \text{Energy}_i) \tag{5}$$

Where:

ATC_j: represents the sum of the annualized total cost of the consumers electrified with the electrification alternative j.

$LCOE_i$: corresponds to the LCOE for system i expressed in USD/kWh.

$Energy_i$: refers to the energy produced by system i.

n: represents the total number of installations of the electrification alternative j.

The charts shown in Figs. 10 and 11, illustrate the relationship between the ATC and the execution time (an average of several executions) for each clustering experiment. In the execution time graph, C2 refers to Component 2, which is responsible for performing the clustering, and C3 is the component responsible for calculating the LCOE.

In previous versions of IntiGIS based on raster data, centralized systems were computed for each cell. However, with the introduction of clustering in the new version, the computation time of Component 3 is reduced, as centralized systems are now calculated exclusively for clusters considered valid. This adjustment partially compensates for the time spent generating the clusters within C2.

4.3. Statistical analysis of the overall performance of the algorithms

Table 4 contains the NCV, ATC, and execution time for each clustering experiment, averaged over multiple runs.

Across the 11 experiments, the Spearman correlation coefficient between NCV and ATC was −0.982 in Scenario 1 and -0.989 in Scenario 2. These values indicate a very strong negative correlation between the two variables, suggesting that as NCV increases, ATC tends to decrease. Additionally, the two-tailed significance (p-value) is 0.00086 in Scenario 1 and 0.0001 in Scenario 2, indicating a very low probability that the observed correlation is due to chance.

Considering this finding, a combined metric was calculated to evaluate the overall performance of the algorithms. This research aims to minimize both cost and execution time. However, these objectives can conflict, as a faster algorithm may result in a more costly solution, and vice versa.

To address this challenge, the "Sum of Ranks" method, an adaptation of the "Weighted Sum" approach for multi-objective optimization, was adopted (Eichfelder, 2021). In this method, both factors, ATC and execution time, are equally weighted, reflecting their equal importance in evaluating algorithm performance. The sum of ranks is calculated by summing the normalized values of ATC and execution time for each algorithm, according to equation (6). The Min-Max normalization method was employed to transform the cost and execution time values to a common scale. The algorithm with the lowest sum of ranks is considered the most efficient, as it offers the best balance between ATC and execution time.

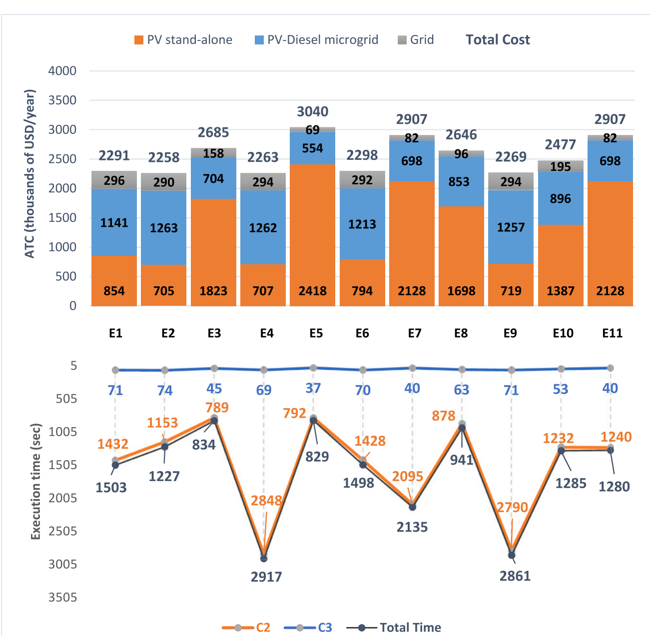

Fig. 10. Plot of ATC vs Execution time for scenario 1.

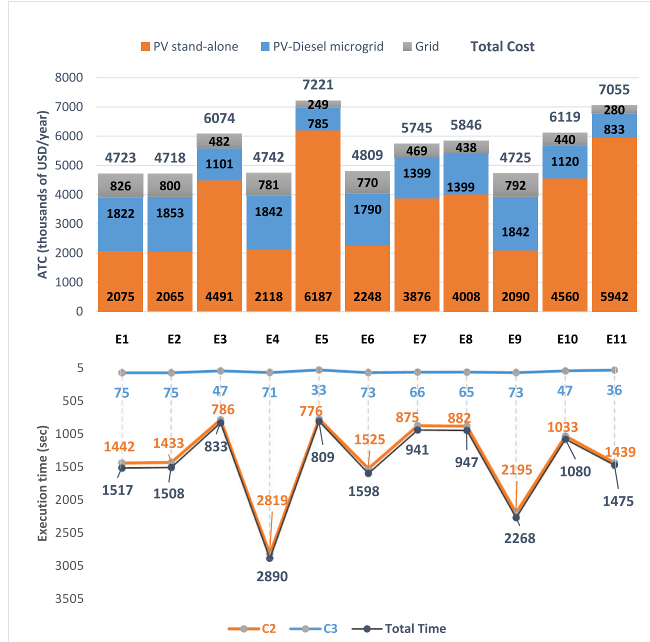

Fig. 11. Plot of ATC vs Execution time for scenario 2.

$$\text{Sum of Ranks} = \text{Normalized ATC} + \text{Normalized Time} \quad (6)$$

The analysis of Table 5 reveals that experiments E2 and E1, which employ the RElect_MGEC algorithm, achieved the lowest sum of ranks in both scenarios, indicating the best balance between cost and execution time. These results demonstrate the proposed method's ability to achieve lower costs within a shorter period.

This finding is corroborated by the tables in Appendix C and Appendix D, which contain statistical summaries of the electrification solutions. The "Total" column in these tables shows that as the "NCV" increases in an electrification solution, profitability also increases, evidenced by the decrease in the totals for "ATC" and the "Weighted average LCOE".

4.3.1. Statistical analysis to determine significant differences in the variables NCV, ATC, and execution time

The RElect_MGEC algorithm, corresponding to Experiment E1, demonstrated the best performance in both scenarios (see Table 5). The objective of the statistical analysis is to determine whether the results of this algorithm in terms of NCV, ATC, and execution time are statistically significant compared to other algorithms.

Several experiments were conducted on some variants with different parameters. From these variants, the experiment with the best overall performance was selected. The selected experiments are presented in Table 4, highlighted in blue. Since there are 7 algorithms, 6 pairs are generated for comparison with the algorithm of interest: V1 vs. V2, V1 vs. V3, V1 vs. V4, V1 vs. V5, V1 vs. V6, and V1 vs. V7.

The comparative evaluation of the algorithms was conducted using established statistical tests and IBM SPSS Statistics version 25. A significance level of 0.05 and a 99% confidence interval were employed. The following procedure was defined.

1. Conduct a normality test using the Shapiro-Wilk test, considering that the sample size $n \leq 50$.

Table 4
Experiments conducted.

Experiments	E1. V1	E2. V1	E3. V1	E4. V2	E5. V3	E6. V4	E7. V4	E8. V5	E9. V6	E10. V6	E11. V7
Scenario 1											
NCV	179	194	123	193	98	188	112	175	192	146	112
ATC (thousands of USD/Year)	2291	2258	2685	2263	3040	2298	2907	2646	2269	2477	2907
Execution time (sec.)	1503	1227	834	2917	829	1498	2135	941	2861	1285	1280
Scenario 2											
NCV	197	197	125	195	90	190	175	173	196	126	98
ATC (thousands of USD/Year)	4723	4718	6074	4742	7221	4809	5745	5846	4725	6119	7055
Execution time (sec.)	1517	1508	833	2890	809	1598	941	947	2268	1080	1475

Table 5
Experiments ranked by overall performance.

Scenario 1											
Experiments	**E2**	**E1**	**E6**	**E10**	**E3**	**E8**	**E9**	**E5**	**E4**	**E11**	**E7**
Sum of Ranks	0.19	0.365	0.37	0.50	0.548	0.549	0.99	1.0	1.01	1.05	1.46
Scenario 2											
Experiments	**E2**	**E1**	**E6**	**E7**	**E8**	**E3**	**E10**	**E9**	**E5**	**E4**	**E11**
Sum of Ranks	0.336	0.342	0.42	0.47	0.52	0.55	0.69	0.70	1.0	1.01	1.25

2. If the samples follow a normal distribution, apply parametric tests using the Student's t-test for two related samples.
3. If the samples do not follow a normal distribution, apply non-parametric tests using the Friedman test for n related samples.
4. If the Friedman test indicates no significant differences among the n related samples, it is concluded that there are no significant differences.
5. If the Friedman test reveals significant differences among the n related samples, conduct post-hoc tests using the Wilcoxon test for two related samples.

4.3.2. Normality test

Table 6 contains the normality test results using the Shapiro-Wilk test for each variable in both scenarios. The variable "Execution Time" in scenario 2 follows a normal distribution, as its p-value of 0.517 is greater than 0.05. Therefore, the null hypothesis of normality is not rejected for this variable. Consequently, parametric tests will be applied using the Student's t-test for related samples.

The "Execution Time" variable in scenario 1 and the NCV and ATC variables in both scenarios obtained p-values less than 0.05, indicating that they do not follow a normal distribution. Non-parametric tests using the Friedman test for n related samples will be applied to these variables.

4.3.3. Friedman test for n related samples

Table 7 presents the results of the Friedman test for the variables Execution Time (Scenario 1), NCV, and ATC in both scenarios. The p-values are notably small: 4.5×10^{-8}, indicating significant differences between the related samples for these variables.

4.3.4. Post-hoc tests using the Wilcoxon test for paired samples

The Wilcoxon rank-sum test is a non-parametric statistical method frequently used to assess significant differences and distribution patterns in clustering outcomes (Ran et al., 2021). This test compares the medians of two related groups to determine if they differ statistically significantly.

The significance level was adjusted using the Bonferroni correction to control the false positive rate during the six post-hoc tests (Rubin, 2021). This correction involves dividing the original significance level 0.05 by the number of tests. In this case, the adjusted significance level is $0.05/6 = 0.008$. Thus, for a result to be considered statistically significant, the p-value obtained in each test must be less than 0.008.

Table 8 presents the results of the Wilcoxon test for the variables "Execution Time" (Scenario 1), "NCV," and "ATC" in both scenarios. The p-values are below 0.008, suggesting that the proposed RElect_MGEC algorithm exhibits statistically significant performance compared to the other algorithms for these variables.

4.3.5. T-student test for two related samples: "execution time"

Table 9 presents the results of the T-Student test for the "Execution Time" variable in Scenario 2. The p-values are below 0.008, indicating a statistically significant difference in execution time between the proposed RElect_MGEC algorithm and the others.

5. Conclusions

This study demonstrates that conventional clustering algorithms like DBSCAN, K-means, and HCA are insufficient for identifying microgrid and grid-extension clusters when considering electrical constraints and geographical barriers. The RElect_MGEC algorithm overcomes these limitations by ensuring that the final clusters meet all given constraints, thereby facilitating the development of appropriate distributed infrastructures. The primary findings of this research encompass.

Table 7
Results of the Friedman test.

	NCV	ATC	Execution Time
Scenario 1	4.5×10^{-8}	4.5×10^{-8}	4.5×10^{-8}
Scenario 2	4.5×10^{-8}	4.5×10^{-8}	–

Table 6
Significance (Sig.) values of the Shapiro-Wilk normality test.

	NCV	ATC	Execution Time
Scenario 1	0.008	0.035	0.045
Scenario 2	0.007	0.016	0.051

Table 8
Significance (Sig.) values of the Wilcoxon test.

	NCV	ATC	Execution Time
Scenario 1	0.0015	0.0015	0.0050
Scenario 2	0.0015	0.0015	–

Table 9
Significance (Sig.) values from the T-Student test for the "Execution Time" variable in Scenario 2.

V1.E2 vs. V2.E4	V1.E2 vs. V3.E5	V1.E2 vs. V4.E6	V1.E2 vs. V5.E8	V1.E2 vs. V6.E9	V1.E2 vs. V7.E11
1.8×10^{-23}	9.1×10^{-18}	1.9×10^{-17}	2.5×10^{-17}	1.6×10^{-22}	1.5×10^{-9}

- The RElect_MGEC algorithm, combining density-based and graph clustering methods, has shown robust performance with real-world data from rural Yoro, Honduras. It handles equal and varying demands effectively, achieving cost savings and efficient clustering within shorter timeframes.
- The algorithm leverages the strengths of DBSCAN and graph partitioning, using a bottom-up merging process based on a greedy heuristic that prioritizes high-power, close-proximity consumers. This ensures areas with sufficient demand density are grouped for constructing larger energy systems like microgrids or grid extensions.
- Fine-tuning constraints like PotMax and MaxLongC allows the assessment of distribution infrastructures while maintaining a balance between power generation, consumption, and distribution. Unlike K-means and HCA, which require predefined cluster numbers, RElect_MGEC dynamically adapts to the data's needs.
- The distribution network based on the Minimum Spanning Tree (MST) design connects consumers efficiently, avoiding obstacles and minimizing energy losses.
- Implementing the RElect_MGEC algorithm within the IntiGIS model and QGIS environment enhances data processing, analysis, accessibility, and usability. The output maps provide clear visual representations, improving problem domain interpretability and aiding users in extracting valuable information from the clustering results.

6. Future research directions and study limitations

- **Algorithm Applicability**: While the geospatial clustering algorithm has shown promising results in this study, its application has been tested only within the IntiGIS model. Future research could explore its adaptability to other models, broadening its use in large-scale techno-economic and geospatial planning for rural electrification and similar studies. Assessing the algorithm's performance across different planning contexts will be crucial for understanding its generalizability.
- **Customization of Clustering Analyses** The algorithm currently offers limited customization options. Future enhancements could allow users to select specific constraints, such as disabling PotMax or MaxLongC, to explore different clustering scenarios. Expanding the algorithm's flexibility would increase its utility in diverse planning situations.
- **Parallel Computing for Enhanced Performance**: Implementing parallel computing techniques could significantly improve the algorithm's efficiency when handling large-scale datasets. Future research could focus on developing a parallelized version of the algorithm and evaluating its performance gains in various planning scenarios.

CRediT authorship contribution statement

Mirelys Torres-Pérez: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation. **Javier Domínguez:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization. **Luis Arribas:** Writing – review & editing, Methodology, Conceptualization. **Julio Amador:** Writing – review & editing, Methodology, Conceptualization. **Pedro Ciller:** Writing – review & editing, Validation, Methodology, Conceptualization. **Andrés González-García:** Writing – review & editing, Methodology, Conceptualization.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Appendix A. Complementary algorithms

Algorithm A.1. MaxLongCGen

Input: G, a weighted tree
Input: center_g, represents the ID of a specific node, and it is an optional parameter that enables the calculation of MaxLongC specifically from this node
Input: upd, optional bool parameter for updating the center and MaxLongC of the tree for later access
Output: MaxLongCGen, a resulting value with the MaxLongC for a tree or a specific node

```
if is_tree(G)://Check if G is a tree
    MaxLongCGen = 0
    if center_g is null://If center_g is not provided
        center_g, MaxLongCGen = center_Graph(G)
    else://Calculate MaxLongCGen from the provided center_g
        MaxLongCGen = node_MaxLongCGen(G, center_g)
    end if
    if upd://Set or update the center and MaxLCG attributes of the graph
        G.graph['center'] = center_g
         G.graph['MaxLCG'] = MaxLongCGen
     end if
     return MaxLongCGen
 else:
     return False
 end if
```

Algorithm A.2. center_Graph

```
Input: G, a weighted tree
Output: center_g, the node that represents the center of the graph
Output: minMaxLongC, value that denotes the MaxLongCGen
e = dict(.)//initialize an empty dictionary
for each node in G:
  e[node] = node_MaxLongCGen(G, node)
end for
minMaxLongC = minimum value in e
center_g = list of nodes with the minimum MaxLongC
if length of center_g > 1:
  maxDegree = 0
  centerNode = None
  for each c_node in center_g:
    if degree of c_node > maxDegree:
      maxDegree = degree of c_node
      centerNode = c_node
    end if
  end for
  center_g = centerNode
else:
  center_g = center_g[0]
end if
return center_g, minMaxLongC
```

Algorithm A.3. node_MaxLongCGen

```
Input: G, a weighted tree
Input: node, a node for which the MaxLongCGen will be calculated.
Output: MaxLongCGen, value of the MaxLongC obtained
MaxLongCGen = 0
T = dfs_tree(G, node)//Obtain oriented tree T constructed from a depth-first-search starting from the specified node
for each nbr in G[node]://Iterate through the neighbors of the central node
  path_edges_bfs = bfs_edges(T, nbr)//Get the breadth-first search edges from nbr
  LongCGen = G[node][nbr]['length']//Get the distance from the center to nbr
  for each edge in path_edges_bfs:
    LongCGen + = G.edges[edge]['length']
  if LongCGe > MaxLongCGen:
    MaxLongCGen = LongCGen
  end if
  end for
end for
return MaxLongCGen
```

Appendix B. Descriptive Statistics of the Clusters

Table 10
Exploring clustering methods: metrics examination for Scenario 1.

Exp.	Method/Input parameters	NC/Total cons	NCV/Total cons	Cluster Max/Mean/Min [a]				Total LV lines[a]
				Size	MaxLongC	P_{agr}	LV lines	
E1	V1/[b], Eps = 800 m Map_barrier	192/8966	192/8966	135/47/20	999.21/739.55/158.74	13500/4669.7/2000	2636.6/1521/280.23	292031
E2	V1/[b], Eps = 800 m	194/9001	194/9001	141/46.4/20	999.21/736.57/158.74	14100/4639.6/2000	2636.6/1514.6/280.23	293841
E3	V1/[b], Eps = 150 m	123/5837	123/5837	138/47/20	995.21/444.74/56.07	13800/4745.5/2000	2812.6/938.76/133.68	115467
E4	V2/[b]	193/8994	193/8994	141/47/20	999.21/747.46/158.74	14100/4660.1/2000	2636.6/1534.6/280.23	296179
E5	V3/[b], Eps = 150 m	113/5910	98/4156	134/42/20	968.17/393.14/56.07	13400/4240.8/2000	2130.9/818.93/133.68	80255.5
E6	V4/[b], NC = 18	188/8749	188/8749	141/47/20	999.21/738.98/158.74	14100/4653.7/2000	2636.6/1509.9/280.23	283879

(continued on next page)

Table 10 (*continued*)

Exp.	Method/Input parameters	NC/Total cons	NCV/Total cons	Cluster Max/Mean/Min [a]				Total LV lines[a]
				Size	MaxLongC	P_{agr}	LV lines	
E7	V4/[b], NC = 1000	112/4975	112/4975	134/44/20	978.14/49,63/56.07	13400/4441.9/ 2000	2175.1/1010.2/ 133.68	113149
E8	V5/[b], NC = 1000	1000/10995	175/6193	134/35/20	847.84/361.92/76.49	13400/3538.8/ 2000	1691/740.68/158.72	129619
E9	V6/[b], NC = 18	192/8961	192/8961	141/47/20	999.21/749.68/ 158.74	14100/4667.1/ 2000	2636.6/1539.4/ 280.23	295567
E10	V6/[b], NC = 1000	146/7073	146/7073	141/48.4/ 20	995.21/533.21/56.07	14100/4844.5/ 2000	2636.6/1105.1/ 133.68	161353
E11	V7/[b], NC = 1000	1000/10995	112/4975	134/44.4/ 20	978.14/493.63/56.07	13400/4441.9/ 2000	2175.1/1010.2/ 133.68	113149

[a] Value that refers to the set of valid clusters.
[b] MinCons = 20, MaxLongC = 1000 m, PotMax = 61090 W.

Table 11
Exploring clustering methods: metrics examination for Scenario 2.

Exp.	Method/Input parameters	NC/Total cons	NCV/Total cons	Cluster Max/Mean/Min [a]				Total LV lines[a]
				Size	MaxLongC	P_{agr}	LV lines	
E1	V1/[b], Eps = 800 m Map_barrier	197/8373	197/8373	135/43/ 20	998.22/728.24/ 102.84	57600/15079/ 4000	3023.7/1571.6/ 198.69	309623
E2	V1/[b], Eps = 800 m	197/8387	197/8387	135/43/ 20	998.22/728.17/ 102.84	57600/15100/ 4000	3023.7/1579.5/ 198.69	311178
E3	V1/[b], Eps = 140 m	125/5518	125/5518	134/44/ 20	997.08/497.16/68.28	57500/15472/ 4000	2525.8/1082/139.35	135259
E4	V2/[b]	195/8330	195/8330	134/43/ 20	998.22/720.89/ 158.74	57500/15137/ 4000	3023.7/1564.8/ 280.23	305140
E5	V3/[b], Eps = 140 m	111/5677	90/3437	134/38/ 20	989.92/413.21/68.28	57500/13440/ 4000	2172.2/882.34/ 139.35	79412
E6	V4/[b], NC = 63	190/8173	190/8173	135/43/ 20	998.22/717.34/ 102.84	57600/15243/ 4000	3023.7/1542.7/ 198.69	293129
E7	V4/[b], NC = 1000	175/6189	175/6189	134/35/ 20	975.95/445.32/ 102.84	57500/12573/ 4000	2188.8/937.49/ 198.69	164061
E8	V5/[b], NC = 1000	1000/10995	173/6043	134/35/ 20	892.64/439.67/ 102.84	57500/12389/ 4000	1936.8/924.54/ 198.69	159945
E9	V6/[b], NC = 63	192/8961	192/8961	141/47/ 20	999.21/749.68/ 158.74	14100/4667/2000	2636.6/1539.4/ 280.23	295567
E10	V6/[b], NC = 1550	1550/10995	126/5409	134/43/ 20	997.08/478.18/68.28	57500/15128/ 4000	2384.5/1035.7/ 139.35	130509
E11	V7/[b], NC = 1550	1550/10995	98/3732	134/38/ 20	990.2/414/68.28	57500/13417/ 4000	2172.2/886.76/ 139.35	86903

[a] Value that refers to the set of valid clusters.
[b] MinCons = 20, MaxLongC = 1000 m, PotMax = 61090 W.

Appendix C. Statistical Summary of Electrification Solutions in Scenario 1

Table 12
Electrification solution S1 (E1.V1), scenario 1.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	2418	6496	2081	10995
NCV	–	142	37	179
Number of installed systems	2418	142	37	2597
Total annual energy produced (kWh)	639863	1719004	550685	2909552
ATC (USD)	854351	1141104	295618	2291074
Weighted Average LCOE (USD/kWh)	1.34	0.66	0.54	0.79

Table 13
Electrification solution S2 (E2.V1), scenario 1.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	1994	6928	2073	10995
NCV	–	157	37	194
Number of installed systems	1994	157	37	2188
Total annual energy produced (kWh)	527662	1833322	548568	2909552

(*continued on next page*)

Table 13 (*continued*)

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
ATC (USD)	704542	1263014	290432	2257988
Weighted average LCOE (USD/kWh)	1.34	0.69	0.53	0.78

Table 14
Electrification solution S3 (E3.V1), scenario 1.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	5158	4520	1317	10995
NCV	–	103	20	123
Number of installed systems	5158	103	20	5281
Total annual energy produced (kWh)	1364936	1196105	348511	2909552
ATC (USD)	1823422	703855	158064	2685341
Weighted average LCOE (USD/kWh)	1.34	0.59	0.45	0.92

Table 15
Electrification solution S4 (E4.V2), scenario 1.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	2001	6911	2083	10995
NCV	–	156	37	193
Number of installed systems	2001	156	37	2194
Total annual energy produced (kWh)	529515	1828823	551214	2909552
ATC (USD)	707015	1261774	294112	2262901
Weighted average LCOE (USD/kWh)	1.34	0.69	0.53	0.78

Table 16
Electrification solution S5 (E5.V3), scenario 1.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	6839	3572	584	10995
NCV	–	87	11	98
Number of installed systems	6839	87	11	6937
Total annual energy produced (kWh)	1809770	945241	154541	2909552
ATC (USD)	2417630	554159	68672	3040461
Weighted average LCOE (USD/kWh)	1.34	0.59	0.44	1.04

Table 17
Electrification solution S6 (E6.V4), scenario 1.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	2246	6675	2074	10995
NCV	–	151	37	188
Number of installed systems	2246	151	37	2434
Total annual energy produced (kWh)	594348	1766372	548832	2909552
ATC (USD)	793720	1212634	291897	2298252
Weighted average LCOE (USD/kWh)	1.34	0.69	0.53	0.79

Table 18
Electrification solution S7 (E7.V4), scenario 1.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	6020	4315	660	10995
NCV	–	99	13	112
Number of installed systems	6020	99	13	6132
Total annual energy produced (kWh)	1593043	1141857	174653	2909552
ATC (USD)	2128328	697540	81554	2907422
Weighted average LCOE (USD/kWh)	1.336	0.6108	0.4669	1.00

Table 19
Electrification solution S8 (E8.V5), scenario 1.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	4802	5382	811	10995
NCV	–	155	20	175
Number of installed systems	4802	155	20	4977
Total annual energy produced (kWh)	1270729	1424212	214611	2909552
ATC (USD)	1697568	852765	95725	2646059
Weighted average LCOE (USD/kWh)	1.34	0.60	0.45	0.91

Table 20
Electrification solution S9 (E9.V6), scenario 1.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	2034	6878	2083	10995
NCV	–	155	37	192
Number of installed systems	2034	155	37	2226
Total annual energy produced (kWh)	538247	1820091	551214	2909552
ATC (USD)	718623	1256725	294112	2269460
Weighted average LCOE (USD/kWh)	1.34	0.69	0.53	0.78

Table 21
Electrification solution S10 (E10.V6), scenario 1.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	3922	5524	1549	10995
NCV	–	122	24	146
Number of installed systems	3922	122	24	4068
Total annual energy produced (kWh)	1037859	1461789	409904	2909552
ATC (USD)	1386577	896102	194604	2477283
Weighted average LCOE (USD/kWh)	1.34	0.61	0.47	0.85

Table 22
Electrification solution S11 (E11.V7), scenario 1.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	6020	4315	660	10995
NCV	–	99	13	112
Number of installed systems	6020	99	13	6132
Total annual energy produced (kWh)	1593043	1141857	174653	2909552
ATC (USD)	2128328	697540	81554	2907422
Weighted average LCOE (USD/kWh)	1.34	0.61	0.47	1.00

Appendix D. Statistical Summary of Electrification Solutions in Scenario 2

Table 23
Electrification solution S1 (E1.V1), scenario 2.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	2622	5183	3190	10995
NCV	–	131	66	197
Number of installed systems	2622	131	66	2819
Total annual energy produced (kWh)	1515252	3095592	1961118	6571962
ATC (USD)	2075328	1821826	826246	4723400
Weighted average LCOE (USD/kWh)	1.37	0.59	0.42	0.72

Table 24
Electrification solution S2 (E2.V1), scenario 2.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	2608	5286	3101	10995
NCV	–	133	64	197
Number of installed systems	2608	133	64	2805
Total annual energy produced (kWh)	1508016	3148079	1915867	6571962
ATC (USD)	2065488	1853169	799524	4718181
Weighted average LCOE (USD/kWh)	1.37	0.59	0.42	0.72

Table 25
Electrification solution S3 (E3.V1), scenario 2.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	5477	3450	2068	10995
NCV	–	86	39	125
Number of installed systems	5477	86	39	5602
Total annual energy produced (kWh)	3275601	2018331	1278029	6571962
ATC (USD)	4491069	1101219	482178	6074466
Weighted average LCOE (USD/kWh)	1.37	0.55	0.38	0.92

Table 26
Electrification solution S4 (E4.V2), scenario 2.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	2665	5266	3064	10995
NCV	–	133	62	195
Number of installed systems	2665	133	62	2860
Total annual energy produced (kWh)	1546560	3142659	1882743	6571962
ATC (USD)	2118373	1842138	781049	4741560
Weighted average LCOE (USD/kWh)	1.37	0.59	0.41	0.72

Table 27
Electrification solution S5 (E5.V3), scenario 2.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	7558	2432	1005	10995
NCV	–	67	23	90
Number of installed systems	7558	67	23	7648
Total annual energy produced (kWh)	4511884	1445573	614505	6571962
ATC (USD)	6186667	785351	248747	7220765
Weighted average LCOE (USD/kWh)	1.37	0.54	0.40	1.10

Table 28
Electrification solution S6 (E6.V4), scenario 2.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	2822	5153	3020	10995
NCV	–	129	61	190
Number of installed systems	2822	129	61	3012
Total annual energy produced (kWh)	1641022	3060598	1870342	6571962
ATC (USD)	2248299	1789901	770486	4808686
Weighted average LCOE (USD/kWh)	1.37	0.58	0.41	0.73

Table 29
Electrification solution S7 (E7.V4), scenario 2.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	4806	4219	1970	10995
NCV	–	126	49	175
Number of installed systems	4806	126	49	4981
Total annual energy produced (kWh)	2827838	2532352	1211773	6571962
ATC (USD)	3876365	1399401	469058	5744823
Weighted average LCOE (USD/kWh)	1.37	0.55	0.39	0.87

Table 30
Electrification solution S8 (E8.V5), scenario 2.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	4952	4219	1824	10995
NCV	–	126	47	173
Number of installed systems	4952	126	47	5125
Total annual energy produced (kWh)	2923988	2532352	1115623	6571962
ATC (USD)	4008338	1399401	438062	5845800
Weighted average LCOE (USD/kWh)	1.37	0.55	0.39	0.89

Table 31
Electrification solution S9 (E9.V6), scenario 2.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	2641	5266	3088	10995
NCV	–	133	63	196
Number of installed systems	2641	133	63	2837
Total annual energy produced (kWh)	1526238	3142659	1903064	6571962
ATC (USD)	2090482	1842138	792047	4724666
Weighted average LCOE (USD/kWh)	1.37	0.59	0.42	0.72

Table 32
Electrification solution S10 (E10.V6), scenario 2.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	5586	3471	1938	10995
NCV	–	88	38	126
Number of installed systems	5586	88	38	5712
Total annual energy produced (kWh)	3325716	2057888	1188358	6571962
ATC (USD)	4559662	1119501	439577	6118740
Weighted average LCOE (USD/kWh)	1.37	0.54	0.37	0.93

Table 33
Electrification solution S11 (E11.V7), scenario 2.

Characteristics	Pv stand-alone	Pv-Diesel microgrid	Grid	Total
Number of consumers	7263	2566	1166	10995
NCV	–	72	26	98
Number of installed systems	7263	72	26	7361
Total annual energy produced (kWh)	4333098	1535500	703364	6571962
ATC (USD)	5941559	833304	279808	7054670
Weighted average LCOE (USD/kWh)	1.37	0.54	0.40	1.07

References

Abdul-Salam, Y., Phimister, E., 2016a. How effective are heuristic solutions for electricity planning in developing countries. Soc. Econ. Plann. Sci. 55, 14–24. https://doi.org/10.1016/j.seps.2016.04.004.

Abdul-Salam, Y., Phimister, E., 2016b. The politico-economics of electricity planning in developing countries: a case study of Ghana. Energy Pol. 88, 299–309. https://doi.org/10.1016/j.enpol.2015.10.036.

Aggarwal, C.C., Wang, H., 2010. A survey of clustering algorithms for graph data. In: Managing and Mining Graph Data. Springer, pp. 275–301. https://doi.org/10.1007/978-1-4419-6045-0_9.

Ahmed, M., Seraj, R., Islam, S.M.S., 2020. The k-means algorithm: a comprehensive survey and performance evaluation. Electronics 9 (8), 1295. https://doi.org/10.3390/electronics9081295.

Akbas, B., Kocaman, A.S., Nock, D., Trotter, P.A., 2022. Rural electrification: an overview of optimization methods. Renew. Sustain. Energy Rev. 156, 111935 https://doi.org/10.1016/j.rser.2021.111935.

Amador, J., 2000. Análisis de los parámetros técnicos en la aplicación de los sistemas de información geográfica a la integración regional de las energías renovables en la producción descentralizada de electricidad. Universidad Politécnica de Madrid, Madrid. http://oa.upm.es/147/.

Amador, J., Domínguez, J., 2005. Application of geographical information systems to rural electrification with renewable energy sources. Renew. Energy 30 (12), 1897–1912. https://doi.org/10.1016/j.renene.2004.12.007.

Ammari, C., Belatrache, D., Touhami, B., Makhloufi, S., 2022. Sizing, optimization, control and energy management of hybrid renewable energy system—a review. Energy and Built Environment 3 (4), 399–411. https://doi.org/10.1016/j.enbenv.2021.04.002.

Banks, D., Mocke, F., Jonck, E., Labuschagne, E., Eberhard, R., 2000. *Electrification planning decision support tool* domestic use of energy conference. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.195.2137&rep=rep1&type=pdf.

Bertheau, P., Oyewo, A.S., Cader, C., Breyer, C., Blechinger, P., 2017. Visualizing national electrification scenarios for sub-saharan african countries. Energies 10 (11), 1899. https://doi.org/10.3390/en10111899.

Bhattacharjee, P., Mitra, P., 2021. A survey of density based clustering algorithms. https://doi.org/10.1007/s11704-019-9059-3.

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

Chaudhry, M., Shafi, I., Mahnoor, M., Vargas, D.L.R., Thompson, E.B., Ashraf, I., 2023. A systematic literature review on identifying patterns using unsupervised clustering algorithms: a data mining perspective. Symmetry 15 (9), 1679. https://doi.org/10.3390/sym15091679.

Cheong, D.M.L.K., Fernando, T., Iu, H.C., Reynolds, M., Fletcher, J., 2017. Review of clustering algorithms for microgrid formation, 2018/06//. https://doi.org/10.1109/ISGT-Asia.2017.8378350.

Ciller, P., 2021. *The rural electrification planning problem: strategies and solutions* [Programa de Doctorado en Energía Eléctrica, Universidad Pontificia Comillas]. https://universalaccess.mit.edu/paper-repo/Tesis_Pedro_Ciller_v9.pdf.

Ciller, P., Lumbreras, S., 2020. Electricity for all: the contribution of large-scale planning tools to the energy-access problem. Renew. Sustain. Energy Rev. 120 https://doi.org/10.1016/j.rser.2019.109624.

Ciller, P., Lumbreras, S., González-García, A., 2021. Network cost estimation for mini-grids in large-scale rural electrification planning. Energies 14 (7382). https://doi.org/10.3390/en14217382.

Corigliano, S., 2022. *Geospatial based methodology for rural electrification planning* Politecnico di Milano]. http://hdl.handle.net/10589/186486.

Dagnachew, A.G., Lucas, P.L., Hof, A.F., Gernaat, D.E., de Boer, H.-S., van Vuuren, D.P., 2017. The role of decentralized systems in providing universal electricity access in Sub-Saharan Africa–A model-based approach. Energy 139, 184–195. https://doi.org/10.1016/j.energy.2017.07.144.

Deichmann, U., Meisner, C., Murray, S., Wheeler, D., 2011. The economics of renewable energy expansion in rural Sub-Saharan Africa. Energy Pol. 39 (1), 215–227. https://doi.org/10.1016/j.enpol.2010.09.034.

Domínguez Bravo, J., Pinedo-Pascua, I., González, J.A., 2008. Herramientas SIG para la integración de energías renovables en electrificación rural. Inti-GIS. Tecnologías de la Información Geográfica para el Desarrollo Territorial. Las Palmas de Gran Canaria.

Eichfelder, G., 2021. Twenty years of continuous multiobjective optimization in the twenty-first century. EURO Journal on Computational Optimization 9, 100014. https://doi.org/10.1016/j.ejco.2021.100014.

ESCAP, 2016. Visualisation map of the interlinkages between SDG 7 and other SDGs. https://doi.org/10.1016/j.ejco.2021.100014. (Accessed 16 July 2024).

Ezugwu, A.E., Ikotun, A.M., Oyelade, O.O., Abualigah, L., Agushaka, J.O., Eke, C.I., Akinyelu, A.A., 2022. A comprehensive survey of clustering algorithms: state-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. Eng. Appl. Artif. Intell. 110, 104743 https://doi.org/10.1016/j.engappai.2022.104743.

Ezugwu, A.E., Shukla, A.K., Agbaje, M.B., Oyelade, O.N., José-García, A., Agushaka, J.O., 2021. Automatic clustering algorithms: a systematic review and bibliometric analysis of relevant literature. Neural Comput. Appl. 33, 6247–6306. https://doi.org/10.1007/s00521-020-05395-4.

Fletcher, J.R., Fernando, T., Iu, H.H., Reynolds, M., Fani, S., 2017. Economic feasibility of stand-alone power systems for existing distribution networks in rural areas. 2017 IEEE Innovative Smart Grid Technologies-Asia (ISGT-Asia). https://doi.org/10.1109/ISGT-Asia.2017.8378356.

Ghosal, A., Nandy, A., Das, A.K., Goswami, S., Panday, M., 2020. A short review on different clustering techniques and their applications. Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 69–83. https://doi.org/10.1007/978-981-13-7403-6_9, 2018.

Golalipour, K., Akbari, E., Hamidi, S.S., Lee, M., Enayatifar, R., 2021. From clustering to clustering ensemble selection: a review. Eng. Appl. Artif. Intell. 104, 104388 https://doi.org/10.1016/j.engappai.2021.104388.

Hagberg, A., Swart, P., S Chult, D., 2008. Exploring network structure, dynamics, and function using NetworkX. 7th Python in Science Conference (SciPy 2008). In: https://conference.scipy.org/proceedings/SciPy2008/paper_2/.

Hall, O., Ohlsson, M., Rögnvaldsson, T., 2022. A review of explainable AI in the satellite data, deep machine learning, and human poverty domain. Patterns 3 (10). https://doi.org/10.1016/j.patter.2022.100600.

Hassan, W., Alshameri, B., Nawaz, M.N., Ijaz, Z., Qasim, M., 2022. Geospatial and statistical interpolation of geotechnical data for modeling zonation maps of Islamabad, Pakistan. Environ. Earth Sci. 81 (24), 547. https://doi.org/10.1007/s12665-022-10669-2.

Hassan, W., Raza, M.F., Alshameri, B., Shahzad, A., Khalid, M.H., Nawaz, M.N., 2023. Statistical interpolation and spatial mapping of geotechnical soil parameters of District Sargodha, Pakistan. Bull. Eng. Geol. Environ. 82 (1), 37. https://doi.org/10.1007/s10064-022-03059-2.

IEA, 2021. India energy outlook 2021. https://www.iea.org/reports/india-energy-outlook-2021. (Accessed 6 August 2024).

IEA, 2022. *For the first time in decades, the number of people without access to electricity is set to increase in 2022* IEA. https://www.iea.org/commentaries/for-the-first-time-in-decades-the-number-of-people-without-access-to-electricity-is-set-to-increase-in-2022. (Accessed 14 April 2023).

IEA, 2024. Progress on basic energy access reverses for first time in a decade. https://www.iea.org/news/progress-on-basic-energy-access-reverses-for-first-time-in-a-decade. (Accessed 8 August 2024).

Ikotun, A.M., Ezugwu, A.E., Abualigah, L., Abuhaija, B., Heming, J., 2022. K-means clustering algorithms: a comprehensive review, variants analysis, and advances in the era of big data. Inf. Sci. https://doi.org/10.1016/j.ins.2022.11.139.

Innovation Énergie Développement, I.E.D., 2021. Geosim - geospatial rural electrification planning. https://www.ied-sa.com/en/products/planning/geosim-gb.html. (Accessed 5 July 2023).

Jejeniwa, T.O., Mhlongo, N.Z., Jejeniwa, T.O., 2024. AI solutions for developmental economics: opportunities and challenges in financial inclusion and poverty alleviation. International Journal of Advanced Economics 6 (4), 108–123. https://doi.org/10.51594/ijae.v6i4.1073.

Jenkner, J., 2020. Cluster points. https://jjenkner.com/ClusterPoints/. (Accessed 28 June 2023).

Kabudi, T.M., 2022. Artificial intelligence for quality education: successes and challenges for AI in meeting SDG4. International Conference on Social Implications of Computers in Developing Countries. https://doi.org/10.1007/978-3-031-19429-0_21.

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

Korkovelos, A., 2020. Advancing the State of Geospatial Electrification Modelling: New Data, Methods, Applications, Insight and Electrification Investment Outlooks. KTH Royal Institute of Technology (Publication Number 978-91-7873-562-4) [Doctoral Thesis in Energy and Environmental Systems. https://www.diva-portal.org/smash/record.jsf?pid=diva2:1431484.

Lenssen, L., Schubert, E., Krivošija, A., Schubert, E., Lang, A., Hess, S., 2023. 5 cluster analysis. In: Katharina, M., Peter, M. (Eds.), Volume 1 Machine Learning under Resource Constraints - Fundamentals. De Gruyter, pp. 179–248. https://doi.org/10.1515/9783110785944-005.

Leonard, A., 2022. *Scalable spatial design of electricity access systems* University of Oxford. https://ora.ox.ac.uk/objects/uuid:7bf531ac-8eff-429e-8dc9-3f64fc4ab164.

Levin, T., Thomas, V.M., 2012. Least-cost network evaluation of centralized and decentralized contributions to global electrification. Energy Pol. 41, 286–302. https://doi.org/10.1016/j.enpol.2011.10.048.

Li, X., Ren, Z., Sun, Q., Xu, Z., 2023. Auto-weighted tensor schatten p-norm for robust multi-view graph clustering. Pattern Recogn. 134, 109083 https://doi.org/10.1016/j.patcog.2022.109083.

Li, X., Sun, Y., Sun, Q., Ren, Z., 2022. Consensus cluster center guided latent multi-kernel clustering. IEEE Trans. Circ. Syst. Video Technol. https://doi.org/10.1109/TCSVT.2022.3229356.

Liao, X.Q., Su, T., Ma, L., 2020. Application of neutrosophic minimum spanning tree in electrical power distribution network. CAAI Transactions on Intelligence Technology 5 (2), 99–105. https://doi.org/10.1049/trit.2019.0100.

Lin, C.-C., Huang, A.Y., Lu, O.H., 2023. Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. Smart Learning Environments 10 (1), 41. https://doi.org/10.1186/s40561-023-00260-y.

Lopez-Vargas, A., Ledezma-Espino, A., Sanchis-de-Miguel, A., 2022. Methods, data sources and applications of the artificial intelligence in the energy poverty context: a review. Energy Build. 268, 112233 https://doi.org/10.1016/j.enbuild.2022.112233.

Lukes, J.A., 1974. Efficient algorithm for the partitioning of trees. IBM J. Res. Dev. 18 (3), 217–224. https://doi.org/10.1147/rd.183.0217.

Mahapatra, S., Dasappa, S., 2012. Rural electrification: optimising the choice between decentralised renewable energy sources and grid extension. Energy for Sustainable Development 16 (2), 146–154. https://doi.org/10.1016/j.esd.2012.01.006.

Mahmmud, F., Watson, S., Woods, J., Halliday, J., Hossain, J., 1996. The economic potential for renewable EnergySources in Karnataka, India. European Wind Energy Association Conference and Exhibition, Exeter (UK).

Martínez Sarmiento, R.A., Domínguez, J., Amador, J., 2014. Tecnologías de información geográfica para la electrificación rural con energías renovables. Caso de estudio: Carchi, Ecuador (Informes Técnicos Ciemat, Issue. http://documenta.ciemat.es/handle/123456789/108.

Mehmood, H., Liao, D., Mahadeo, K., 2020. A review of artificial intelligence applications to achieve water-related sustainable development goals. 2020 IEEE/ITU International Conference on Artificial Intelligence for Good (AI4G). https://doi.org/10.1109/AI4G50087.2020.9311018.

Min, Y., Perucci, F., 2020. UN/DESA Policy Brief #81: impact of COVID-19 on SDG progress: a statistical perspective. https://www.un.org/development/desa/dpad/

publication/un-desa-policy-brief-81-impact-of-covid-19-on-sdg-progress-a-statistical-perspective/.
Moner-Girona, M., Bódis, K., Huld, T., Kougias, I., Szabó, S., 2016. Universal access to electricity in Burkina Faso: scaling-up renewable energy technologies. Environ. Res. Lett. 11 (8), 084010 https://doi.org/10.1088/1748-9326/11/8/084010.
Monteiro, C., Saraiva, J.T., Miranda, V., 1998. Evaluation of electrification alternatives in developing countries - the SOLARGIS tool. Melecon '98 - 9th Mediterranean Electrotechnical Conference 2, 1037–1041, 1 and 2. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=699387.
Morrissey, J., 2019. Achieving universal electricity access at the lowest cost: a comparison of published model results. Energy for Sustainable Development 53, 81–96. https://doi.org/10.1016/j.esd.2019.09.005.
Nascimento, M.C., De Carvalho, A.C., 2011. Spectral methods for graph clustering–a survey. Eur. J. Oper. Res. 211 (2), 221–231. https://doi.org/10.1016/j.ejor.2010.08.012.
Nawaz, M.N., Akhtar, A.Y., Awan, T.A., Nawaz, M.M., Qamar, S.U., Shehzad, T., Farooq, R., 2024a. Multivariate formulation to predict the frictional strength of fiber reinforced soils using gene expression programming. Eng. Appl. Artif. Intell. 134, 108660 https://doi.org/10.1016/j.engappai.2024.108660.
Nawaz, M.N., Alshameri, B., Maqsood, Z., Hassan, W., 2024b. Predictive modelling of cohesion and friction angle of soil using gene expression programming: a step towards smart and sustainable construction. Neural Comput. Appl. 36 (18), 10545–10566. https://doi.org/10.1007/s00521-024-09626-w.
NetworkX Developers, 2020. NetworkX. Version v2.5, Version 2.5. https://networkx.org/documentation/stable/release/release_2.5.html.
Oladeji, O., 2018. *Network partitioning algorithms for electricity consumer clustering* Massachusetts Institute of Technology]. https://hdl.handle.net/1721.1/122917.
Oyelade, J., Isewon, I., Oladipupo, O., Emebo, O., Omogbadegun, Z., Aromolaran, O., Olawole, O., 2019. Data clustering: algorithms and its applications. 2019 19th International Conference on Computational Science and its Applications (ICCSA). https://doi.org/10.1109/ICCSA.2019.000-1.
Page Arias, A., 2015. IntiGIS 2.0: Objetivos y propuesta metodológica. Análisis de competitividad tecnológica en Ghana. [MSc, Universidad Complutense de Madrid], Madrid. https://eprints.ucm.es/33324/.
Parreno Jr, R.P., Del Mundo, R., 2015. Clustering algorithm as A planning support tool for rural electrification optimization. Int. J. Sci. Technol. Res. 4 (12), 112–120. http://refhub.elsevier.com/S1364-0321(19)30831-7/sref24.
Pinedo-Pascua, I., 2010. IntiGIS: propuesta metodológica para la evaluación de alternativas de electrificación rural basada en SIG [PhD, Universidad Politécnica de Madrid]. Madrid. https://dialnet.unirioja.es/servlet/tesis?codigo=185480.
Pons, J.L.P., Valdés, J.E., Maroto, L.D., 2013. Planificación integrada de electrificación mediante SIG (Integrated electrification planning using GIS). An. Mec. Electr. XC (1), 6. http://hdl.handle.net/11531/5002.
QGIS Development Team, 2022. Changelog for QGIS 3.22. https://qgis.org/en/site/forusers/visualchangelog322/index.html.
QGIS Development Team, 2023. QGIS Python API. https://qgis.org/pyqgis/3.22/.
Qin, Y., Ding, S., Wang, L., Wang, Y., 2019. Research progress on semi-supervised clustering. Cognitive Computation 11 (5), 599–612. https://doi.org/10.1007/s12559-019-09664-w.
Quevedo Saldias, A. (2022). *Uso de datos GIS para electrificación rural y su aplicación para medir el potential de Microrredes y Sistemas Fotovoltaicos Individuales en Yoro – Honduras* Universidad Internacional de Valencia].
Raj, S., Bhattacharyya, B., 2016. Weak bus determination and real power loss minimization using Grey wolf optimization. 2016 IEEE 6th International Conference on Power Systems (ICPS). https://doi.org/10.1109/ICPES.2016.7584143.
Raj, S., Bhattacharyya, B., 2018. Optimal placement of TCSC and SVC for reactive power planning using Whale optimization algorithm. Swarm Evol. Comput. 40, 131–143. https://doi.org/10.1016/j.swevo.2017.12.008.
Ran, X., Xi, Y., Lu, Y., Wang, X., Lu, Z., 2023. Comprehensive survey on hierarchical clustering algorithms and the recent developments. Artif. Intell. Rev. 56 (8), 8219–8264. https://doi.org/10.1007/s10462-022-10366-3.
Ran, X., Zhou, X., Lei, M., Tepsan, W., Deng, W., 2021. A novel K-means clustering algorithm with a noise algorithm for capturing urban hotspots. Appl. Sci. 11 (23), 11202. https://www.mdpi.com/2076-3417/11/23/11202.
Rodriguez, M.Z., Comin, C.H., Casanova, D., Bruno, O.M., Amancio, D.R., Costa, L.d.F., Rodrigues, F.A., 2019. Clustering algorithms: a comparative approach. PLoS One 14 (1). https://doi.org/10.1371/journal.pone.0210236. Article e00205.
Romero Otero, L., 2016. Sistemas de Información Geográfica y electrificación rural. *Análisis, desarrollo y estudio de caso con IntiGIS* [Master. Universidad Complutense de Madrid, Madrid. http://eprints.ucm.es/41911/.
Rosenberg, M., French, T., Reynolds, M., While, L., 2022. Finding an optimised infrastructure for electricity distribution networks in rural areas - a comparison of different approaches. Swarm Evol. Comput. 68 https://doi.org/10.1016/j.swevo.2021.101018. Article 101018.
Rouault, E., Warmerdam, F., Schwehr, K., Kiselev, A., Butler, H., Łoskot, M., Dawson, N., 2023. GDAL. Zenodo. https://doi.org/10.5281/zenodo.8144520, Version v3.7.1.
Rubin, M., 2021. When to adjust alpha during multiple testing: a consideration of disjunction, conjunction, and individual testing. Synthese 199 (3), 10969–11000. https://doi.org/10.1007/s11229-021-03276-4.
Sahai, D., 2013. Toward Universal Electricity Access: Renewable Energy-Based Geospatial Leastcost Electrification Planning (No. 84314). The World Bank. https://www.esmap.org/sites/default/files/esmap-files/ASTAE_Indonesia_TUEA_Brief_1.pdf.
Sahlberg, A., 2023. *New Methods and Applications to Explore the Dynamics of Least-Cost Technologies in Geospatial Electrification Modelling* [Doctoral Thesis. KTH Royal Institute of Technology]. DiVA, Stockholm. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-337676.
Schaeffer, S.E., 2007. Graph clustering. Computer science review 1 (1), 27–64. https://doi.org/10.1016/j.cosrev.2007.05.001.
Schubert, E., Sander, J., Ester, M., Kriegel, H.P., Xu, X., 2017. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. ACM Trans. Database Syst. 42 (3) https://doi.org/10.1145/3068335.
Shaikh, M.S., Ansari, M.M., Jatoi, M.A., Arain, Z.A., Qader, A.A., 2020. Analysis of underground cable fault techniques using MATLAB simulation. Sukkur IBA Journal of Computing and Mathematical Sciences 4 (1), 1–10. https://doi.org/10.30537/sjcms.v4i1.566.
Shaikh, M.S., Hua, C., Hassan, M., Raj, S., Jatoi, M.A., Ansari, M.M., 2022a. Optimal parameter estimation of overhead transmission line considering different bundle conductors with the uncertainty of load modeling. Optim. Control Appl. Methods 43 (3), 652–666. https://doi.org/10.1002/oca.2772.
Shaikh, M.S., Hua, C., Raj, S., Kumar, S., Hassan, M., Mohsin Ansari, M., Ali Jatoi, M., 2022b. Optimal parameter estimation of 1-phase and 3-phase transmission line for various bundle conductor's using modified whale optimization algorithm. Int. J. Electr. Power Energy Syst. 138, 107893 https://doi.org/10.1016/j.ijepes.2021.107893.
Shaikh, M.S., Raj, S., Babu, R., Kumar, S., Sagrolikar, K., 2023. A hybrid moth–flame algorithm with particle swarm optimization with application in power transmission and distribution. Decision Analytics Journal 6, 100182. https://doi.org/10.1016/j.dajour.2023.100182.
Shaikh, M.S., Raj, S., Ikram, M., Khan, W., 2022c. Parameters estimation of AC transmission line by an improved moth flame optimization method. Journal of Electrical Systems and Information Technology 9 (1), 25. https://doi.org/10.1186/s43067-022-00066-x.
Solargis, 2019. Longterm yearly average of global irradiation at optimum tilt -Honduras - global Solar Atlas 2. https://globalsolaratlas.info/download/honduras.
Sun, Y., Ren, Z., Hu, P., Peng, D., Wang, X., 2023a. Hierarchical consensus hashing for cross-modal retrieval. IEEE Trans. Multimed. https://doi.org/10.1109/TMM.2023.3272169.
Sun, Y., Wang, X., Peng, D., Ren, Z., Shen, X., 2023b. Hierarchical hashing learning for image set classification. IEEE Trans. Image Process. 32, 1732–1744. https://doi.org/10.1109/TIP.2023.3251025.
Swarndeep Saket, J., Pandya, S., 2016. An overview of partitioning algorithms in clustering techniques. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) 5 (6), 1943–1946.
Szabó, S., Bódis, K., Huld, T., Moner-Girona, M., 2013. Sustainable energy planning: leapfrogging the energy poverty gap in Africa. Renew. Sustain. Energy Rev. 28, 500–509. https://doi.org/10.1016/j.rser.2013.08.044.
Torres-Pérez, M., Domínguez, J., Hernández Leyva, C., Peña Abreu, M., 2021. Freeware GIS tool for the techno-economic evaluation of rural electrification alternatives. Acta Sci. Pol. Administratio Locorum 20 (1), 47–58. https://doi.org/10.31648/aspal.5821.
Torres-Pérez, M., Domínguez, J., Rodríguez, M., Peña Abreu, M., 2019. Tool for the planning of rural electrification taking into account criteria of the territorial ordering. Revista Cubana de Ciencias Informáticas 13 (3). https://rcci.uci.cu/?journal=rcci&page=article&op=view&path%5B%5D=1886.
Torres-Pérez, M., Peña Abreu, M., Domínguez, J., 2024. Improvements to the IntiGIS model related to the clustering of consumers for rural electrification. In: Progress in Artificial Intelligence and Pattern Recognition. IWAIPR 2023, vol. 14335. Springer, Nature Switzerland, pp. 260–272. https://doi.org/10.1007/978-3-031-49552-6_23.
Torres-Pérez, M., Rodríguez Gamez, M., Vázquez Pérez, A., Castro Fernández, M., Vilaragout Llanes, M., Domínguez, J., 2022. The territorial ordering for the diversification of the electrical system. Case study: Isla de la Juventud. Acta Sci. Pol. Administratio Locorum 21 (2), 241–266. https://doi.org/10.31648/aspal.7005.
UN, 2020. The sustainable development goals report 2020: goal 7. https://unstats.un.org/sdgs/report/2020/goal-07/. (Accessed 6 August 2024).
UN, 2024. Sustainable development goals: goal 7 - affordable and clean energy. https://www.un.org/sustainabledevelopment/energy/. (Accessed 10 March 2024).
UNDP, 2024. Energy access. https://www.undp.org/energy/our-work-areas/energy-access. (Accessed 10 June 2024).
Van Ruijven, B.J., Schers, J., van Vuuren, D.P., 2012. Model-based scenarios for rural electrification in developing countries. Energy 38 (1), 386–397. https://doi.org/10.1016/j.energy.2011.11.037.
Vandenbergh, M., Neirac, F.P., Turki, H., 1999. A GIS approach for the siting of solar thermal power plants application to Tunisia. J. Phys. IV 9 (P3), 223–228. https://doi.org/10.1051/jp4:1999333.
Vinicius, G.T.F., Silvia, C., Aleksandar, D., Massimo, B., Marco, M., 2021. Rural electrification planning based on graph theory and geospatial data: a realistic topology oriented approach. Sustainable Energy, Grids and Networks 28. https://doi.org/10.1016/j.segan.2021.100525, 2021.
Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Fuso Nerini, F., 2020. The role of artificial intelligence in achieving the Sustainable Development Goals. Nat. Commun. 11 (1), 233. https://doi.org/10.1038/s41467-019-14108-y.
William, L., 2019. Geospatial python tutorial - PyQGIS in jupyter Notebook. Retrieved from. https://lerryws.xyz/posts/PyQGIS-in-Jupyter-Notebook. (Accessed 26 June 2023).

World Bank Group, 2016. Achieving universal access in the Kaduna electric service area. http://documents.worldbank.org/curated/en/782491487138851242/Achieving-universal-access-in-the-Kaduna-electric-service-area.

Xu, D., Tian, Y., 2015. A comprehensive survey of clustering algorithms. Annals of Data Science 2 (2), 165–193. https://doi.org/10.1007/s40745-015-0040-1.

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