MASTER EN INGENIERÍA EN TECNOLOGÍAS INDUSTRIALES

TRABAJO FIN DE MASTER

The impact of building location accuracy on electrification planning

Autor: Fidel Sanz Azuara

Director: Rafael Palacios Hielscher

Co-Director: Pablo Dueñas Martínez

Madrid

Agosto de 2023

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título

The impact of building location accuracy on electrification planning

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

curso académico 2022/2023 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos. El Proyecto no es

plagio de otro, ni total ni parcialmente y la información que ha sido tomada

de otros documentos está debidamente referenciada.

Fdo.: Fidel Sanz Azuara   Fecha: 29/ 08/ 2023

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

Fdo.: Rafael Palacios Hielscher   Fecha: 29/ 08/ 2023

Fdo.: Pablo Dueñas Martínez   Fecha: 29/ 08/ 2023

MASTER EN INGENIERÍA EN TECNOLOGÍAS INDUSTRIALES

TRABAJO FIN DE MASTER

The impact of building location accuracy on electrification planning

Autor: Fidel Sanz Azuara

Director: Rafael Palacios Hielscher

Co-Director: Pablo Dueñas Martínez

Madrid

Agosto de 2023

Agradecimientos

Quisiera provechar la oportunidad para agradecer a las personas que me han ayudado tanto durante mi recorrido académico como en este proyecto.

En primer lugar, a mi padre y a mi madre que me han apoyado en todo momento.

Igualmente, agradecer a mi familia por mostrarme siempre su interés y apoyo.

También agradecer a mis amigos y compañeros del grado y master con los que tantos buenos momentos he vivido a lo largo de estos años.

Asimismo, me gustaría agradecer a los profesores de la universidad por la formación que me han proporcionado, tanto técnica como en valores.

Finalmente, agradecer a mis directores del proyecto Rafael Palacios Hielscher y Pablo Dueñas Martínez por su disponibilidad y consejo a lo largo de todo el proceso.

THE IMPACT OF BUILDING LOCATION ACCURACY ON ELECTRIFICATION PLANNING

Autor: Sanz Azuara, Fidel.
Director: Palacios Hielscher, Rafael
Co-director: Dueñas Martínez, Pablo
Entidad Colaboradora: IIT-Comillas

RESUMEN DEL PROYECTO

Este proyecto, aborda el impacto de la obtención de la ubicación de demanda en la planificación de redes eléctricas. Se desarrolla un método que, mediante una función, genera puntos de ubicación de forma aleatoria, sujetos a restricciones, que reflejan edificios. La evaluación de este método se realiza mediante la herramienta de planificación de redes eléctricas REM (Reference Electrification Model). Se comparan los resultados obtenidos con otros datos reales de puntos que se introducen en la misma herramienta. Los resultados obtenidos sugieren que este enfoque podría simplificar la expansión de la electrificación en regiones con información limitada sobre ubicaciones de edificios.

El método propuesto tiene el potencial de mejorar la eficiencia en la creación de redes eléctricas, reduciendo la necesidad de estudios costosos in situ para identificar ubicaciones de demanda. Al utilizar datos satelitales y métodos de generación de puntos aleatorios, se puede agilizar la planificación y la expansión de redes eléctricas en áreas remotas y con acceso limitado a información precisa.

Aunque el enfoque puede tener limitaciones debido a la falta de datos precisos de ubicación, sigue siendo una herramienta valiosa para abordar desafíos técnicos y financieros en la electrificación global.

En resumen, este proyecto muestra la viabilidad de un método que utiliza funciones e información geográfica básica para generar ubicaciones de demanda eléctrica, utilizables en la planificación de redes eléctricas. Esta aproximación ofrece la posibilidad de ampliar el acceso a la electricidad en áreas con desafíos de infraestructura, al agilizar la planificación y reducir los costes.

Palabras clave: Electrificación, distribución, redes eléctricas, ubicación, geolocalización, aleatorio, edificio, desarrollo

1. **Introducción**

   El acceso a la electricidad juega una parte importante en el día a día de la población mundial, aunque muchas veces pasa desapercibido. El acceso a la electricidad en la tierra es muy desigual, estando los países más desarrollados completamente cubiertos

por las redes eléctricas mientras que en los países en desarrollo el porcentaje con acceso a una toma eléctrica es inferior, especialmente en las zonas rurales.[1]

Aunque la información no es concluyente sobre la relación entre la disponibilidad y el acceso a la electricidad y la mejora de la economía, existen estudios que muestran que un mayor acceso a la electricidad en países en desarrollo está directamente relacionado con mejoras en la educación, la sanidad y el desarrollo económico.[2][3]

Por ello, múltiples organizaciones e instituciones trabajan por posibilitar que todos los ciudadanos de estos países cuenten con electricidad. Además del desarrollo de sistemas de generación eléctrica más eficientes y de fácil instalación que se puedan llevar a lugares remotos, es necesario ampliar las redes existentes. Esto es un reto importante con ciertas dificultades técnicas y, principalmente, financieras. Es por ello que se han creado herramientas que calculan los costes de la creación de la infraestructura a partir de la información de la demanda y la generación, así como de la red existente.[4]

La información de la demanda es uno de las principales carencias. Es posible realizar estimaciones de cuánto necesita una unidad familiar pero es preciso ubicar donde se encuentra dicha demanda para llevar el cable hasta ella. Varios métodos de detección de casas mediante imágenes satelitelitales han sido desarrollados y existen varias bases de datos con dicha información.[5][6] Esta información tiene cierto grado de error y no está ampliamente disponible en todas las regiones. Estas bases de datos permiten abaratar el coste de calcular la red a generar, ya que no es necesario hacer un costoso estudio in situ, más preciso, sobre la ubicación exacta de los edificios. Para poder usar dichas bases de datos u otras metodologías de ubicación de la demanda es necesario evaluar el impacto que las ubicaciones, obtenidas con distintos métodos, tienen sobre la red eléctrica generada.

2. **Definición del Proyecto**

En este proyecto se evalúa el impacto del método utilizado en la obtención de los puntos que se emplean como ubicaciones de demanda en el proceso de creación de una red eléctrica. Para ello se desarrolla una metodología de generación de puntos aleatorios que, cumpliendo ciertas restricciones, representan dichos edificios. Para evaluar la viabilidad de este método se utiliza una herramienta de generación/planificación de redes eléctricas, llamada REM, (Reference Electrification Model). Los resultados obtenidos al utilizar REM, introduciendo los puntos generados, son comparados con las redes eléctricas que se obtienen en la misma herramienta para otros datos de entrada.

Además, el método desarrollado puede facilitar el proceso de electrificación en lugares donde la información sobre la ubicación de los edificios es escasa.

3. Descripción del sistema

El sistema desarrollado genera los puntos que representan la ubicación de los edificios a partir de los datos del contorno de la región a analizar, el número de edificios en la zona y una base de datos con información de la existencia o ausencia de asentamientos en superficies cuadradas de 30x30 metros.

En la Ilustración 1 se muestra un diagrama con las distintas entradas y salidas del sistema desarrollado, llamado Sprinkle.
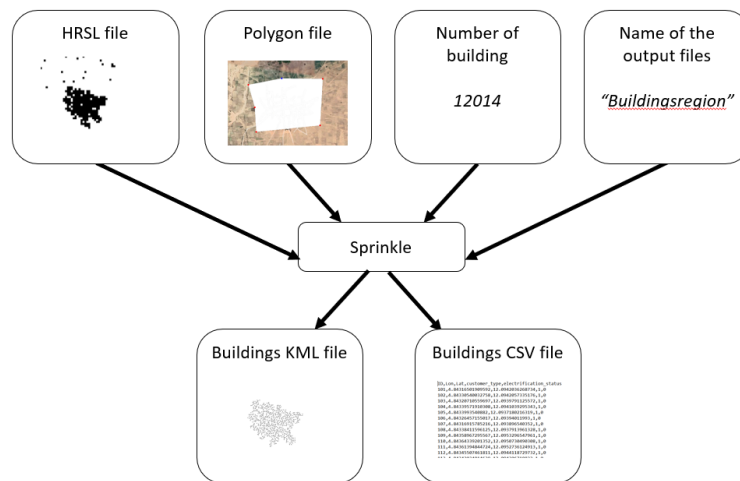

Ilustración 1 - Esquema del sistema desarrollado.

El desarrollo y ejecución de Sprinkle es una primera fase del proyecto. Seguidamente, en la segunda fase, se utilizan los puntos generados y otros puntos, obtenidos de imágenes de satélites, como entradas a la herramienta de planificación de redes eléctricas. Los resultados obtenidos serán analizados para ver la viabilidad del sistema propuesto.
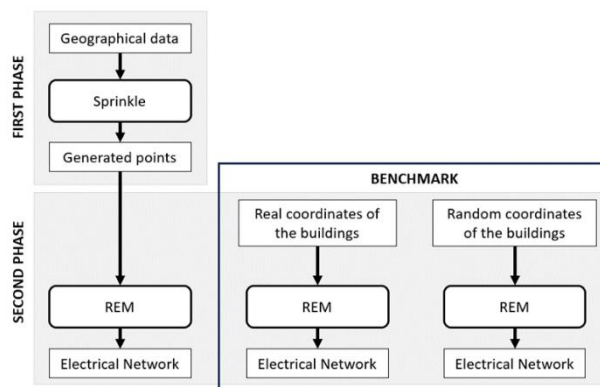

Ilustración 2 – Fases del proceso de desarrollo y evaluación de Sprinkle

En la Ilustración 2 se muestra un esquema con las distintas fases del proyecto: desarrollo de la herramienta de generación de puntos que representan edificios y evaluación de los resultados usando REM.

4. Resultados

Los resultados son prometedores, mostrando que, tanto en un área de estudio pequeña como en un área grande, el impacto en los kilómetros y costes de la red a generar son considerablemente pequeños. Esto permite que se pueda utilizar el sistema creado como estudio preliminar en la creación de una red eléctrica, especialmente útil en el planteamiento de los costes previsibles en la construcción de las infraestructuras eléctricas.

Además, extrapolando lo obtenido a una base de datos más fiable que la generada con el método desarrollado en este proyecto, se puede llegar a usar dichos datos en fases más avanzadas de la creación de la red eléctrica.

5. Conclusiones

La expansión de las redes eléctricas a lugares remotos tiene como aliados los múltiples implicados en el mapeado del planeta y más específicamente en aquellos que tratan de tener detectadas y almacenadas las ubicaciones de los edificios en países en desarrollo. El sistema desarrollado y analizado en este proyecto aporta otra alternativa para definir la ubicación de la demanda en lugares remotos. Dicha información es crucial en el cálculo de la ampliación de las redes existentes y en la obtención de fondos internacionales. Todo ello contribuye a llevar la electricidad a países en desarrollo.

6. Referencias

[1] World Bank - Access to electricity (% of population), https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS?most_recent_value_desc=false&view=chart

[2] Fatih Karanfil, Yuanjing Li, "Electricity consumption and economic growth: Exploring panel-specific differences",https://doi.org/10.1016/j.enpol.2014.12.001

[3] Jimenez, R, "Development Effects of Rural Electrification", https://doi.org/10.18235/0000629

[4] "Universal Energy Access Lab - MIT & IIT-Comillas." https://universalaccess.mit.edu/#/main (accessed May 29, 2023).

[5] Y. Aksoy and A. Aydin Alatan, "Utilization of false color images in shadow detection," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7584 LNCS, no. PART 2, pp. 472–481, 2012, doi: 10.1007/978-3-642-33868-7_47/COVER.

[6] O. Borkovska, J. Gill, L. Pistolesi, J. Squires, and G. Yetman, "Mapping the Distribution of Human Population: The High Resolution Settlement Layer (HRSL) Analysis Country Global Urban Footprint Global Human Settlement Layer High Resolution Settlement Layer", Accessed: Jan. 24, 2023. [Online]. Available: https://sedac.ciesin.columbia.edu/binaries/web/global/news/2018/borkovska_hrsl_ob_gy_v3.pdf

THE IMPACT OF BUILDING LOCATION ACCURACY ON ELECTRIFICATION PLANNING

Author: Sanz Azuara, Fidel.
Supervisor: Palacios Hielscher, Rafael
Co-supervisor: Dueñas Martínez, Pablo
Collaborating Entity: IIT-Comillas

ABSTRACT

This Project aims to assess the impact that the location of the electricity demand has on the planification of the electrical grid. A method is developed that, by means of a function, generates random location points, subject to constraints, that represent buildings. The evaluation of this method is carried out using the REM (Reference Electrification Model) electrical network planning tool. The results obtained are compared with other real point data entered into the same tool. The results obtained suggest that this approach could simplify electrification expansion in regions with limited information on building locations.

The proposed method has the potential to improve efficiency in the creation of power grids by reducing the need for costly on-site surveys to identify demand locations. By using satellite data and random point generation methods, planning and expansion of power grids in remote areas with limited access to accurate information can be expedited.

Although the approach may have limitations due to the lack of accurate location data, it remains a valuable tool for addressing technical and financial challenges in global electrification.

To summarize, this project demonstrates the feasibility of a method that uses basic geographic functions and information to generate electricity demand locations for use in power grid planning. This approach offers the potential to expand access to electricity in areas with infrastructure challenges by streamlining planning and reducing costs.

Keywords: Electrification, distribution, power grids, location, geolocation, random, building, development

1. Introduction

Access to electricity plays an important part in the daily life of the world's population, although it often goes unnoticed. Access to electricity on Earth is very unequal, with the most developed countries being completely covered by electricity grids while in developing countries the percentage with access to an electrical outlet is lower, especially in rural areas.[1] In the developing countries, the percentage with access to an electrical outlet is lower.

Although the data is inconclusive on the relationship between availability and access to electricity and economic improvement, there are studies that show that greater access to electricity in developing countries is directly related to improvements in education, health and economic development.[2][3]

For this reason, many organizations and institutions are working to make it possible for all citizens of these countries to have access to electricity. In addition to the development of more efficient and easy-to-install electricity generation systems that can be taken to remote locations, it is necessary to expand the existing networks. This is a major challenge with certain technical and, above all, financial difficulties. This is why tools have been created that calculate the costs of infrastructure creation based on information on demand and generation, as well as the existing grid.[4]

Demand information is one of the main shortcomings. It is possible to make estimates of how much a household needs but it is necessary to pinpoint where the demand is located in order to run the cable to it. Several methods of house detection using satellite images have been developed and there are several databases with such information.[5][6] This information has a certain degree of error and is not widely available in all regions. These databases allow lowering the cost of calculating the network to be generated, since it is not necessary to make a costly and more accurate on-site survey of the exact location of buildings. In order to use such databases or other demand location methodologies, it is necessary to evaluate the impact that the locations, obtained with different methods, have on the generated power grid.

2. Project definition

This project evaluates the impact of the method used to obtain the points used as demand locations in the process of creating an electrical network. For this purpose, a methodology is developed for the generation of random points that, by complying with certain restrictions, represent these buildings. To evaluate the feasibility of this method, an electrical network generation/planning tool called REM (Reference Electrification Model) is used. The results obtained by using REM, introducing the generated points, are compared with the electrical networks obtained in the same tool for other input data.

Moreover, the developed method can facilitate the electrification process in places where information about the location of buildings is scarce.

3. Description of the system

The developed system generates the points representing the location of the buildings from the contour data of the region to be analyzed, the number of buildings in the area and a database with information on the existence or absence of settlements in square areas of 30x30 meters.
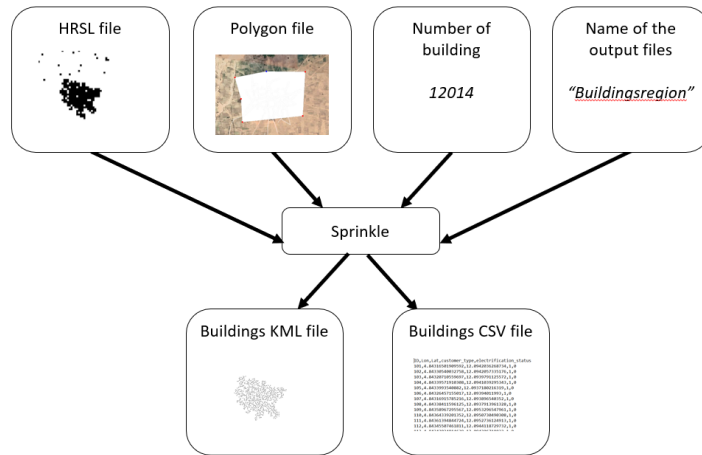

Figure 1 - Inputs and outputs of Sprinkle.

Figure 1 shows a diagram with the different inputs and outputs of the developed system, called Sprinkle.

The development and implementation of Sprinkle is the first phase of the project. Then, in the second phase, the generated points and other points, obtained from satellite images, are used as inputs to the electrical network planning tool. The results obtained will be analyzed to see the feasibility of the proposed system.
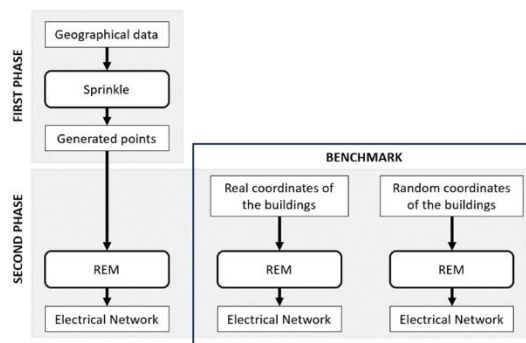

Figure 2 – Phases of the Sprinkle development and evaluation process

In Figure 2 a diagram of the different phases of the project is shown: development of the tool (Sprinkle) for the generation of points representing buildings and evaluation of the results using REM.

4. Results

The results are promising, showing that, both in a small study area and in a large area, the impact on the kilometers and costs of the network to be generated are considerably small. This allows the created system to be used as a preliminary study in the creation of an electrical network, especially useful in the approach of foreseeable costs in the construction of electrical infrastructures.

Furthermore, by extrapolating the obtained data to a more reliable database than the one generated with the method developed in this project, it is possible to use such data in more advanced phases of the creation of the electrical network.

5. Conclusions

The expansion of power grids to remote locations has as allies the multiple allies involved in the mapping of the planet and more specifically in those who try to have detected and stored the locations of buildings in developing countries. The system developed and analyzed in this project provides another alternative to define the location of demand in remote locations. Such information is crucial in calculating the extension of existing networks and in obtaining international funding. All this contributes to bringing electricity to developing countries.

6. References

[1] World Bank - Access to electricity (% of population), https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS?most_recent_value_desc=false&view=chart

[2] Fatih Karanfil, Yuanjing Li, "Electricity consumption and economic growth: Exploring panel-specific differences",https://doi.org/10.1016/j.enpol.2014.12.001

[3] Jimenez, R, "Development Effects of Rural Electrification", https://doi.org/10.18235/0000629

[4] "Universal Energy Access Lab - MIT & IIT-Comillas." https://universalaccess.mit.edu/#/main (accessed May 29, 2023).

[5] Y. Aksoy and A. Aydin Alatan, "Utilization of false color images in shadow detection," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7584 LNCS, no. PART 2, pp. 472–481, 2012, doi: 10.1007/978-3-642-33868-7_47/COVER.

[6] O. Borkovska, J. Gill, L. Pistolesi, J. Squires, and G. Yetman, "Mapping the Distribution of Human Population: The High Resolution Settlement Layer (HRSL) Analysis Country Global Urban Footprint Global Human Settlement Layer High Resolution Settlement Layer", Accessed: Jan. 24, 2023. [Online]. Available: https://sedac.ciesin.columbia.edu/binaries/web/global/news/2018/borkovska_hrsl_ob_gy_v3.pdf

Index

Figures Index

Tables Index

Chapter 1. Introduction

Humans have long relied on energy provided by various sources, ranging from heat obtained from burning logs of wood to harnessing the mechanical power of animals or natural elements like water currents and wind. In contemporary times, the sources of energy have become varied, with oil, coal and gas being the primary contributors. These three sources combined account for more than 78% of the global energy production [1]. Moreover, the global consumption of energy was 28,466.3TWh (year 2021)[2]. Notably, countries within the Organisation for Economic Cooperation and Development (OECD), comprising some of the most developed countries, consume almost 40% of this energy, that is 11,210.2 TWh [2].

All this energy is used in different ways and to different objectives. One of the most notable, and the one that is most relevant to this project, is the use of the energy in the form of electricity.

Electricity is an essential element of modern life, playing an important role in powering houses, businesses and industries around the world. It has become normal to expect that when a button is pressed a light will turn on or a machine will start working. However, we take for granted that electricity is always available, but unfortunately that is not the case in developing countries who lack access to reliable electricity.

This problem is mainly exacerbated in rural areas where it is more expensive to bring cables to, more so than to densely populated areas.

The use of electricity can improve the lives of the people without access to it in various ways, from the possibility of lighting their houses and streets at night to the preservation and cooking of food.

It also improves the economy of the region where the electricity infrastructure is well-established. In fact, according to The World Bank the OECD's GDP was 58,268,353.65 million dollars in 2021, that is close to 60% of the global GDP (which was 96,527,425.92 million dollars in 2021)[3][4]. Taking into account that the member countries of the OECD are some of the most developed countries in the world and that it does not include other countries such as China, which had the second largest GPD in 2021 (17,734,062.65million dollars [3][4]), that leaves the developing countries with less than the 22% of global GDP in 2021. There is not only one factor that can be singled out as the cause of this difference in the economic performance, but electricity access and reliability is certainly playing a part in it [5].

Indeed, the percentage of people that have access to electricity in low income countries is far lower than that of countries with high or middle income. Low income countries can only provide electricity to 41.4% of their population while middle and high income countries provide electricity to 94% and 100% of their population respectively.[6] To put it another way, the OECD member states provide electricity to 99.9% of their population while only 90.4% of the world population has access to electricity.

Some countries, knowing the advantages it provides, are accelerating the electrification of their countries. A great example is India, which has gone from 58.7% of its population having access to electricity in the year 2000 to 99% in 2020, based on the information provided by The World Bank.[6]

As well as countries, many institutions and organizations are developing different tools to make the expansion of electrification possible at the optimal cost. This project has been made in collaboration with IIT-Comillas ("Instituto de Investigación Tecnológica" in Spanish, Institute for Research in Technology, IRT-Comillas in English), which is part of the Universal Energy Access Lab, that aims to make access to electricity for everyone by creating tools that plan detailed electrical networks in developing countries [7]. Academics, including engineers, working in this institute as well as other academics from other universities have developed REM [8] (Reference Electrification Model) which develops optimal electrical networks.

There are other technologies, programs and databases that can further accelerate the implementation of electrical grids in developing countries.

This project uses REM as the tool to assess the impact of building location accuracy on electrification planning. Besides this program there are other instruments and technologies relevant to the project.

Chapter 2. State of the art

In order to develop this project, there are three areas of study that need to be taken into account: Geographic Information Systems (GIS), electrification systems and random points distribution methods.

2.1. Geographic Information Systems (GIS)

Geographic Information Systems (GIS) is a computer system for storing, managing, analyzing, editing, and displaying data related to positions on Earth`s surface [9]. GIS allows individuals and organizations to work with spatial patterns layering different information so it is more manageable. There are many ways of getting the data to employ in GIS, there are online sources of maps and different kinds of data such as shapes of countries, rivers and road tracks and it can also be added manually or through programming.

2.1.1. Geolocating buildings using satellite imagery

Various initiatives have been dedicated to the creation of software programs designed for building identification. These efforts involve a variety of methods, ranging from the application of machine learning techniques to the implementation of advanced image recognition algorithms.

Some have used LIDAR to obtain two masks with which they can recognize void areas and filled areas and multispectral imagery to find the limits of the buildings.[10]

Others have employed modified satellite images to identify shadows. To do so they have taken only certain frequencies of the images as they are the ones that show shadows more accurately.[11] With these shadows and other information, building locations can be obtained.[12]

Using machine learning, more specifically support vector machines (SVM), researchers managed to recognize man-made objects from high resolution images. For this, images of resolution better than 5 meters were used and several classes of objects could be classified from them.[13]

Another interesting idea to classify buildings from aerial images is to assume that the buildings have a convex rooftop as other team of researchers do in "Automatic Vegetation Identification and Building Detection from a Single Nadir Aerial Image".[14]

Using a variety of machine learning methods on high resolution satellite images the database of Open Buildings, covering Africa and parts of Asia, was obtained [15]; furtherly developed to improve the results and expand the area explored to south America an other regions.[16]

2.1.2. Demographics by region data

The High Resolution Settlement Layer (HRSL) provides estimates of human population distribution at a resolution of approximately 30 m. The settlement extent data were developed by the Connectivity Lab at Facebook using computer vision techniques to classify blocks of optical satellite data as settled (containing buildings) or not. Satellite imagery from DigitalGlobe was used. The Center for International Earth Science Information Network (CIESIN) used proportional allocation to distribute population data from subnational census data to the settlement extents.[17][18][19]

2.2. Electrification systems

Electrification has been a topic for many studies and for the development of systems that calculate parameters related to the electricity network, such as the grid structure, demand and load.

Integration with GIS functionalities has been developed for electrical grid planning tools which enhances the capability to see and calculate the impact connection of renewable generation on the existing grid [20]. Simplified dimensioning for photovoltaic and wind system, as well as the management of the systems, has also been developed with modules that also use GIS to visualize the information.[21]

Optimization of the tools to generate optimal electrification projects has been widely investigated. The design of minimum cost wind-photovoltaic electrification systems that combines micro-grids and stand-alone systems was improved using "two heuristics calculated from opposite and extreme solutions: a single micro-grid connecting all points and a solution with all stand-alone systems"[22].

2.2.1. About REM

The Reference Electrification Model (REM) supports large-scale electrification planning and local (smaller-scale) electrification projects. The model requires inputs about building locations, solar irradiance, topography, grid extent and reliability, expected consumer demand, fuel costs, and infrastructure costs. After running a series of clustering and optimization algorithms specifically designed for electrification planning, REM produces lowest-cost system designs. A user can then tune input parameters to

enable sensitivity analyses and derive policy-driven insights or entrepreneurially-driven plans. [15]

2.3. Random point distribution methods

The distribution of geographical points is an important part of developing an electrical grid. Normally these points are taken from a database where the information is available or studies are carried to know the location and the specific requirements of each of them.

As there is no literature of methods to create random points to use in tools that generate electrical grids, a study of different methods to create points randomly in different surfaces is studied.

The main focus of these studies is developing efficient methods to generate random points. Algorithms to create random points uniformly distributed in a hyperellipsoid have been optimized and applied to Monte Carlo simulations.[23]

The study of the generation of random points in different surfaces is a constant in the literature, covering Bezier surfaces [24], hyperellipsoids [25], tetrahedrons [26] and circles [27].

Chapter 3. Project definition / Developing the system

The developed system consists of a generator of geographical points, in the form of geographical coordinates, that represent buildings. These generated points can be used in the development of an electrical grid and for other purposes where the location of buildings is essential.

In this case, REM tool, which calculates an electrical grid, is used to evaluate if the geographical points generated can replace more precise geographical points obtained using other methods.

3.1. Objectives and specifications

The objective of the developed system, called Sprinkle, is to generate geographical points to be used in the generation of an electrical grid. These points will represent the location of a building.

These generated points are an output of the system in the form of a CSV file that follows the structure required for the inputs in REM. REM is the tool used to evaluate if the data generated by the developed system is equivalent to the data obtained using another method such as selecting the location of the building based on satellite images.

3.2. Data used

The data used ranges from the data used to develop the system to the data needed to run the system and to evaluate the system's capabilities.

3.2.1. Data used to develop the system

The data used to develop the system are MATLAB functions, including those functions created by other users, such as a function that exports the created geographical points to a KML file. (This allows the representation of the points in GIS, in a map like Google Earth.)

3.2.2. Data used to run the system

To run the system some inputs are necessary. These are required to make the generated points as reliable as possible.

3.2.2.1. HRSL

The HRSL stands for High Resolution Settlement Layer and is a database that has the information of whether there is a settlement or not in a region. The database divides

the surface of the Earth in squares of a size around 30 by 30 square meters, and with satellite images and using computer vision classifies if each of them contains buildings or not.

The HRSL database is divided into many TIF files covering each one different areas of the world and it does not cover all the countries. The countries covered are mainly developing countries as these are the ones where this data could be more useful. These files are updated from time to time; therefore, the current files could be based on old satellite images.

3.2.2.2. Polygon

In order to define the area where the geographical points have to be created, a polygon that covers the area has to be drawn to use as an input to the system. The polygon is created using Google Earth as a KML file.

3.2.2.1. Number of buildings in the area

To distribute a correct amount of buildings in the area that are compatible with the amount of buildings obtained by other methods, the number of geographical points (that represent the buildings) need to be provided.

3.2.3. Data used to evaluate the system

To evaluate the system the REM tool is used. This makes it necessary that the data used as an input to REM has to follow certain format. This format has to be the same for all inputs.

As the impact of the new method of obtaining the geographical points for the buildings needs to be measured, the REM tool has to be used with more reliable information that allows to make a suitable comparison. For this reason, three other methods to obtain the geographical points have been analyzed: manual method, buildings databases and random distribution of points covering all the area inside the polygon.

3.2.3.1. Manual data

This data has been obtained by looking at satellite images and manually marking the points where a house was identified. This method has been used in a small area with a low number of houses because it is not a very efficient system.

3.2.3.2. Building databases

There are some organizations that have created databases that include information about the location of buildings. To do this many methods have been used, such as image recognition software using satellite images or using satellites with LIDAR technology.

This information has a high level of accuracy but is not as precise as manual data. This is used for bigger areas where the manual data would take too long to obtain.

3.2.3.3. Random distribution of points

This method considers that the distribution of houses in the area is uniform along all of it and does not take into account any other information. This is the least reliable method as it relies on the polygon created around the area to be close to the limits of a settlement where there are houses. This method can work well for small areas with equally distributed buildings and a well-defined contour.

3.2.4. Regions to analyze

To determine the region or regions where the system is to be tested there are different details to be considered, mainly geographical data of the region: information about the number and location of buildings and the availability of the HRSL database of the region.
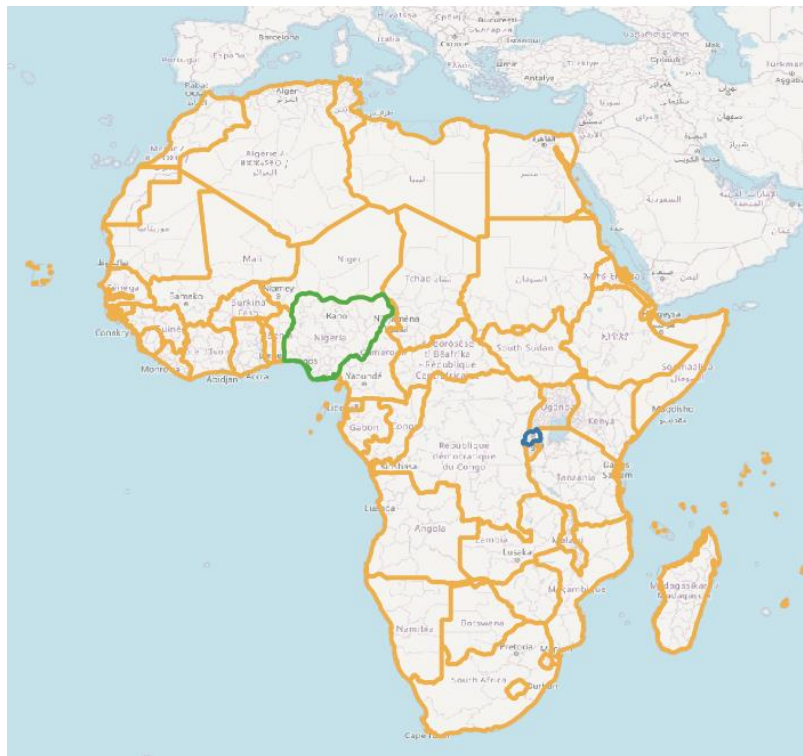
Figure 1: Map of the regions where the system is executed; Marked in green Nigeria, where Zugu is analyzed; marked in blue Rwanda where Huye district is analyzed

Taking into consideration this parameters, two regions in Africa have been selected: Zugu in Nigeria and Huye in Rwanda. Zugu is a town in the Sokoto region in the north west of Nigeria. This helps observe the impact of the method developed in small areas.

Huye is a district in the Southern Province in Rwanda. It allows to study the impact of the method in bigger areas.

3.3. Algorithm

The algorithm employed in this project consist in two main phases:

1. Generation of the geographical points using the developed system.
2. Execution of the REM tool to assess its impact on the results.

To compare the results, geographical points generated manually, randomly and those taken from buildings databases are also used as input in REM. The results obtained from these serve as a benchmark to which the developed system can be compared.
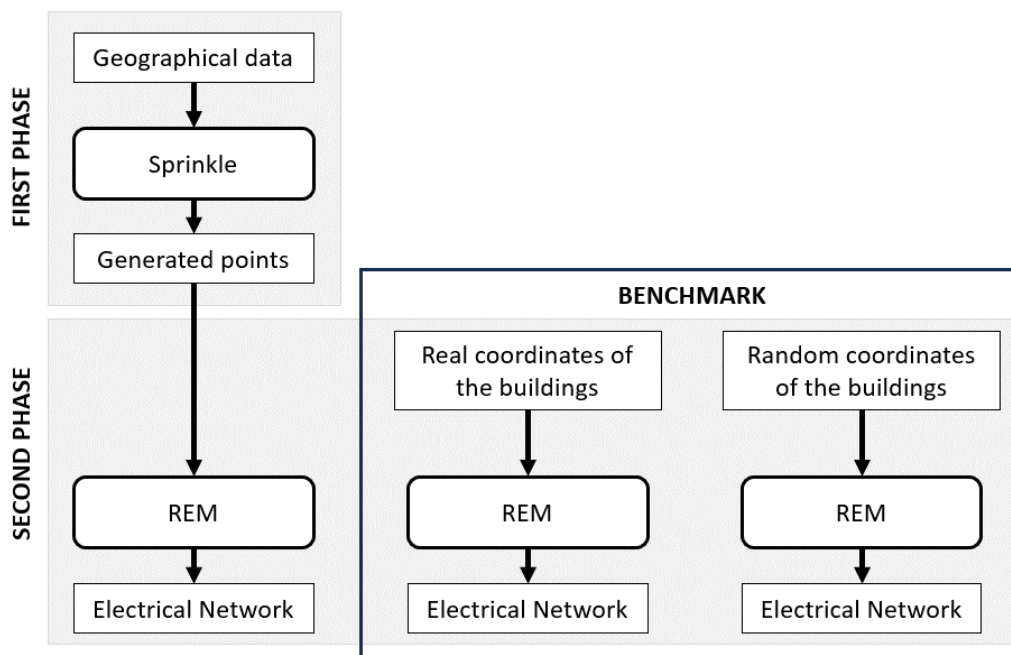
Figure 2: Algorithm followed in the project

3.3.1. Distribution of the geographical points

The developed system, that will be also known as Sprinkle in this project, generates the coordinates for geographical points that represent buildings.

There are three main inputs:

- The HRSL of the region. This is introduced as a TIF file and it has to be the file of the region analyzed.
- Polygon of the area where the geographical points are going to be distributed. This is created using Google Earth and encompasses the area where the electrical grid is going to be developed.
- Number of buildings that is going to be distributed by the program. This is done to make the results comparable with other available methods.

The fourth input is the name of the output files.
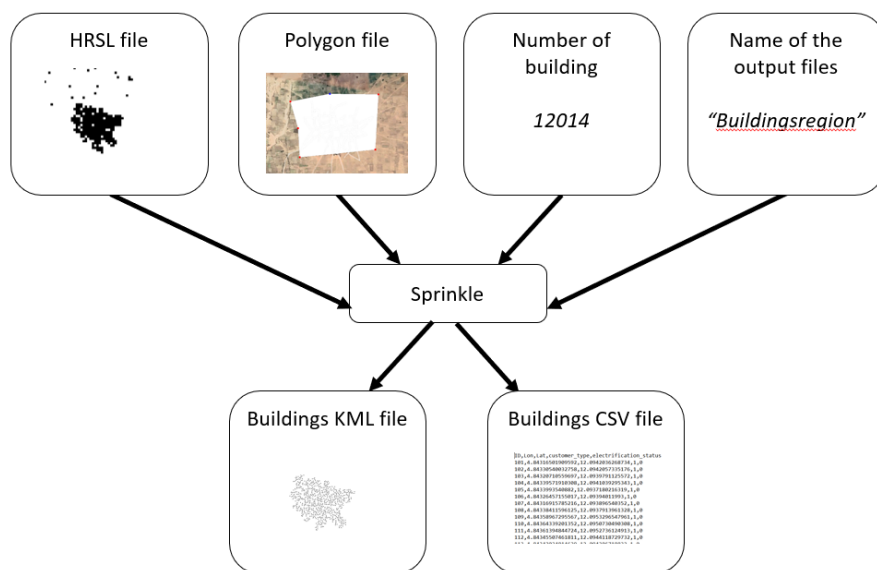
Figure 3: Inputs and outputs of Sprinkle

With these four inputs the function, which is included in ANNEX II, follows five steps:

- Obtain the data from the HRSL file.
- Obtain the squares inside the polygon where there is population.
- Distribute the random points inside the obtained squares.
- Delete the excess points.
- Export the data of the sprinkled points in CSV and KML files.
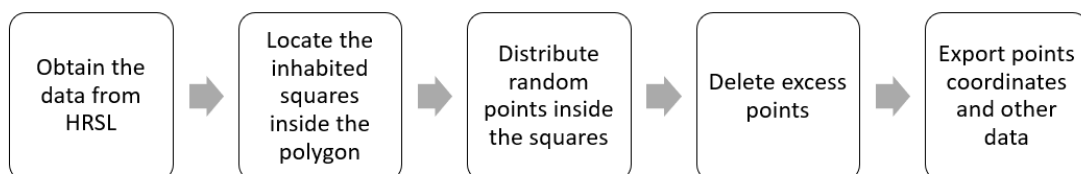
Figure 4: Steps followed by Srinkle, the function created to generate geographical points.

3.3.1.1. Obtain the data from the HRSL file.

The HRSL file comes in a TIF format. TIF stands for Tagged Image File, and is a type of image file which in this case contains the information of the coordinates of the corners of the image and the geographical dimensions of the pixels.

This type of files can be opened with MATLAB using the function "readgeoraster". The outputs of this function are a matrix with the values for each of the pixels in the image (also known as colormap) and a second argument containing spatial referencing information for the matrix, as shown in Figure 5.

Property ▲	Value
RasterInterpretation	'cells'
XIntrinsicLimits	[0.5000,4.6617e+04]
YIntrinsicLimits	[0.5000,3.2906e+04]
CellExtentInLatitude	2.7778e-04
CellExtentInLongitude	2.7778e-04
LatitudeLimits	[-4.9221,4.2182]
LongitudeLimits	[28.1251,41.0740]
RasterSize	[32905,46616]
AngleUnit	'degree'
ColumnsStartFrom	'north'
RowsStartFrom	'west'
RasterExtentInLatitude	9.1403
RasterExtentInLongit...	12.9489
CoordinateSystemTy...	'geographic'
GeographicCRS	1x1 geocrs
AngleUnits	'degrees'

Figure 5: TIF file properties

From this data obtained there are two items to be used afterwards: the array containing the information of each pixel and the "refmat" matrix which is a reference matrix that contains the main geographical information about the TIF file.

3.3.1.2. Obtain the squares inside the polygon where there is population.

With the information obtained in the previous step the function "get_active_squares_hrslf.m" creates squares that surround the areas where the HRSL file states that there are settlements if they are inside the polygon created to cover the area where the geographical points need to be created.

The function first finds all the squares where there is a settlement, then it loads the polygon file and analyzes which squares with settlements are inside the polygon using the function "inpolygon".

The final steps are to count the number of squares inside the polygon with settlements and then a loop that marks the edges of each square to store it in an array and also exports it as a KML file.

3.3.1.3. Distribute the random points inside the obtained squares.

A function called "sprinkle_randomf" has been created to distribute the random points inside each square. The name sprinkle comes from the act that the points created are scattered over an area where the buildings are expected to be.

The function takes the number of squares and their coordinates from the previous step. With this information and the number of buildings/geographical points, that need to be distributed, a ratio of house per square is calculated. This number is rounded up so the number of houses is always equal or greater to the required number of buildings.

In a loop, a number of points equal to the ratio of houses per square are distributed inside each of the squares and these points are stored in an array.

The number of geographical points is always equal or greater than the number of buildings used as input, therefore some of them have to be deleted.

3.3.1.4. Delete the excess points.

The excess points created need to be deleted. This can be done with various methods such as deleting the last added points or deleting points randomly. This two options can alter the results severely, for instance when deleting the last added points it can create a gap in an area due to the way the points are created in the previous steps. Taking this into account and given that REM, the tool used to evaluate the system, considers points within a distance of less than 3 meters to be the same, a third method has been developed. This method calculates which are the closest pairs of points and deletes one of them.

The developed function, called "DeleteExtraPointsDf", initiates by calculating the distance between one of the points and all the others. To be able to find the closest points among all the available points a loop is created. In this loop:

- First a copy of the array of the previously generated points is created and the first point in it is eliminated in the copy. In the second iteration of the loop the second point is deleted; in the third iteration the third point and so on.
- Then "delaunayn" function is used on the array with the deleted point. This function calculates the Delaunay triangulation, which maximizes the minimum

of all angles of the triangles in the triangulation; this step is not essential but reduces the time required to execute the next steps.

- The deleted point from the copied array is defined as the "queypoint"; meaning that the distance between this point and all the rest will be measured.
- "dsearchn" function is used with the copied array, the Delaunayn triangulation and the "querypoint" as inputs and the output obtained are the position in the array of the point closest to the "querypoint" and the distance between them.
- The closest point to the "querypoint" and the distance between them is stored in an array "closestp".

The loop is done for all the points created.

When the array "closestp" is completed it is put in order from smallest to biggest distance.

The closest points are deleted until the amount of points is equal to the required amount of points. The points in the list must not be consecutively deleted because then the two points that are closest would both be deleted and deleting just one would be enough to make that small distance disappear.

Finally the array of the points is the output of the step.

3.3.1.5. Export the data of the sprinkled points in CSV and KML files.

When the final points have been obtained they need to be exported as a CSV and as KML; the CSV file is used as input to the tool that evaluates the results and the KML file allows the visualization of the points on a map.

The CSV file is built in a loop and using the function "writetable". It is important to note that the file need to follow a structure that contains:

- ID, this is an identification number which is created from 100 upwards.
- Longitude, is the geographical coordinate that specifies east-west.
- Latitude, is the geographical coordinate that specifies north-south.
- Customer Type, describes the type of consumer that is connected to the point.
- Electrification status. Describes if there is already a cable to the house, that is, if there is need to be electrified and so calculated by REM. If it is 0, it means that there is no electricity and REM will use it to calculate the grid.

3.3.2. Evaluating the results using REM

To analyze how well the system developed works compared with other systems available the results obtained are used as inputs in REM; this allows for a better analysis as the results are obtained in dollars and in kilometers of electric wires (Network length).

3.3.2.1. Points in REM

In order to get the generated data of the points to REM and be able to execute it the information has to follow a structure.

Points are given in a CSV file where the information has to be in the following order:

- Number ID: an identifier for the building
- Longitude: coordinate, measured in degrees, that specifies east-west position
- Latitude: coordinate, measured in degrees, that specifies north-south position
- Customer type: different patterns of electricity consumption can be created and then applied to different households; for simplicity all buildings will be the same customer type.
- Electrification status: specifies if the house has already access to electricity; 1 if it does, 0 It does not. All buildings are assumed to lack access to electricity in this study.

3.3.2.2. Other inputs

There are many inputs that can be modified in REM depending on the electricity consumption curve, the type of cables, the existing grid, the consumption per house…

In this case most of the data has been kept equal and constant to make the comparison more standardized. For this project, the only data that has to be changed, when calculating in a new area, is the existing grid and the coordinates to the region.

3.3.2.3. REM execution

REM is executed in a server provided by IIT-Comillas which contains the code and can be operating and calculating continuously, without interruption. The calculation of the network takes time as it is an iterative process trying to optimize the electrical grid.

3.3.2.4. REM results

When REM finishes executing it generates a folder with different files, including KML files with the representation of the grid and various CSV files with different information about the costs and other parameters, like power factors, contracted power and

Network length. A summary of all the different parameters available is generated both TXT and HTML format for its visualization.

These results are analyzed using Excel and MATLAB to visualize the dimensions and costs of the electrical network. To see the distribution of the network on a map GIS software, such as QGIS or Google Earth, can be used.

3.3.3. Creating the benchmark

In order to establish the references for comparison with the developed system, it is essential to generate two datasets that serve as inputs for REM. The first dataset is selected as a pseudo-perfect case, where all the coordinates of the points are taken as highly accurate. The second dataset is generated randomly using a system less advanced than the one developed for this project. This last dataset is considered to be less accurate than either of the other methods discussed, but it needs to be verified once the results are compared.

The first kind of dataset is taken either manually, from satellite images, or from a building database. The coordinates of the points are not 100% accurate but are as high as the dataset can get.

The second dataset is created using the function "sprinkle_random_totalf". This function has similarities with the function "sprinkle_randomf", explained in 3.3.1.3, with the difference that this simpler one disregards the squares obtained from HRSL,. and disperses the points in all the surface inside the polygon.

The datasets are then inputted in REM and the results are used to assess the quality of the system developed.

3.4. Implementation

To evaluate the system created, two implementations have been carried out. First one in a small area of Nigeria, in a town of less than one thousand buildings called Zugu. Then, in a larger area of Rwanda with more than one hundred thousand buildings, in the department of Huye.

3.4.1. Zugu, Sokoto, Nigeria

The system is first tried with a small region of Nigeria where the information gathered manually is available so it is possible to have a basis for the development of a greater area.

3.4.1.1. Location

The location selected is Zugu located in the northwest region of Nigeria called Sokoto.

Figure 6: From left to right, up to down: Image of Nigeria in Africa; Sokoto in Green inside of Nigeria; Zugu in red inside of Sokoto; Zugu with a black polygon marking the populated region

The interest in this area comes from the region being in a country where the electrification is still ongoing. It is also a place where the manual detection of the buildings has been made so there is no need to repeat it.
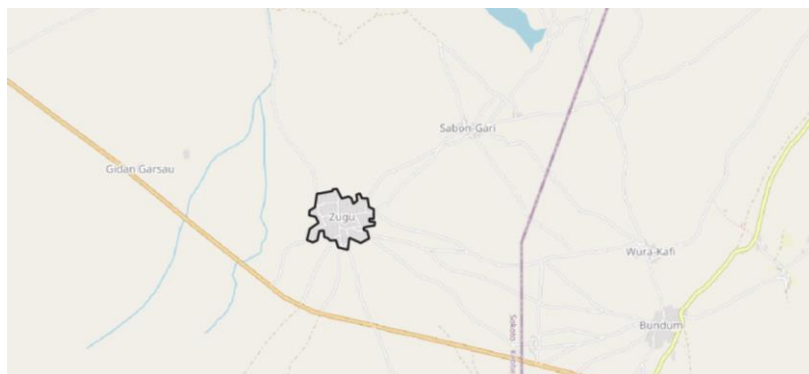
Figure 7: Zugu populated region marked with a black line

3.4.1.2. Data inputs

In this case, the geographical data regarding the location of the buildings is acquired using three methods.

- In the first method every point where a building could be seen in the town of Zugu has been manually marked. This can be seen in Figure 8.

- The second approach has been to distribute the points in a random way, where the input data was the number of buildings in the region and the limits of area where they are located.

- The third method is similar to the second method improved by the use of the HRSL data, which gives further constraints to the area where points can be scattered. This method is the system developed in this project and has been executed ten times to get different set of points to evaluate.

Figure 8: Points representing the buildings selected manually. The google earth satellite image is more recent than the selection of the points therefore the mismatch in the area reached.

3.4.1.3. Processing the data and getting the points that represent buildings

To distribute the geographical points representing the buildings, the data is processed as explained in section 3.3.1. of this project.

First, the information obtained from the HRSL is used to see where there are settlement.
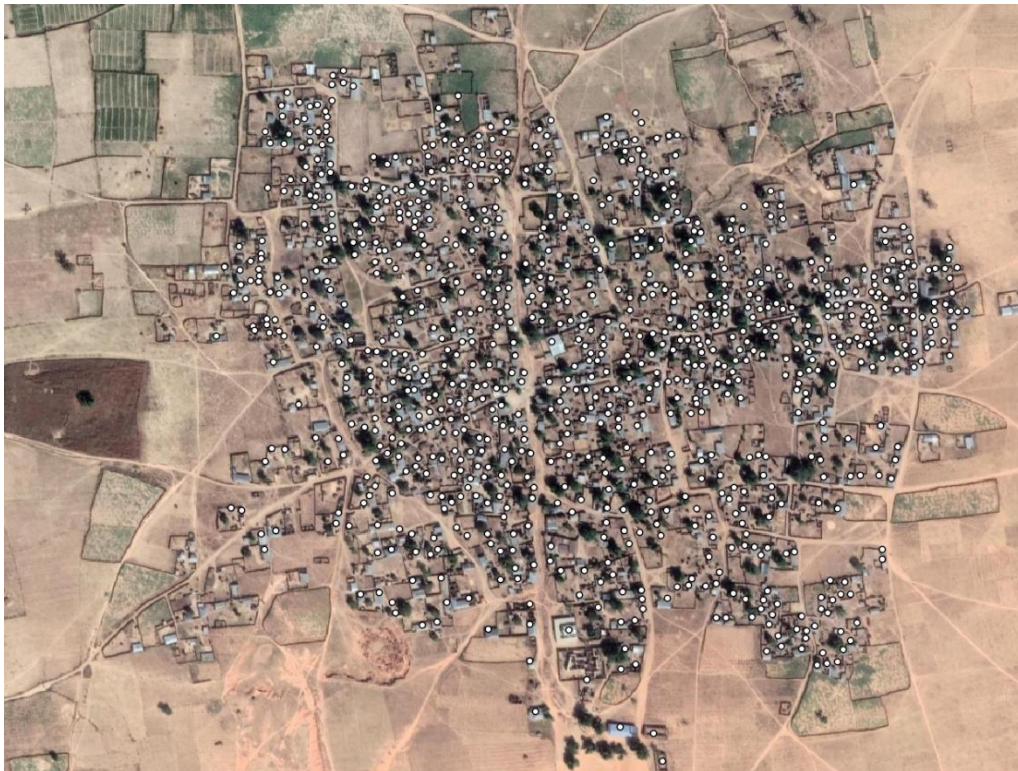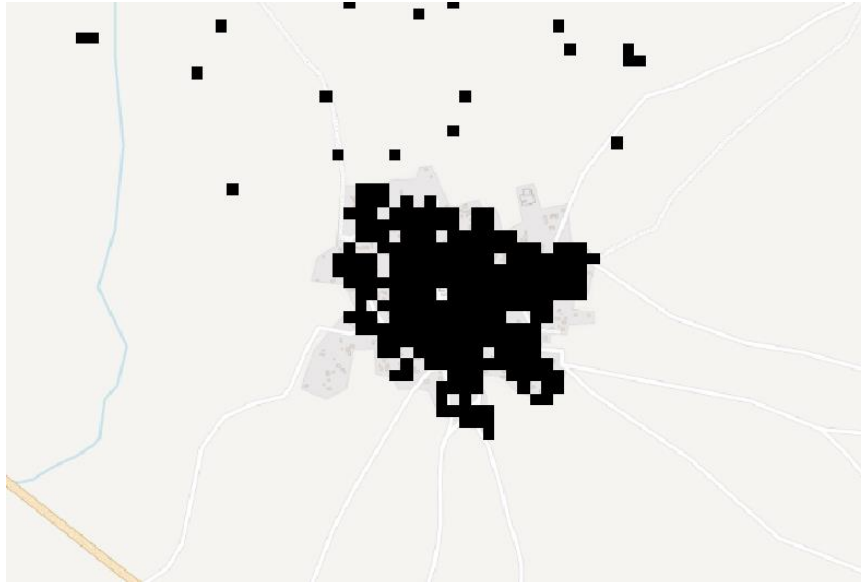

Figure 9: HRSL visualization on Zugu

- **High Resolution Settlement Layer (HRSL)** provides the information of where there were settlements (at least one house) at the time of analyzing the satellite images. The algorithm that detects the houses is not perfect and some trees and other objects were mistakenly detected as houses. Furthermore, the area of Zugu has changed and new houses have been constructed that can be seen in recent satellite images but that HRSL could not detect as they did not exist at the time of its creation.

A polygon, delimiting the area to be analyzed, is important to avoid distributing points in regions where it is known that there are no buildings and to mark the extent to which the electrical grid needs to be developed. In Figure 10, the polygon used to delimit Zugu, has been drawn; the black squares are the marks that HRSL gives to indicate that at least one house has been detected inside the square. Black squares outside of the drawn polygon will not be used for the calculation of the buildings location; the quality of the polygon drawn will impact the quality of the results. For this study, a simple polygon with just four edges has been used.
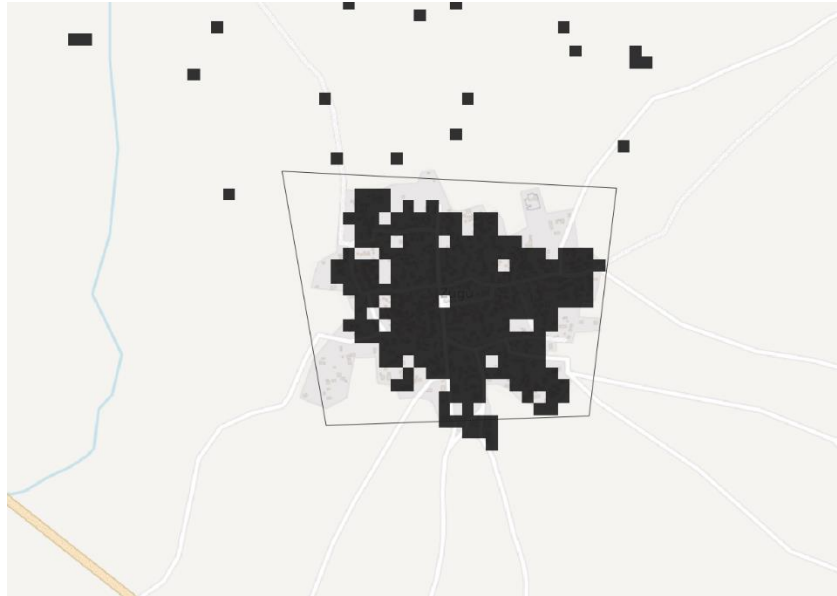
Figure 10: HSRL visualization on Zugu and Polygon

Afterwards, the system created selects the relevant squares for the calculation. These are the ones that are inside the previously created polygon. The calculated squares can be seen in Figure 11.
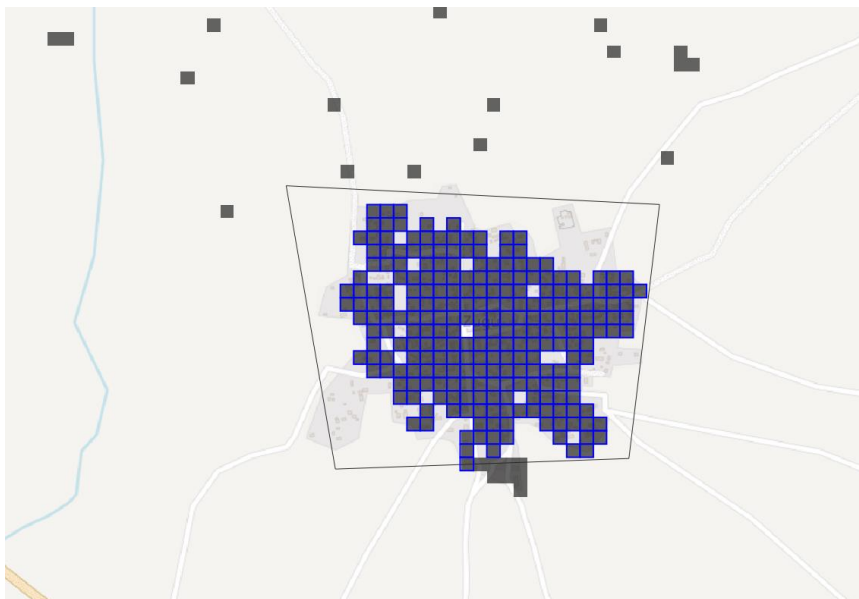
Figure 11: HSRL visualization on Zugu with marked blue squares inside the polygon

Finally, the points that represent the buildings are generated from the information of the number of houses and the location of the squares where there is a settlement. In Figure 12, these points, representing buildings, can be seen as green dots.
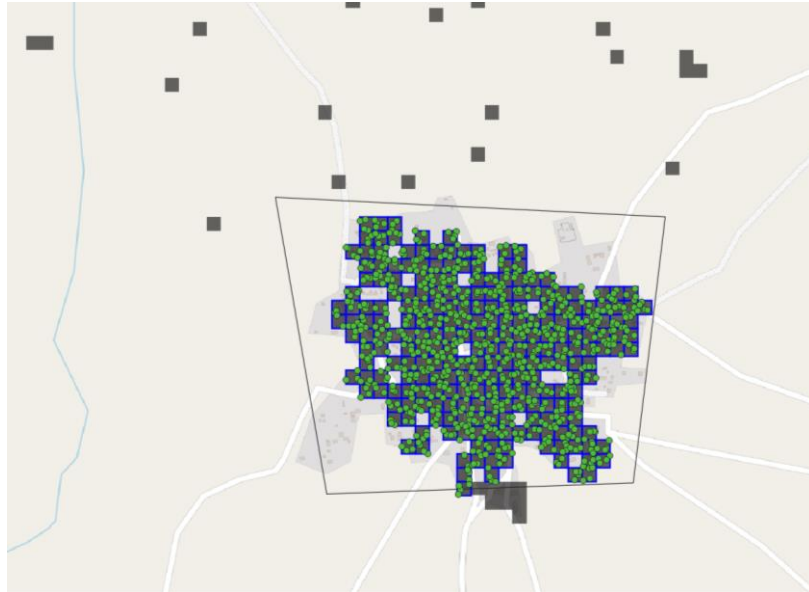
Figure 12: HSRL visualization on Zugu with generated green points representing the buildings

3.4.1.4. REM output

The Reference Electrification Model (REM) is executed with the information of the coordinates of the buildings. REM outputs many different files with varied information such as demand profiles, dispatches, cost parameters, the representation of the grid path and the network length, among other information. In Figure 13, the representation of the electrical network is shown in yellow while the buildings it connects are shown as small green circles.
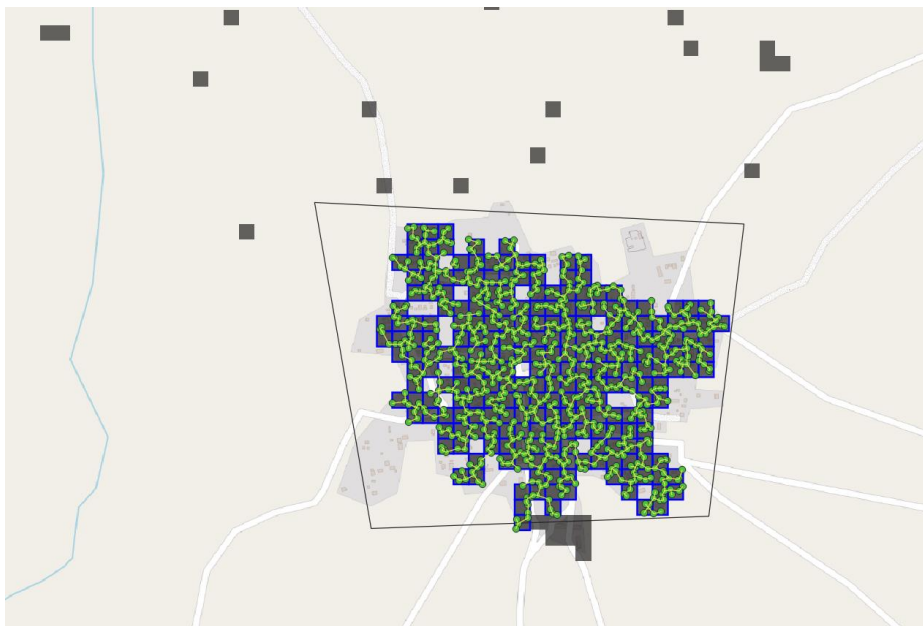
Figure 13: HSRL visualization on Zugu with generated green points representing the buildings and the electrical network

3.4.1.5. Analyzing the difference between the three methods

For this implementation, twelve results are obtained: one from the manual input, one from the fully random input and ten from the multiple executions of the developed system.

	Number of houses	Number of supply points	Network length (km)
Manual	898	898	10.65
Random	898	881	13.54
Sprinkle 1	898	897	10.45
Sprinkle 2	898	896	10.45
Sprinkle 3	898	896	10.48
Sprinkle 4	898	898	10.58
Sprinkle 5	898	898	10.56
Sprinkle 6	898	897	10.59
Sprinkle 7	898	897	10.59
Sprinkle 8	898	897	10.61
Sprinkle 9	898	897	10.50
Sprinkle 10	898	898	10.53
Sprinkle Average	898	897	10.53

Table 1: Results from REM in Zugu

In all these methods, the amount of houses/buildings used as input was 898, but if two points are too close by (within less than 3 meters from the closest point) REM would join them as one. Then new points, resulting from the concentration of the points in this situation, known as supply points (points to supply with electricity), are used by REM to generate the electrical grid. Therefore, there are always the same or less number of supply points than number of houses.
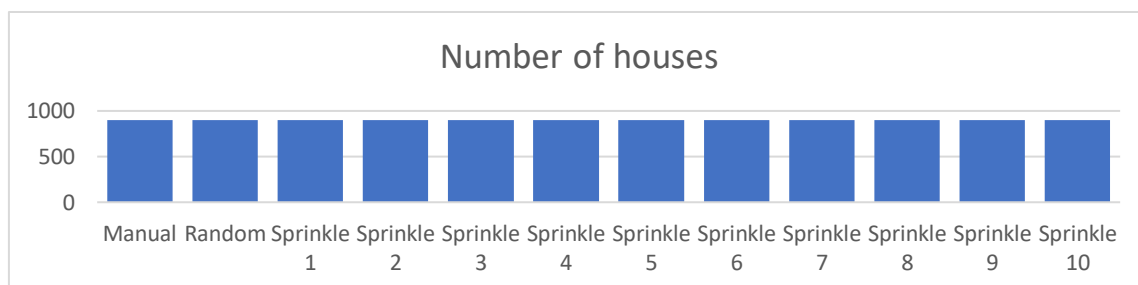
Figure 14: Number of houses for each execution in Zugu

Number of supply points

Figure 15: Number of supply points for each execution in Zugu

The first method, the manual one, is considered the most accurate approach to obtain the coordinates of the buildings in a region. It is time-consuming, as a person is needed to manually mark the location of the buildings. Despite being very accurate, it can still have errors from the person selecting the points and from the satellite images not being well georeferenced.

The second method, the random distribution of points inside the polygon, is the quickest and the one that needs the least information. The consequence of this is that the results are hardly optimized and the calculated costs rise.

The third method, the Sprinkle method developed in this project, is a compromise between the other two methods, trying to find the most accurate solution with the limited available information.

Comparing the manual method and Sprinkle method, the difference between them is quite small, as shown in the results given in Table 1. Looking at the network length results, the distance for the manual method is 10.65 km and for the average of all the executions of the Sprinkle is 10.53 km; there is only a 2% difference between them.

Comparing the manual method and the random method, the kilometers needed to create the electric network increase by 27%, from the 10.65 km in the manual method to the 13.54 km in the random method; so if the random method were to be used to study whether to develop the network the results would show a cost much higher than the real one.
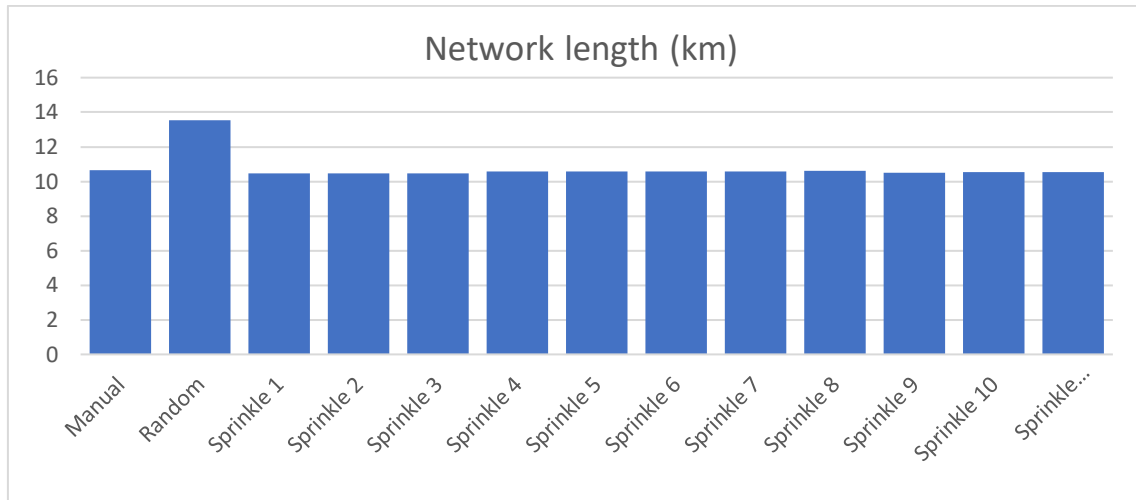
Figure 16: Network length obtained for each execution in Zugu

When comparing the results of using the Sprinkle method and the random method, that does not use the information of the HRSL data, it is possible to see in Figure 17 that, when using the polygon as the only boundary for the distribution of the points, the network expands to occupy all the area. When the points are scattered in a bigger area the network needs to expand to reach all the points, this leads to a 28% increase in the kilometers of cable required compared with the sprinkle method, from 10.53 km to 13.54 km.
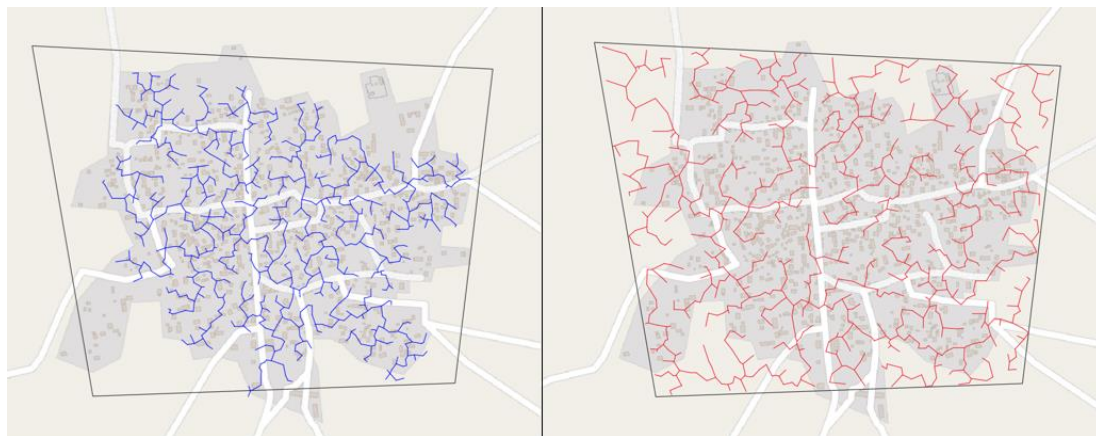
Figure 17: Zugu electrical network comparison between Sprinkle and random method; in blue the network created using Sprinkle's output and in red the network created using the fully randomized method of creating geographical points.

3.4.2. Huye, Southern Province, Rwanda

The method is further examined in a bigger region, this time located in Rwanda, where the information of the coordinates of the buildings is gathered from a building database.

3.4.2.1. Location

The location selected is Huye which is in the Southern Province of Rwanda.
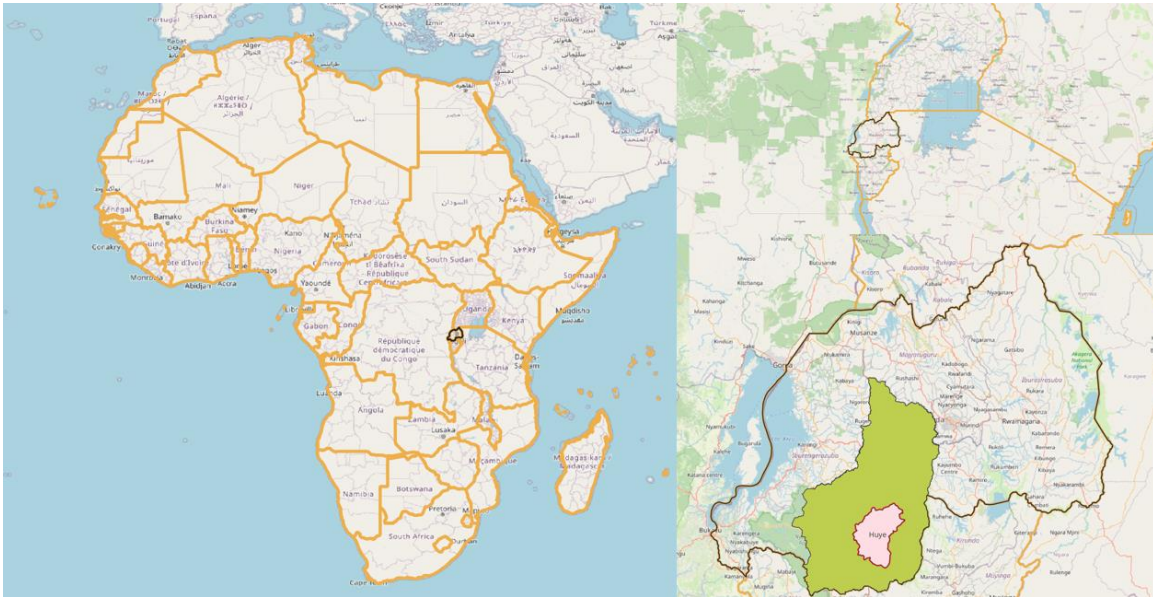
Figure 18: Image of Rwanda in Africa; Southern Province in Green; Huye in red inside of Rwanda

The study of this area is attractive because the country is still developing its electrical grid. The region and the country are suitable for the project as there are building databases available.

3.4.2.2. Data inputs

In this case the information of the location of the buildings is obtained using three different methods:

- First method uses an existing building database.
- The second approach has been to distribute the points in a random way, where the input data was the number of buildings in the region and the limits of area where they are located.
- The third method is similar to the second method improved by the use of the HRSL data, which gives further constraints to the area where points can be scattered. This method is the system developed in this project.

3.4.2.3. Processing the data and getting the points that represent buildings

To distribute the geographical points representing the buildings, the data is processed following the steps explained in section 3.3.1 of this project.

First the information obtained from the HRSL is used to see where there are settlements.
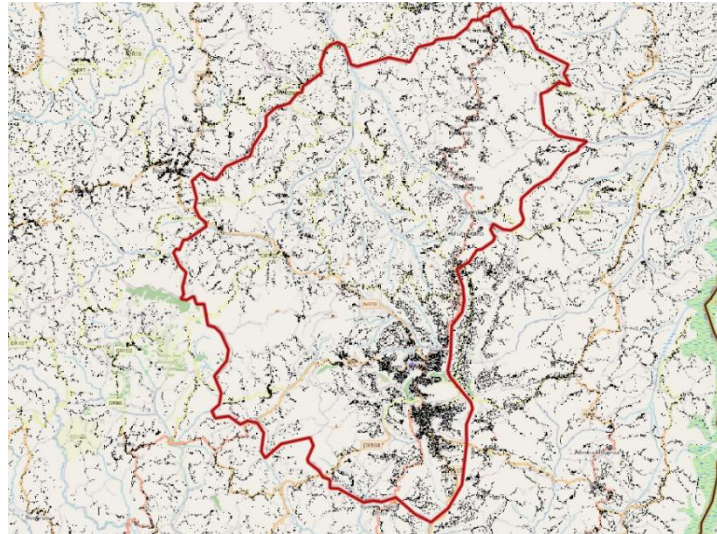
Figure 19: HRSL visualization on Huye, Rwanda; each black point represents a 30x30 m square where at least one house has been detected

- **High Resolution Settlement Layer (HRSL)** provides the information of where there were settlements (at least one house) at the time of analyzing the satellite images. The algorithm that detects the houses is not perfect and some trees and other objects were mistakenly detected as houses.

A polygon, delimiting the area to be analyzed, is important to mark the extent to which the electrical grid needs to be developed. In Figure 20 the black polygon, used to delimit Huye, has been drawn; the black squares outside of the polygon will not be used for the calculation of the buildings location. The polygon has been manually created and, as can be seen in Figure 20, it does not fit perfectly with Huye's shape marked in red. This has been made to keep the polygon as simple as possible.
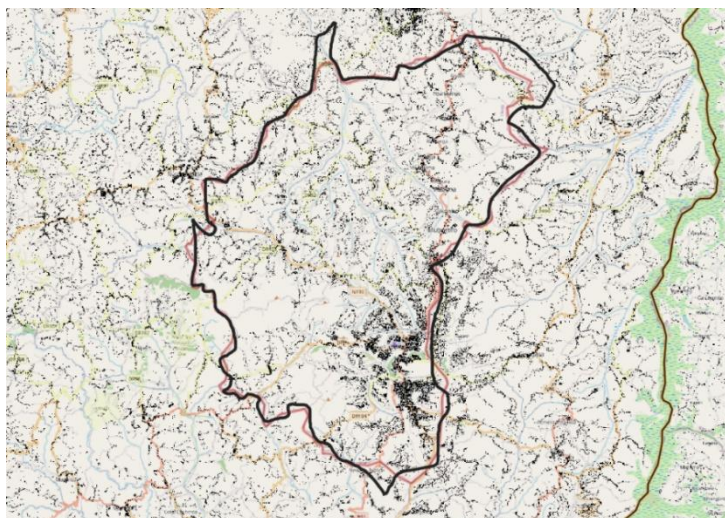
Figure 20: HSRL visualization on Huye and Polygon

Afterwards, the system created selects the relevant squares for the calculation, these are the ones that are inside the polygon. The calculated squares can be seen in Figure 21.
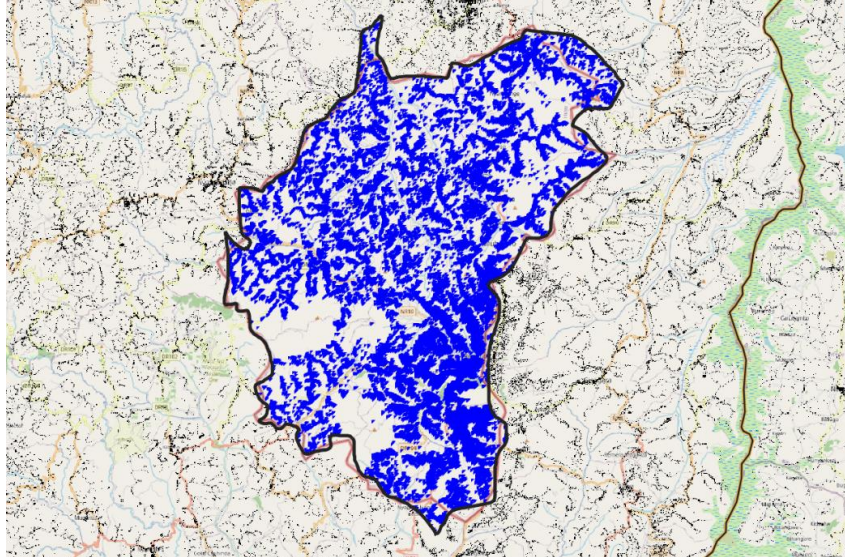
Figure 21: HSRL visualization on Huye with marked blue squares inside the polygon

Finally, the points that represent the buildings are generated from the information of the number of houses and the location of the squares where there is a settlement. In Figure 22 these points, representing buildings, can be seen as green dots.
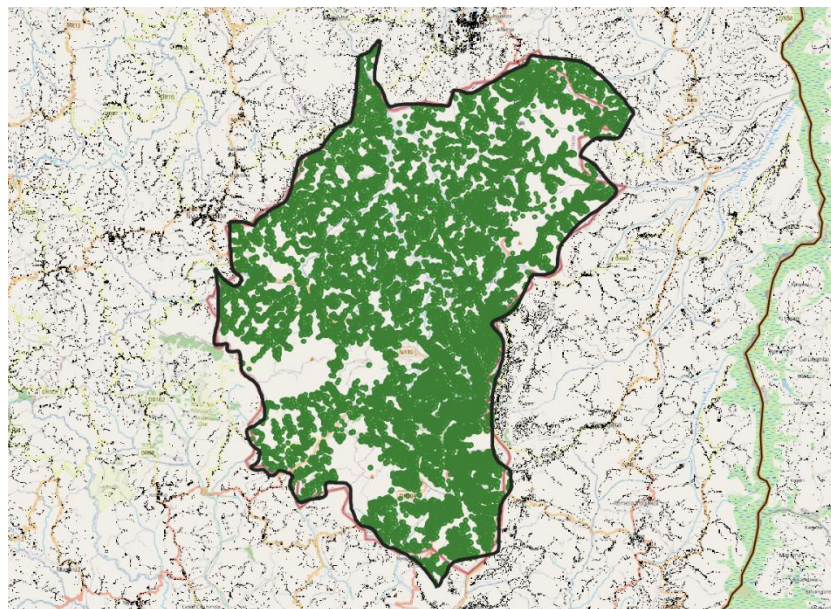
Figure 22: HSRL visualization on Huye with generated green points representing the buildings

3.4.2.4. REM output

The Reference Electrification Model (REM) is executed with the information of the coordinates of the buildings. REM outputs many different files with varied information such as demand profiles, dispatches, cost parameters, the representation of the grid path and the network length, among other information. In Figure 23 the representation of the electrical network is shown in yellow while the buildings it connects are shown as small green circles.
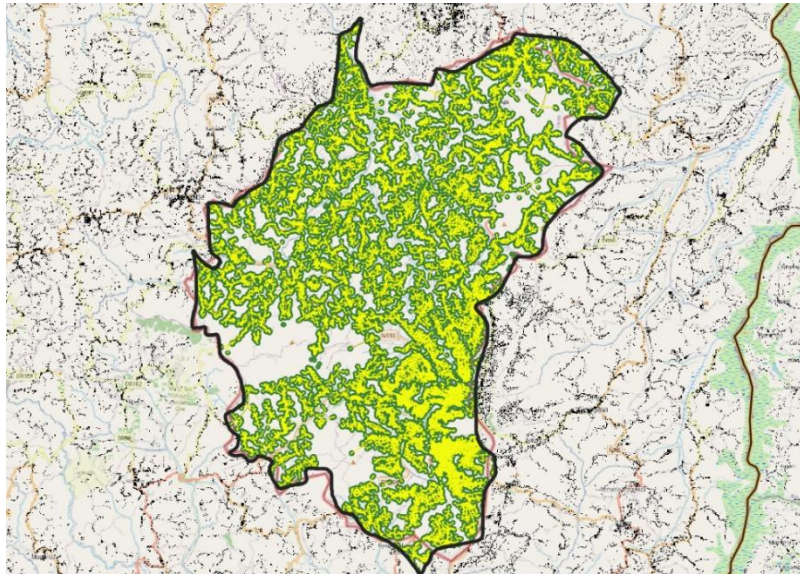
Figure 23: HSRL visualization on Huye with generated green points representing the buildings and the electrical network

3.4.2.5. Analyzing the difference between the three methods

For this implementation three results are obtained: one from the manual input, one from the fully random input and one from the executions of the developed system. There are two reasons why in this case only one execution of the developed system has been made; firstly, because the previous implementation (3.4.1.5) shows that the different executions of the developed system do not cause a great variation in the results; secondly, because the amount of buildings to be distributed in this case is higher than in the previous one (100500 compared to 898) and that has an impact in the time the developed system takes to generate the points.

In this case, being the area much larger than in Zugu (section 3.4.1) and having more buildings/houses, the network is divided into multiple low-voltage mini-grids connected by medium-volage links.

	Number of houses	Number of supply points	Network length (km) LV	Network length (km) MV	Network length (km)
Building Database	100449	99766	2967.60	485.20	3452.80
Random	99163	98949	5286.45	738.70	6025.15
Sprinkle	100479	100454	2821.87	508.77	3330.64

Table 2: Results from REM in Huye

In all these methods the amount of houses/buildings used as input was 100500, but if two points are too close by (within less than 3 meters from the closest point) REM would join them as one. Then new points, resulting from the concentration of the points in this situation, known as supply points (points to supply with electricity), are used by REM to generate the electrical grid. Therefore, there are always the same or lower number of supply points than number of houses.
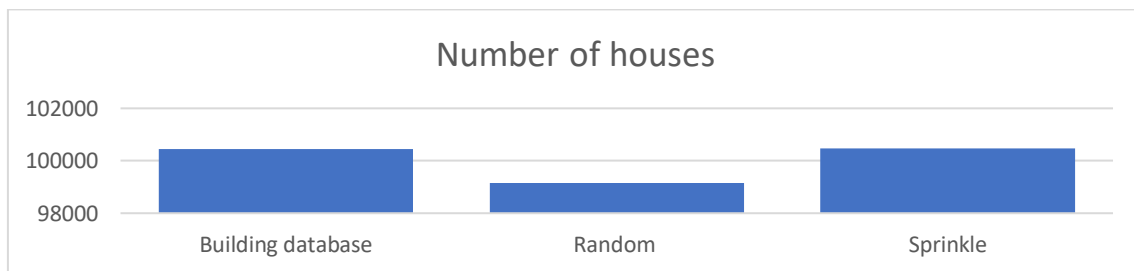
Figure 24: Number of supply points for each execution in Huye
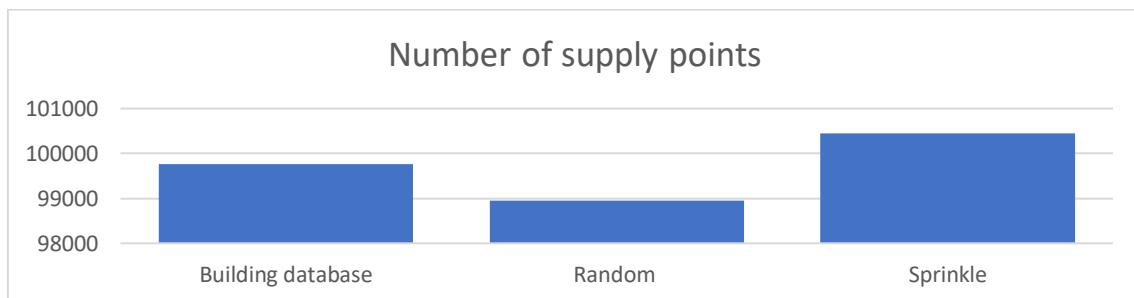
Figure 25: Number of supply points for each execution in Huye

The first method, that uses an existing building location database, is considered the most accurate approach to obtain the coordinates of the buildings in this region. Satellite images and machine learning have been used to detect the buildings. The location of the building is as precise as the system developed to obtain it. Despite being very accurate, it can still have errors from the software selecting the points and from the satellite images not being well georeferenced or being outdated.

The second method, the random distribution of points inside the polygon, is the quickest and the one that needs the least information. The consequence of this is that the results are hardly optimized and the calculated costs rise.

The third method, the sprinkle method developed in this project, is a compromise between the other two methods, trying to find the most accurate solution with the limited available information.
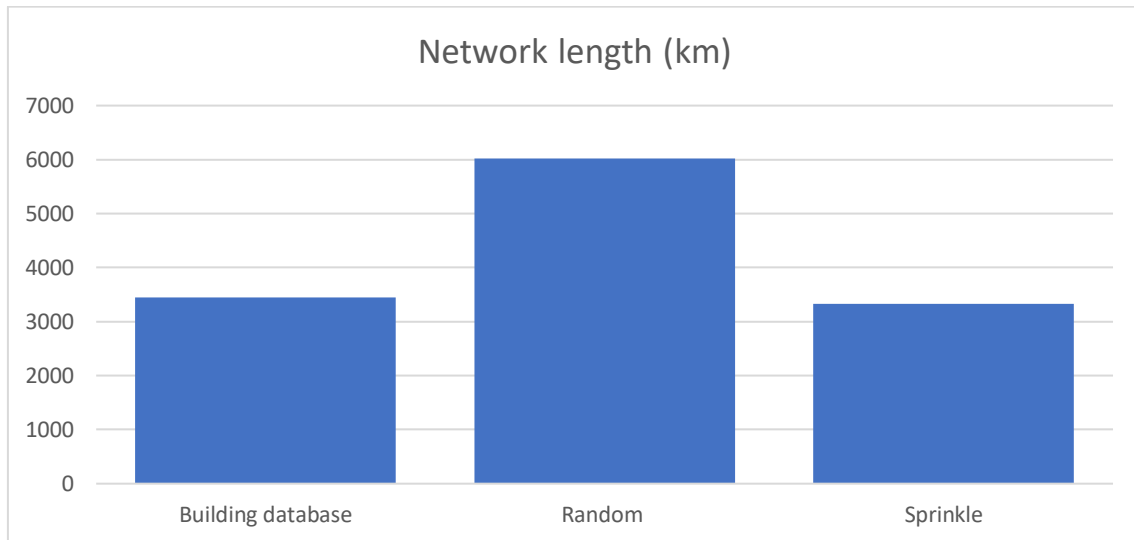
Figure 26: Network length obtained for each execution in Huye

Comparing the method that uses the buildings database and the Sprinkle method, the difference In the network length is relatively small, as shown in the results presented in Table 2. Taking a closer look at the network length results, the distance is 3452.80 km for the method using the buildings database and 3330.64 km for the Sprinkle method; there is less than a 4% difference between them.

Comparing the method that uses buildings database and the random method, the kilometers to create the electric network increase by 75%, from the 3452.80 km in the method with the building database to 6025.15 km in the random method. The random method overestimates the kilometers of cable needed, oversizing the network and the estimated costs to develop the infrastructure, making it harder to receive the required funds.

When comparing the results of using the sprinkle method and the random method, that does not use the information of the HRSL data, it is possible to see in Figure 27 that, when using the polygon as the only boundary for the distribution of the points, the network expands to occupy all the area. The network's length increases 80% from 3330.65 km in Sprinkle method to 6025.15 km in the random method.
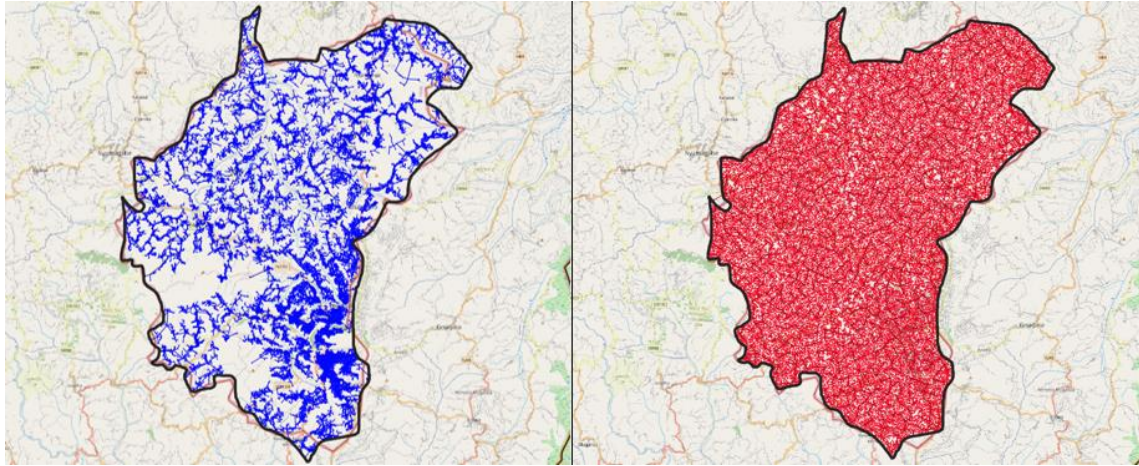
Figure 27: Huye electrical network comparison between Sprinkle and random method; in blue the network created using Sprinkle's output and in red the network created using the fully randomized method of creating geographical points.

Chapter 4. Results analysis

The results, obtained from the implementation on two different cases of the developed system (Sprinkle) and two other methods of sampling buildings for electrification planning, helped determine the impact that the method created has. The results for each case have been discussed in sections 3.4.1.5 and 3.4.2.5. Studying the combination of these results broadens the analysis of the feasibility of Sprinkle.

The first case was a small-scale examination to determine the viability of Sprinkle and, as seen in section 3.4.1.5, it was mostly satisfactory, obtaining very similar results between the developed system and the method that used the exact location of the houses.

The second case was an evolution of the first one, studying a larger area. The necessity to distribute more points made the process of distributing the points in Sprinkle much harder and computer intensive, allowing for just one execution of the method. The fact that it was only possible to obtain one set of locations is not relevant because, in the first case, the variance among all the ten executions of Sprinkle was 0.0033.

Another interesting result obtained from the analysis is that, as Sprinkle creates more points separated further apart than 3 meters than the other methods due to its design that aims to have the least amount of cojoined points, the network created by REM, using as input the Sprinkle points, has more supply points than the ones created by the other methods, excluding the manual one, which is the most precise method.

Using a random generation of points in an area to generate an electrical network is not efficient, generating up to 75% bigger networks than those networks created using data obtained from satellite images. This difference is lower when studying small areas and when the polygon that delimits the area is as close to the limits of the inhabited area as possible, but still 27% bigger than the results using data acquired from satellite images.

Sprinkle's network length is shorter in average to the length of the network created with the more reliable locations. This means that, when using Sprinkle method, a factor should be used. Seeing that in the first case studied the network length, using Sprinkle, was 0.98 times the one using the satellite images and in the second case it was 0.96, it is possible to assume a factor ranging between 1.04 and 1.02. These factors should be further studied.

Chapter 5. Conclusions and future projects

In conclusion, the different methods available to locate the electricity demand play an important role in the planning of the electrical grid. The proposed method, of distributing the location of buildings, is promising in simplifying the planification of the expansion of areas where information is scarce.

The use of randomized points in the development of an electric grid is as good as the quality of the randomized inputs and the size of the area analyzed. This is, when the points are distributed uniformly across the whole area of study, without using any other input or constraint like the population density, the grid is distributed correctly as long as the area is small enough and the contour of the urbanized area is well determined. If the surrounding of the area is not fit to the towns and cities, the grid is then distributed with a higher cost and it does not adjust correctly to the real scenario.

On the other hand, when using the same randomized method in a large area as input to the electricity planning tool, the network obtained is overestimated. This is because large areas can have deeply populated and completely empty areas; when using a random generation of building coordinates the distribution, inside the area of study, is uniform and the network is required to cover all the coordinates increasing the kilometers of cable to be deployed.

Using more sophisticated methods, like the Sprinkle method developed, increases the precision of the network created. Just adding extra constraints, such as databases of the existence of settlements in a grid of 30 by 30 meters squares, to the fully random method has proved to come closer to the results obtained using the accurate locations. The Sprinkle method, though quite precise, tends to slightly underestimate the length and cost of the planned electrical network.

For future projects there could be further analysis of the different methods as well as improving the Sprinkle tool.

Analyzing more cases in more locations would help finding a conversion factor to transform the values of the Sprinkle result to a closer to the real value without underestimating the costs and length of the network.

To improve the time Sprinkle takes to run, which makes it hard to execute problems with more than 100000 buildings, the function that deletes the excess points could be optimized. To do so, other, more efficient, mathematical functions could be examined. An alternative, to optimizing the function that deletes extra points in Sprinkle, would be

to create a function that distributes the exact number of points, that represent buildings, inside the area defined by the polygon and the HRSL data.

Other ideas are functions to locate the center of highly populated areas and distribute more points in that area compared to its surrounding and the outskirts of the city; and functions that distribute the points, that represent buildings, inside the HRSL squares taking into account the population density and the latest censuses available.

Further analysis, of the impact of building location accuracy in electrification planning, can be made in the future using new geospatial information big tech companies are developing and launching as open source.

All these projects could bring electricity a little bit closer to those who still lack access to what has become something essential to most of us.

Bibliography

[1]   "IEA – International Energy Agency - Sankey Diagram."
        https://www.iea.org/sankey/ (accessed Jan. 24, 2023).

[2]   bp, "Statistical Review of World Energy 2022," 2022, Accessed: Jan. 20,
        2023.   [Online].   Available:
        https://www.bp.com/content/dam/bp/business-
        sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-
        stats-review-2022-full-report.pdf

[3]   World   Bank,   "GDP   (current   US$) | Data."
        https://data.worldbank.org/indicator/NY.GDP.MKTP.CD (accessed Jan. 21,
        2023).

[4]   "List of OECD Member countries - Ratification of the Convention on the
        OECD."   https://www.oecd.org/about/document/ratification-oecd-
        convention.htm (accessed Jan. 21, 2023).

[5]   International Energy Agency, "WEO-2017 Special Report: Energy Access
        Outlook," Paris, 2017. Accessed: Jan. 29, 2023. [Online]. Available:
        www.iea.org/t&c/

[6]   World Bank, "Access to electricity (% of population) | Data."
        https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS (accessed Jan. 14,
        2023).

[7]   "Universal   Energy   Access   Lab   -   MIT   &   IIT-Comillas."
        https://universalaccess.mit.edu/#/main (accessed May 29, 2023).

[8]   P. Ciller *et al.*, "Optimal Electrification Planning Incorporating On- And Off-
        Grid Technologies- And Reference Electrification Model (REM),"
        Proceedings of the IEEE, vol. 107, no. 9, 2019, doi:
        10.1109/JPROC.2019.2922543.

[9]   "GIS   (Geographic   Information   System)."
        https://education.nationalgeographic.org/resource/geographic-
        information-system-gis/ (accessed Jan. 04, 2023).

[10]   M. Awrangjeb, M. Ravanbakhsh, and C. S. Fraser, "Automatic detection of
        residential buildings using LIDAR data and multispectral imagery," *ISPRS*

Journal of Photogrammetry and Remote Sensing, vol. 65, no. 5, pp. 457–467, Sep. 2010, doi: 10.1016/J.ISPRSJPRS.2010.06.001.

[11]   Y. Aksoy and A. Aydin Alatan, "Utilization of false color images in shadow detection," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7584 LNCS, no. PART 2, pp. 472–481, 2012, doi: 10.1007/978-3-642-33868-7_47/COVER.

[12]   B. Sirmaçek and C. Ünsalan, "Building detection from aerial images using invariant color features and shadow information," *2008 23rd International Symposium on Computer and Information Sciences, Istanbul, Turkey*, 2008, doi: 10.1109/ISCIS.2008.4717854.

[13]   J. Inglada, "Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, no. 3, pp. 236–248, Aug. 2007, doi: 10.1016/J.ISPRSJPRS.2007.05.011.

[14]   N. Shorter and T. Kasparis, "Automatic Vegetation Identification and Building Detection from a Single Nadir Aerial Image," *Remote Sensing* , vol. 1, no. 4, pp. 731–757, Oct. 2009, doi: 10.3390/RS1040731.

[15]   W. Sirko *et al.*, "Continental-Scale Building Detection from High Resolution Satellite Imagery," Jul. 2021, Accessed: Jan. 27, 2023. [Online]. Available: http://arxiv.org/abs/2107.12283

[16]   "Open Buildings." https://sites.research.google/open-buildings/ (accessed Jan. 27, 2023).

[17]   O. Borkovska, J. Gill, L. Pistolesi, J. Squires, and G. Yetman, "Mapping the Distribution of Human Population: The High Resolution Settlement Layer (HRSL) Analysis Country Global Urban Footprint Global Human Settlement Layer High Resolution Settlement Layer", Accessed: Jan. 24, 2023. [Online]. Available: https://sedac.ciesin.columbia.edu/binaries/web/global/news/2018/borkovska_hrsl_ob_gy_v3.pdf

[18]   "Center for International Earth Science Information Network." https://ciesin.columbia.edu/data/hrsl/ (accessed Jan. 24, 2023).

[19]   "Facebook's   Data   For   Good   Program: HDX."
        https://dataforgood.facebook.com/dfg/docs/high-resolution-population-
        density-maps-demographic-estimates-documentation (accessed Jan. 27,
        2023).

[20]   G. Quinonez-Varela, A. Cruden, C. Graham, B. Punton, L. Blair, and J.
        Thomson, "A GIS/PSS planning tool for the initial grid connection
        assessment of renewable generation," *Renew Energy*, vol. 32, no. 5, pp.
        727–737, Apr. 2007, doi: 10.1016/j.renene.2006.03.011.

[21]   J. M. Sánchez-Lozano, C. Henggeler Antunes, M. S. García-Cascales, and L.
        C. Dias, "GIS-based photovoltaic solar farms site selection using ELECTRE-
        TRI: Evaluating the case for Torre Pacheco, Murcia, Southeast of Spain,"
        Renew   Energy,   vol.   66,   pp.   478–494,   Jan.   2014,   doi:
        10.1016/j.renene.2013.12.038.

[22]   A. García-Villoria, B. Domenech, L. Ferrer-Martí, M. Juanpera, and R.
        Pastor, "Ad-hoc heuristic for design of wind-photovoltaic electrification
        systems, including management constraints," *Energy*, vol. 212, p. 118755,
        Dec. 2020, doi: 10.1016/J.ENERGY.2020.118755.

[23]   J. Dezert and C. Musso, "An Efficient Method for Generating Points
        Uniformly Distributed in Hyperellipsoids," 2001.

[24]   K. Han, J. Franke, and Y. T. Feng, "Numerical Methods of Generating
        Random Points with Prescribed Distributional Properties on (Non)Rational
        Bezier Surfaces," *Mathl. Comput. Modelling*, vol. 23, no. 10, pp. 15–28,
        1996.

[25]   X. Rong Li, "Generation of random points uniformly distributed in
        hyperellipsoids," *Proceedings of the 1st IEEE Conference on Control
        Applications,   CCA   1992*,   pp.   654–658,   1992,   doi:
        10.1109/CCA.1992.269770.

[26]   C. Rocchini and P. Cignoni, "Generating Random Points in a Tetrahedron,"
        2001, Accessed: Aug. 24, 2023. [Online]. Available: http://vcg.iei.pi.cnr.it

[27]   "Generate   Uniform   Random   Points   within   a   Circle   –   Corner."
        https://meyavuz.wordpress.com/2018/11/15/generate-uniform-random-
        points-within-a-circle/ (accessed Aug. 24, 2023).

[28]   "SDGs   &   International   Development   Projects | UNDB."
https://devbusiness.un.org/sdgs-international-development-projects
(accessed Jul. 12, 2023).

ANNEXES

ANNEX I. SDG

ANNEX II. MATLAB Sprinkle code

ANNEX III. Webpages of interest

ANNEX IV. Acronyms

ANNEX I. SDG

According to the web of the United Nations:

> "The Sustainable Development Goals (SDGs) are the blueprint to achieve a better and more sustainable future for all. They address the global challenges we face, including those related to poverty, inequality, climate change, environmental degradation, peace and justice. The 17 Goals are all interconnected, and in order to leave no one behind, it is important that we achieve them all by 2030."[28]

For this annex an examination has been conducted on the potential impacts of the project developed with regard to the Sustainable Development Goals. As the system developed is a tool to generate random points following certain constraints and taking into account that it has been evaluated with a tool that generates electrical grids, the main focus of this annex will be on electricity access for developing countries.

Following this consideration, and knowing that the developed system could still be used for other purposes, the main impact this project has is on the SDG 7 (Affordable and Clean Energy), which intends to ensure access for all to energy in an economical, reliable and ecological way.

Figure 28: SDG 7

Inside this goal, there are some specific targets, some of which could be attained using what has been developed in this project. For instance, target 7.1 "ensure universal access to affordable, reliable and modern energy services" is one that is linked with this project. By making electrification easier by not needing so much geographical information about an area could speed up the initial steps of deciding to develop an electrical grid in a region.

Figure 29: SDG 4

Another Sustainable Development Goal that could be impacted from the development of this project is goal 4 "Quality education". The target 4.1 aims to ensure that children get a quality primary and secondary education. It highlights the need to improve the completion rate for those children. Improving access to electricity allows children to study when they arrive home from school and make their homework using electric light and not more costly gas/oil lamp.

ANNEX II. MATLAB Sprinkle code

II.I   General code for Sprinkle

This function is the general script of the developed code. In it all the other functions are called to execute the steps explained in section 3.3.1.

```
Function SprinkleFull(tiff_file,polygon_file,buildings,exit_name)
%% Script to get the sprinkled points from Collumbia HSRL map data
% and them into csv and kml/kmz files
%
% Inputs:
% *tiff_file: Collumbia HSRL tif file were the information about the
%  population is stored
% *polygon_file: kml file that is a polygon surounding the area of study
% *buildings: amount of buildings expected in the region
% *exit_name: name expected for the output files
%
% Example:
% [nga,refmat] =
% SprinkleFull('hrsl_nga_pop.tif','polygon.kml',898,'SprinkleFullPrueba1');
%
% Fidel Sanz Azuara, May 2023
%
%% Obtain data from the HSRL file

[A,R]=readgeoraster(tiff_file);
hrsl = A;
refmat = [0 -R.CellExtentInLongitude; R.CellExtentInLatitude 0;
R.LongitudeLimits(1) R.LatitudeLimits(2)];

% Alternatively, but not recomended:
% [hrsl,refmat,bbox_nga] = geotiffread(tiff_file);

%% Obtain squares where there is population
% This is done inside the area marked by the polygon
[squares_LoLa,num_squares]=get_active_squares_hrslf(hrsl,refmat,polygon_file);

%% Distribute the random points inside the squares
[rdots_new] = sprinkle_randomf (squares_LoLa, buildings);

%% Delete the excess of points in the sprinkle using Delaunay
FinalSprinklePoints = DeleteExtraPointsDf (rdots_new, buildings);

%% Export the csv and kml file of the sprinkled points
ExportCsvKmLf(FinalSprinklePoints,exit_name);

End
```

II.II Get active squares from HRSL

This function obtains the coordinates of the squares inside the drawn polygon which are inhabited, using the data in HRSL, as explained in section 3.3.1.2.

```
function
[squares_LoLa,num_squares]=get_active_squares_hrslf(hrsl,refmat,polygon_file)
%% Getting the square limits from the HRSL data inside a defined polygon
%
% This function uses HRSL data and finds the squares where there are
% settlements; then selects the squares inside the polygon and these are
% returned in an array containing the coordinates of said squares' vertices
%
% Inputs:
% *hrsl: array with information of presence (1) or absence (0) of a
%        settlement in the square
% *refmat: matrix with information of the coordinates of each square
% *polygon_file: kml file with the information of the coordinates of the
%                boundary of the region being analyzed
%
% Outputs:
% *squares_LoLa: coordinates of the squares that inside the polygon
% *num_squares: number of squares inside the polygon
%
% Example:
%
[squares_LoLa,num_squares,rdots_new]=get_active_squares_hrsl_v02(nga,refmat,'p
olygon.kml');
%
% Original: Rafael Palacios, May 2018
% Modified: Fidel Sanz Azuara, May 2023
%
%% Find coordinates of places with Settlements
[rows,cols] = find(hrsl>0);

%% Load KML polygon
s = kml2struct(polygon_file);
poly_lat = s.Lat;
poly_lon = s.Lon;

%% Count num active dots
LL = [rows,cols,ones(size(rows))] * refmat;
IN = inpolygon(LL(:,1),LL(:,2),poly_lon,poly_lat);
num_squares = sum(IN);

%% Create sub-polygons
Lat = [];
Lon = [];
ind = find(IN==1);
pp = [];
for i=1:length(ind)
    pp=[pp;...
        rows(ind(i))-0.5 cols(ind(i))-0.5 1; ...
        rows(ind(i))-0.5 cols(ind(i))+0.5 1; ...
        rows(ind(i))+0.5 cols(ind(i))+0.5 1; ...
        rows(ind(i))+0.5 cols(ind(i))-0.5 1; ...
        rows(ind(i))-0.5 cols(ind(i))-0.5 1; ...
        NaN NaN 1];
end
squares_LoLa= pp * refmat;

%% GEplot polylines
GEplot(['squares_'  polygon_file],squares_LoLa(:,2),squares_LoLa(:,1),'-b');
```

```matlab
end
```

II.III Sprinkle random

This function distributes points inside the area defined by the squares obtained as explained in section 3.3.1.3.

```matlab
function [rdots_new]=sprinkle_randomf(squares_LoLa,buildings)
%% Sprinkle random points inside predefined squares
%
% This function distributes the number of points provided (buildings)
% evenly among all squares.
%
% Inputs:
% *squares_LoLa: squares where the points are distributed in
% *buildings: number of points to distribute inside all of the squares
%
% Outputs:
% *rdots_new: coordinates of the points randomly distributed
%
% Examples:
% [rdots_new] = sprinkle_randomf(squares_LoLa,buildings);
%
% Fidel Sanz Azuara, May 2023
%
%% Take the number of areas with settlements from the squares_t table
areas = size(squares_LoLa,1)/6;
num = size(squares_LoLa,1);
% The ammount of houses/buildings in the area:
houses = buildings;

%% The ratio houses/area:
houseperarea = houses / areas;
hpa = ceil(houseperarea);


%% Generate the random points representing houses
i = 1;
j = 0;
puntos = 0;
rdots = NaN(houses, 2);
while (i<num)
        latmin = squares_LoLa(i+2,2);
        latmax = squares_LoLa(i,2);
        lonmin = squares_LoLa(i,1);
        lonmax = squares_LoLa(i+1,1);

        housesinarea = randi([hpa-2,hpa]);

        while (j<hpa)
            latrand = latmin + (latmax-latmin).*rand(1,1);
            lonrand = lonmin + (lonmax-lonmin).*rand(1,1);

            puntos=puntos+1;
            rdots(puntos,1:2)=[latrand, lonrand];
            j=j+1;
        end
        j = 0;
        i = i + 6;
end

rdots_new = rdots;
end
```

II.IV Delete extra points

This function deletes the extra generated points, as explained in section 3.3.1.4.

```matlab
function [FinalSprinklePoints]=DeleteExtraPointsDf(rdots_new,buildings)
%% Delete the extra points created
%
% This function deletes the excess points from an array. The points are
% deleted to avoid them being too close.
%
% Inputs:
% *rdots_new: array containig coordinates of points created previously
% *buildings: number of points to distribute inside all of the squares
% Outputs:
% *FinalSprinklePoints:
%
% Example:
% FinalSprinklePoints = DeleteExtraPointsDf (rdots_new, buildings);
%
% Fidel Sanz Azuara, May 2023
%
%% Detect points which are closer to each other
X = rdots_new(:,1);
Y = rdots_new(:,2);
points = size (rdots_new,1);
pdist=[];
for i = 1:points
    rows2delete = i;
    % select the rows of the array to keep
    rows2keep = setdiff(1:size(rdots_new, 1), rows2delete);
    rdots_new_2= double(rdots_new(rows2keep, :));
    TN = delaunayn(rdots_new_2); % make Delaunayn triangulation
    Querypoints = double([X(i),Y(i)]);
    % calculate the closest point to the querypoint a the distance
    [closestp,dist] = dsearchn(rdots_new_2,TN,Querypoints);
    % add the information of the point and the distance to pdist array
    if closestp == i
        pdist=[pdist;...
            closestp+1 dist];
    else if closestp > i
            pdist=[pdist;...
            closestp+1 dist];
        else
            pdist=[pdist;...
            closestp dist];
        end
    end
end

%% Sort the points from closest distance to bigest
pdist_orde = sortrows(pdist,2);

%% Delete the closest found points
houses = buildings;
srdots = size(rdots_new,1);
rows2delete = [];
j = 1;

while(srdots>houses)
    i_check = 1;
    check = 1;
    while(i_check<size(rows2delete,2))
        if pdist_orde(j,1) == rows2delete(1, i_check)
            i_check = size(rows2delete,2);
            check = 0;
```

```matlab
        end
        i_check = i_check +1;
    end
    if check == 1
        rows2delete = [rows2delete pdist_orde(j,1)];
        srdots = srdots - 1;
    end
    j = j+2;
end

rows2keep = setdiff(1:size(rdots_new, 1), rows2delete);
rdots_new_limit = rdots_new(rows2keep, :);

%% Export points
FinalSprinklePoints = rdots_new_limit;

end
```

II.V Export to CSV and KML

This function exports the obtained points to a CSV and KML files, as explained in section 3.3.1.5.

```matlab
function ExportCsvKmLf(FinalSprinklePoints,exit_name)
%% Export points to kml and csv files
% This function exports the data created throught sprinkle to a kml and csv
% file. The kml can be open in software (like google earth) to view the
% location of the points. The csv file is created in a format that can be
% used, with minimum changes, in REM and in other softwares.
%
% Inpputs:
% *FinalSprinklePoints: is the array that includes the location of each
%                       point created.
% *exit_name: is the desired name for the files created
%
% Outputs:
% There is two outputs that are created outside MATLAB, these are the KML
% and KML files containing the points information.
%
% Example:
% ExportCsvKmlf(Points, 'Example');
%
% Fidel Sanz Azuara, May 2023
% Uses GEplot created by Rafael Palacios
%
%% Generate KML file with the points generated
GEplot(exit_name,FinalSprinklePoints(:,1),FinalSprinklePoints(:,2),'.b','MarkerSize',20);

%% Create new table for the CSV file
num = size(FinalSprinklePoints,1);
sz = [num 5];
varTypes = {'double','double','double','double','double'};
varNames = {'ID','Lon','Lat','customer_type','electrification_status'};
Tfs_mod = table('Size',sz,'VariableTypes',varTypes,'VariableNames',varNames);

%% Add the points information to the table
ii=0;
for ii = 1 : num
    Tfs_mod.ID(ii) = 100+ii;
    Tfs_mod.Lat(ii) = FinalSprinklePoints(ii,1);
    Tfs_mod.Lon(ii) = FinalSprinklePoints(ii,2);
    Tfs_mod.customer_type(ii) = 1;
    Tfs_mod.electrification_status(ii) = 0;
```

```
end

%% Create the CSV file from the table
writetable(Tfs_mod,[exit_name '.csv'],'Delimiter',',','QuoteStrings',true);

end
```

II.VI Fully random sprinkle

This function distributes points randomly inside a defined polygon, as explained in section 3.3.3.

```
function [rdots_new]=sprinkle_random_totalf(polygon_file,buildings)
%% Sprinkle random points inside a predefined polygon
%
% This function generates random points evenly distributed inside a
% polygon.
%
% Inputs:
% *polygon: closed path that delimits the area where points are created
% *buildings: number of points to distribute inside all of the squares
%
% Outputs:
% *rdots_new: coordinates of the points randomly distributed
%
% Examples:
% [rtdots_new] = sprinkle_random_totalf('polygon.kml',buildings);
%
% Fidel Sanz Azuara, May 2023
%
%% Take the number of areas with settlements from the squares_t table
% The ammount of houses/buildings in the area:
houses = buildings;

%% Load KML polygon
s = kml2struct(polygon_file);
poly_lat=s.Lat;
poly_lon=s.Lon;

poly_lon_max = max(poly_lon);
poly_lat_max = max(poly_lat);
poly_lon_min = min(poly_lon);
poly_lat_min = min(poly_lat);

%% Generate the random points representing houses
i = 0;
j = 0;
puntos = 0;
rdots = NaN(houses, 2);

latmin = poly_lat_min;
latmax = poly_lat_max;
lonmin = poly_lon_min;
lonmax = poly_lon_max;

while (puntos < houses)
    latrand = latmin + (latmax-latmin).*rand(1,1);
    lonrand = lonmin + (lonmax-lonmin).*rand(1,1);

    IN=inpolygon(lonrand,latrand,poly_lon,poly_lat);
    if (IN==1)
        puntos=puntos+1;
        rdots(puntos,1:2)=[latrand, lonrand];
```

```
    end
    if mod(puntos,100)== 0
    fprintf(' There are %d points in the kml file.\n',puntos);
    end
end
fprintf(' There are %d points in the kml file.\n',puntos);
rdots_new = rdots;
end
```

ANNEX III. Webpages of interest

The following webpages contain information of interest containing geographical information, data and tools useful for the project and future developments.

Maps and satellite images:

- Maps of the administrative areas of all countries: https://gadm.org/index.html
- Maps of the global power system: https://gridfinder.rdrn.me/
- Countries shapes: https://open.africa/dataset/africa-shapefiles
- Nigeria shape file: https://africaopendata.org/dataset/shape-file-of-nigeria
- OpenSteetMaps: https://www.openstreetmap.org/#map=5/-1.033/11.997

HRSL

- HRSL Information: https://ciesin.columbia.edu/data/hrsl/
- Interactive map with the HRSL data:
  https://columbia.maps.arcgis.com/apps/View/index.html?appid=ce441db6aa5
  4494cbc6c6cee11b95917
- HRSL data in Africa:
  https://data.humdata.org/dataset/highresolutionpopulationdensitymaps

Buildings databases:

- Open Buildings database: https://sites.research.google/open-buildings/#faq
- Buildings database from Bing Maps:
  https://github.com/microsoft/GlobalMLBuildingFootprints

Other information and tools

- Nigeria spatial distribution of Population:
  https://hub.worldpop.org/geodata/summary?id=49705
- Datasets of various disciplines geographically referenced:
  https://data.humdata.org/
- QGIS, tool to work with and visualize geographical information:
  https://www.qgis.org/en/site/
- World population denstity: https://ourworldindata.org/grapher/population-density

ANNEX IV. Acronyms

In order of appearance :

- GDP: Gross Domestic Product
- OECD: Organisation for Economic Cooperation and Development
- IIT-Comillas: Instituto de Investigación Tecnológica
- IRT-Comillas: Institute for Research in Technology
- REM: Reference Electrification Model
- GIS: Geographic Information Systems
- LIDAR: Light Detection and Ranging
- SVM: Support Vector Machines
- HRSL: High Resolution Settlement Layer
- CIESIN: Center for International Earth Science Information Network
- CSV: Comma Separated Values
- KML: Keyhole Markup Language
- TIF: Tagged Image File format
- HTML: HyperText Markup Language

In alphabetical order:

- CIESIN: Center for International Earth Science Information Network
- CSV: Comma Separated Values
- GDP: Gross Domestic Product
- GIS: Geographic Information Systems
- HRSL: High Resolution Settlement Layer
- HTML: HyperText Markup Language
- IIT-Comillas: Instituto de Investigación Tecnológica
- IRT-Comillas: Institute for Research in Technology
- KML: Keyhole Markup Language
- LIDAR: Light Detection and Ranging
- OECD: Organisation for Economic Cooperation and Development
- REM: Reference Electrification Model
- SVM: Support Vector Machines
- TIF: Tagged Image File format