Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Civil and Environmental Engineering

Mons Erling Mathiesen

Nighttime Lights Based Pathfinding Model for Electrical Line Layout Prediction

Master's thesis in Engineering and ICT - Geomatics
Supervisor: Hongchao Fan
Co-supervisor: Fredrik Moger
August 2021

NTNU
Kunnskap for en bedre verden

Mons Erling Mathiesen

Nighttime Lights Based Pathfinding Model for Electrical Line Layout Prediction

Master's thesis in Engineering and ICT - Geomatics
Supervisor: Hongchao Fan
Co-supervisor: Fredrik Moger
August 2021

Norwegian University of Science and Technology
Faculty of Engineering
Department of Civil and Environmental Engineering

NTNU
Norwegian University of
Science and Technology

Abstract

The world is failing to stay on track to reach Sustainable Development Goal 7, *"ensure access to affordable, reliable, sustainable and modern energy for all"*. Five hundred eighty-five million people in Sub-Saharan Africa are expected to live without access to a reliable energy source by 2030. It is not a question of a greener, more sustainable energy source but the immediate need for energy. This study aimed to simplify the site selection process for new power-producing projects by documenting the existing electrical infrastructure, as distance to the grid is a critical cost parameter. We studied the possibility of inferring power lines using nighttime lights. Small settlements, buildings and even street lamps emit light radiance detectable by satellites. Nighttime lights functioned as nodes in a graph and were connected by Dijkstra's algorithm. The resulting path was the assumed power line layout in a country.

We studied previous research concerning power line inference and planning. The methodology aimed to reproduce and better the most promising research. Further, alternative approaches, such as street network detection using satellite imagery, were investigated to determine how methods could be transferred to power line inference.

The implemented solution makes use of water bodies, populated areas and convolutional filters to preprocess nighttime lights. Road networks, elevation changes, and protected areas determined the path's cost. We gathered open-source power line data from 34 European countries to determine the quality of the predictions. The lack of documentation in Sub-Saharan Africa restricted validation opportunities. The layers and filters aimed to remove natural light sources and noise while cost layers facilitated real-world grid design.

Finally, we scrutinized predicted paths to identify possible method improvements. The detailed maps illustrated complications with the methodology and external factors. The predicted paths of power lines achieved an Intersection-over-Union score of 0.29. The results were below industry standards of 0.5, deeming the predicted paths unsuitable as infrastructure documentation in projects. Errors in methodology partially explained the low evaluation metrics, but a significant share was attributed to low-quality validation data.

Sammendrag

Verden holder ikke følge med FNs bærekraftig utviklingsmål 7, *"ensure access to affordable, reliable, sustainable and modern energy for all"*. Fem hundre åttifem millioner mennesker i Afrika sør for Sahara forventes å leve uten tilgang til en pålitelig energikilde innen 2030. Det er ikke snakk om en grønnere, mer bærekraftig energikilde, men det primære behovet for energi. Denne studien forsøkte å forenkle stedvalgprosessen for nye energiprosjekter ved å dokumentere den eksisterende elektriske infrastrukturen. I denne oppgaven undersøkte vi muligheten for å utlede kraftledninger ved bruk av nattlys. Små bosetninger, bygninger og til og med gatelamper avgir lysstråling oppdagbart av satellitter. Nattlys fungerte som noder i en graf og ble koblet sammen med Dijkstras algoritme. De resulterende koblingene var det antatte strømlinjenettet i et land.

Vi studerte tidligere forskning om predikering av kraftledninger. Den primære kilden til relatert arbeid brukte Dijkstras algoritme og nattlys og forslagene deres for fremtidig arbeid ble implementert. Samtidig gjennomførte vi et litteraturstudie som studerte overlappende problemer, for eksempel deteksjon av gatenettverk ved hjelp av satellittbilder, for å avgjøre hvordan metoder kan overføres til det aktuelle domenet.

Den implementerte løsningen bruker vannforekomster, befolkede områder og konvolusjonsfilter til å forbehandle nattlys. Veinettverk, høydeendringer og naturreservater bestemte kostnaden ved veivalg. Vi samlet data fra ideele organisasjoner i 34 europeiske land for å bestemme kvaliteten på de utledede strømlinjene. Mangelen på dokumentasjon i Afrika sør for Sahara begrenset valideringsmuligheter. Lagene og filtrene hadde som mål å fjerne naturlige lyskilder og støy, mens kostnadslagene muliggjorde strømnettdesign som i virkeligheten.

Til slutt gransket vi de predikerte strømlinjene for å identifisere mulige metodeforbedringer. De detaljerte kartene illustrerte komplikasjoner med metodikken og eksterne faktorer. De predikerte strømlinjene oppnådde en Intersection-over-Union poengsum på 0.29. Resultatene var under industristandarder og egnet seg ikke som dokumentasjon av strømlinjer. Feil i metodikk forklarte delvis de lave poengsummene, men en betydelig andel ble tilskrevet valideringsdata av lav kvalitet.

Preface

This mater thesis marks the final delivery of the study program Engineering and ICT, with a specialization in Geomatics. The program is a part of the Department of Civil and Environmental Engineering at the Norwegian University of Science and Technology. I want to thank Enernite for introducing the thesis problem and my supervisor Hongchao Fan for valuable input.

Mons Erling Mathiesen, Oslo 27.08.2021

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The introduction will present the motivation and background of the study. In addition, the chapter provides an explicit definition of the master thesis' research goal and methods. Finally, we provide a brief outline for the thesis.

1.1 Background & Motivation

Well-documented power line layout data is challenging to obtain in underdeveloped countries. Data is fragmented across different companies and governmental bodies, requiring extensive efforts to get a complete overview. Reliable infrastructure data is a vital factor when planning new projects. In this master thesis, we aim to predict the layout of power lines in a country based on nighttime lights. The predictive model uses nighttime lights as a proxy for electricity consumption and links all light hotspots using the shortest path algorithm. The shortest path algorithm is guided by various data layers aiming to encourage actual electrical grid design. The purpose is to create a model to generate a reliable dataset of power lines where documentation is absent.

Power producers spend significant time and resources to assess a site's suitability. In the early stages of site selection, there is a coarse screening process discarding unsuitable sites. Enernite is a company providing data, data analysis and planning support for power producers in Sub-Saharan Africa (SSA). This master thesis conducts a study to assist Enernite in one of its most common challenges, providing accessible and reliable electricity infrastructure data. Power producers

have expressed difficulties concerning data collection when considering sites in different regions of a country or even multiple countries. Data source fragmentation complicates and delays data collection, forcing producers to spend time and resources on a potentially unsuitable site. An important parameter considered in site selection is the distance to existing electricity infrastructure, indicating investments required to connect. A complete dataset of the transmission and distribution grids in Sub-Saharan Africa facilitates distance-to-grid calculations, increasing the efficiency of the site selection phase.

The emphasis on Sub-Saharan Africa results from the lack of efforts to stay on track to reach Sustainable Development Goal 7. The United Nations established the Sustainable Developments Goals in 2015, where Goal 7 captures the electricity access challenges of SSA: *"ensure access to affordable, reliable, sustainable and modern energy for all"*. In 2019, the Energy Progress Report estimated that 650 million people would remain without access to a reliable energy source in 2030. The population in SSA make up 90% of these 650 million. Enernite has already experience increased demand for projects in SSA, and that the region will be the recipient of increased investments and focus over the next decade.

Well-documented infrastructure is a luxury few countries provide, especially in Sub-Saharan Africa. Data providers exist, but platforms are tailored to individual countries. This master thesis is based solely on open-source data. OpenStreetMap and other geospatial data sources are comparable in completeness and accuracy to commercial providers in certain countries, but infrastructure data is lacking in Sub-Saharan Africa. Algorithms and models created in this master thesis were applied and validated in Europe, where electrical grids are documented. OpenInfraMap stated in an email (R. Garret, personal communication, 25. April 2021) that transmission networks ($> 200kV$) are complete in Europe. However, lower voltages are under-mapped everywhere, except the UK, France and the Czech Republic, which provide some open power network data.

1.2 Research Goals

The work conducted in this thesis aims to study two research goals.

RG1: Improve the state of the art methods for distribution grid prediction by introducing new data layers for filtering and weighting of the shortest path algorithm.

The current state of the art methods using nighttime lights for estimating the layout of power lines only applies OpenStreetMap roads networks to assist Dijkstra's

shortest path. This thesis aims to study the impact of a population filter and an urban area filter to better model human activity. In addition, data layers with protected areas and slopes are introduced to implement constraints experienced in real-world power line layout design.

RG2: Assess the reliability of predicted distribution grid layouts.

The validity and accuracy of predictions must be determined to have any value for solar power producers. A challenge in remote sensing is the validity of a model in various geographical areas. This study plans to validate model outputs in a large geographical area, most European countries, to indicate the model's general validity. All data is available globally, ensuring model applicability in regions other than the studied area.

1.3 Research Methods and Outline

In order to achieve the research goals, a literature review of relevant theory, datasets, and state of the art methods was conducted. The acquired knowledge was then applied to improve the state of the art methods and validate new findings. The thesis starts by studying theory and related work before describing the methodology and validation process. Theory and related work was studied to be able to improve on the state of the art methods, thus contributing to solving **RG1**. Further, the model is implemented and applied in many European countries before validation against open-source data. This approach aims to solve **RG2**.

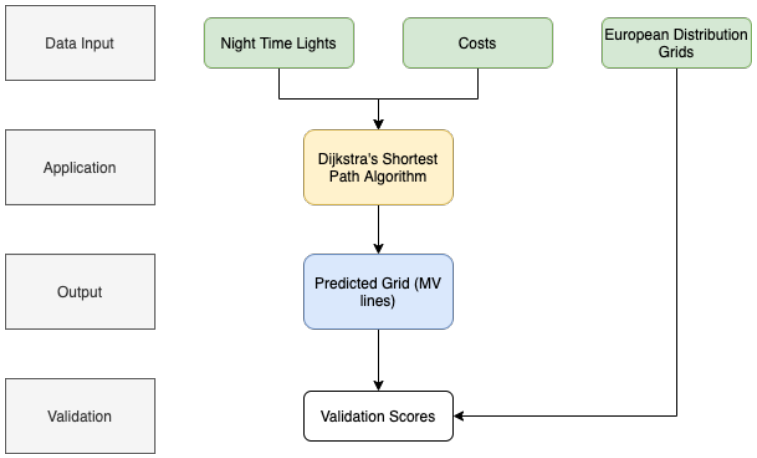


Figure 1.1: Simplified overview. Input data in green, the model in yellow, model outputs in blue, and the evaluation metrics in white.

Chapter 2 makes up the research phase of the report. The chapters aim to give the reader insights into relevant theory, the current state of the art methods and possible improvements. Chapter 3 continues by describing aspects of the methodology and the reasoning behind implemented changes. Chapter 4 covers the practical implementation and the experimental results. Finally, chapters 5 and 6 quantitatively and qualitatively assess the method and results.

Theoretical Background (Chapter 2): This chapter provides the necessary theoretical background to understand the problem and the methods used. In addition, the primary data source, the Suomi National Polar-orbiting Partnership (S-NPP) satellite, is covered. The chapter also presents insight into the most recognized research based on nighttime lights. The theoretical background is explained through a review of related work.

Methodology (Chapter 3): After providing theoretical background and relevant research in chapter 2, chapter 3 presents the implemented methods. This chapter gives the reader reasoning behind applied datasets and methods. Different aspects of the study, costs and filters, are presented and linked to the real-world challenges of grid design. Finally, the chapter presents the validation process and evaluation metrics.

Experiments and Results (Chapter 4): Experiments presents all steps necessary to reproduce the experiments made in the thesis. The chapter provides a detailed description of each dataset, covering data source, resolution and neces-

sary preprocessing before usage. This chapter first presents a baseline model. The baseline model is used as a benchmark for improvements. The chapter continues by presenting the impact of proposed filters and new cost layers.

Discussion (Chapter 5): Prediction of distribution grids using night lights is a method with high uncertainty and many limitations. This chapter discusses the practical applicability of the obtained results.

Conclusion and Future Work (Chapter 6): The last chapter summarizes the main findings and presents our perspective on the results. Finally, we present suggestions for future work.

Chapter 2

Theoretical Background

The following chapter provides an introduction to key concepts, methods, and relevant literature. We first cover prominent research based on nighttime lights and argue for their suitability as a consumption proxy. The chapter gives a brief overview of the satellite program, its surveying capabilities and areas of application. Further, the differentiation of grid categories and studies optimizing power line design are reviewed. Although the amount of research using nighttime lights for distribution grid inference is limited, nightlights and energy transmission line design is widely studied. This chapter is divided into three fundamental sections; Nighttime lights (2.1), pathfinding (2.2), and electrical grids (2.3).

2.1 Night Time Lights

Street lamps, flashing billboards, oil flares, forest fires, northern lights are all sources of light pollution captured by the night light sensor onboard satellites. Despite natural phenomena, night lights have been valuable when used as a proxy of human activity. They can provide insights at finer scales than what is found in national statistics. Most notably are insights into impoverished regions of the world, where statistics are absent or of low-quality [1], [2]. The human activity proxy has been applied to monitor urban expansion [3], [4] and to estimate GDP [5]. Li [6] studied nighttime light dynamics under the Iraqi civil war, and Zhao [7] applied daily nighttime light data to study the impact of natural disasters.

2.1.1 Electricity Consumption Proxy

The failure to stay on track to reach SGD 7 has motivated remote sensing of night lights research. Multiple studies have used nightlights to estimate access rates in rural and developing regions of the world. Doll [8] presented the first satellite-derived estimates of rural electrification rates in developing countries. The study suggested that night lights are sufficient for detecting small human settlements. In addition, their findings include alarmingly low electrification rates in Sub-Saharan Africa, supporting this thesis' purpose. Dugoua and his team [9] applied night lights to study rural electrification rates in India. The paper found that rural electrification rates derived by nighttime luminosity were surprisingly accurate. Night lights were also experimented with as a proxy for poverty and financial inclusion but performed significantly worse than access rates.

Falchetta [10] performed a similar study using higher-resolution data and derives electricity rates in Sub-Saharan Africa from a combination of nighttime lights, land cover, and population data. Their results are broadly consistent at both province- and national-level statistics. All studies come to similar conclusions, suggesting that night lights are a well-functioning proxy for electricity access rates. The alternatives to using night lights as targets would be to use another proxy for electricity consumption. The team behind OpenGridMap [11] suggested targets derived by buildings. A critical difference between using night lights and buildings as targets is the availability of data sets. Nighttime lights are available globally, but open-source building footprints vary greatly in coverage and quality depending on the country.

2.1.2 Night Time Light Data Source

There exists a wide range of satellite sensors capturing nighttime lights. Table 2.1 provides an overview of some sensors capturing nighttime lights. The table is a simplified version of what was provided by Levin [12]. In the process of satellite selection, we evaluated parameters such as spatial resolution, availability, amount of preprocessing required, and research based on the sensor. The parameter excluding most satellite sensors was the availability of data products. Sensors such as EROS-B can provide night lights at resolutions of 0.7 by 0.7 meters [13], but data is captured on-demand.

Sensor	**Spatial Res. (m)**	**Temporal Res.**	**Products**
DMSP/OLS	3000	Daily (1992-2013)	Free, calibrated
VIIRS/DNB	500	Daily (2012 - present)	Free, science grade monthly composites
LuoJia1-01	130	15 day revisit time	Free, calibrated
Landsat 8	15-30	Sporadic (2013)	Free, only very bright objects detected
EROS-B	0.7	On demand	Commercial

Table 2.1: Overview of select night light sensors

Landsat 8 captures night lights through an optical imagery sensor. The sensor provides data at better spatial resolutions than VIIRS/DNB but captures night lights infrequently. Sporadic capturing complicates the use of data in research. Thus most researchers prefer VIIRS/DNB over Landsat 8 for night lights research but are optimistic for Landsat 10 [14]. LuoJia1-01 (LJ1-01) is a small satellite explicitly launched for acquiring nighttime lights. LJ1-01 regularly provides data at 130 m resolution through an online portal [15]. VIIRS/DNB provided daily images, but this study did not require frequencies above monthly, meaning that the biweekly frequency of LJ1-01 was equally suitable.

The DMSP/OLS was considered the primary NTL-data source but was discontinued in 2013 and, therefore, not suitable for this project. The spatial resolution is also inferior compared to the alternatives. The sensor was replaced by VIIRS/DNB, which has been the primary data source in scientific research. The preprocessed night lights published by the Earth Observation Group (EOG) are easily accessed through an online portal. The processing removed natural phenomena and other noise, which were unwanted for this study. All parameters considered, both LuoJia1-01 and VIIRS/DNB were suitable sensors. The preprocessing removes unwanted natural phenomena such as sunlight, moonlight, data affected by lightning, and stray lights. The necessary steps to go from raw sensor data to a science grade radiance raster are thoroughly described in Elvidge [16]. As far as the author knows, the free, high resolution, science grade rasters are one of a kind.

2.1.3 Soumi NPP Mission

The Suomi National Polar-orbiting Partnership (Suomi NPP) is, as of May 2021, a fully operational earth-orbiting satellite. NASA launched the satellite in October 2011 with an initial planned mission duration of 5 years. However, the Suomi NPP continues to monitor the Earth's health today. Five instrument suites comprise the satellite's surveying capabilities.

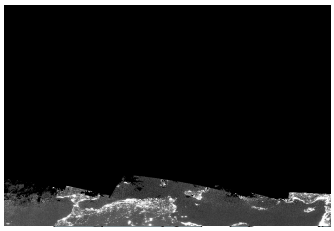
The Ozone Mapping and Profiler Suite's (OMPS) monitors ozone levels. Data captured by OMPS plays a vital role in ozone-related research, such as the monitoring of the ozone hole in the Antarctic [17]. Clouds and the Earth's Radiant Energy System (CERES) keeps track of the Earth's emitted and reflected energy. Research of long term changes in the climate applies energy and reflectance data provided by CERES. Other instruments aboard the Suomi NPP satellite include the Cross-track Infrared Sounder (CrIS) and Advanced Technology Microwave Sounder(ATMS), which acquire temperature, pressure and moisture data used for weather predictions. The last and largest instrument suite flying on the Suomi NPP is the Visible Infrared Imaging Radiometer Suite (VIIRS). VIIRS captures data on the visible and infrared spectra. The radiometric data's most common use cases include monitoring phytoplankton, ocean colour, fires, and vegetation development.

2.1.4 VIIRS/DNB

The launch of Suomi NPP and VIIRS Day/Night Band (DNB) has significantly improved NTL data quality. Before the launch of Suomi NPP, the Defense Meteorological Satellite Program Operational Linescan System (DMSP/OLS) data was the primary data resource for night light related research [16], [18]. Data collected by DMSP/OLS is publicly available from 1992 to 2013 at a spatial resolution of 3000m. The Day/Night band in VIIRS captures natural and artificial lighting from the Earth's surface and atmosphere at a spatial resolution of 15 arc seconds (450m)[19]. Interestingly, the primary purpose of DNB was to observe clouds lit by moonlight, which in turn was to be used for weather predictions. However, the Day/Night Band has increased the opportunities for night lights research by significantly improving night lights' data quality. In addition to improving the spatial resolution, the DNB exceeds the quality of OLS by avoiding saturation in urban areas. Avoidance of light saturation is possible due to a broader dynamic range. Unsaturated data in urban areas is crucial when monitoring changes in urban areas but has a limited impact on this study.

The DNB is capable of detecting radiance values at a minimum of $3 * 10^{-5} \frac{W}{m^2 sr}$, which in practical terms is comparable to a single, isolated street lamp [20]. Despite improvements, natural challenges of night light-capturing persist. Night Light instruments are prone to the midnight sun and short nights as it narrows the satellite's overpass window. The impact of short nights is visualized in Figure 2.1. Monthly composites from January 2020 (left) and May 2020(right) provided by the Earth Observation Group (EOG) were visualized in QGIS. Both composites are of the same area, but short nights resulted in the absence of data in the right image. The top section of Figure 2.1a also illustrates the impact of aurora borealis.[1] The black area in 2.1b does not imply that no night lights were observed, but the quality of the data was too low to be included in the monthly composite.

(a) Data coverage January, Europe



(b) Data coverage May, Europe

Figure 2.1: Side by side visualization of nightlights over Europe impacted by short nights

2.2 Pathfinding

Pathfinding is the problem of finding the shortest path between a start point and a destination. Several variants of the problem exist; the variant depends on the nature of the problem. The most common cases are finding the shortest path between:

- One start node and a single goal node
- One start node and multiple goal nodes
- All nodes

[1] Northern Lights

The number of sources and destinations represent the main differences invariants, but edges and edge weights can introduce other criteria to a problem. Typical applications of pathfinding are satellite navigation [21] and network package routing [22].

This thesis's distribution grid layout problem is similar to finding the shortest paths between all nodes. In addition, a combination of datasets is used to calculate the cost of traversal, thus incorporating multiple criteria in the found path. A random target point, a well-lit pixel, is chosen as the source. The process expands until all nodes have been visited and all target points have been discovered.

Dijkstra Dijkstra's algorithm (DA) is a shortest path algorithm developed by Edsger W. Dijkstra in 1959. The note written by Dijkstra [23] introduces the algorithm as a solution for finding the shortest path between a start and goal node, but it is today commonly applied to find the shortest path between a start node and all other nodes. When predicting the layout of a nation's distribution grid, a randomly selected electrified settlement is set as the source node and the rest as destinations. The cost matrix value gives the expense of travelling from a cell to a neighbour in the specified position. The cost of a cell is calculated based on distance, road networks, protected areas and slope. Cost calculations are further discussed in Section 3.2.

A fundamental feature of Dijkstra's is the usage of a priority queue. The algorithm sorts a queue of unvisited nodes by costs in increasing order, continuously expanding the closest or cheapest node. An edge's weight gives the cost of traversing a node. The original algorithm's logic applies to a graph with a set of vertices and edges but also holds for matrix-based problems.

$\sqrt{2}$	1	$\sqrt{2}$
1	0	1
$\sqrt{2}$	1	$\sqrt{2}$

Table 2.2: Cost calculation kernel

Table 2.2 shows the kernel for calculating traversal costs from the current location. Traversal costs are the product of the kernel value and the cost raster value at the same index. The kernel accounts for increased travel distance along the diagonal by a factor of $\sqrt{2}$. The center cell in Table 2.2 is our current position. Table 2.3 shows an example cost raster and calculated traversal costs from the starting position. Cell S indicates the starting cell. The first iteration of Dijkstra's shortest path will traverse either directly west, east, or south as they offer the

cheapest travel.

1	1	1	S	1
1	2	2	1	1
1	2	2	5	1
1	2	2	1	1
1	T	1	1	1

-	-	-
1*1	0	1*1
$\sqrt{2} * 2$	1*1	$\sqrt{2} * 1$

Table 2.3: Example cost raster and kernel applied to the starting location

Dijkstra's expands greedily, ensuring the shortest path to the current location. In other words, the greedy property implies that once the current location is a target point, the shortest path is also discovered. When a target point, cell T in Table 2.4, is explored, the costs along the path connecting S and T are zeroed. The link between target points is classified as a power line, and the model can reuse the link for further exploration without additional cost. This modified version of DA creates a minimum spanning tree between all target points.

1	1	1	S	1
1	2	2	1	0
1	2	2	5	0
1	2	2	0	1
1	T	0	1	1

Table 2.4: Shortest path

Facebook Engineering introduced the modified version of DA described above in 2019. The engineers published the method encouraging further research using DA and nighttime lights for distribution grid prediction [24].

2.3 Electrical Grids

Electrical grids can be divided into two main categories; transmission grid and distribution grid. The transmission grid operates at high voltage levels because power transfer is more efficient at higher voltages. While the primary purpose of the transmission network is bulk transportation of power, the distribution grid provides access and usable energy to smaller entities such as solar farms and rural areas. The voltage level distinguishing transmission from distribution varies from

country to country. For practical purposes, this report makes a clear separation. The distribution grid concerns all power lines carrying less than 69 kilovolts (kV), and everything above is classified as the transmission grid.[2]

The separation of the two is crucial for the practical applicability of the project. Transmissions grids, hereafter called high voltage (HV) lines, are provided by a single national agency in most countries, making data collection simple. Data collection of distribution grids, hereafter called medium voltage (MV) lines, suffer from more fragmentation as governments tender regional work to regional suppliers. This paper's experiments and purpose predict the layout of medium voltage lines while a country's high voltage lines are an essential input parameter.

2.3.1 Energy Transmission Line Design

Multiple studies have applied pathfinding algorithms for optimizing power line design. The goals of these studies were not to validate the layout of an existing grid but to suggest a path for new builds. The methodology varies from linear programming to evolutionary algorithms [25], [26]. Regardless of methodology, most studies implement a penalty function or cost structure aiming to mimic economic and environmental costs, for instance, in [27], [28]. Two commonly cited cost structures were proposed in [29], [30]. The cost structures consist of both environmental and economic parameters. The primary factor is the length of the path, and the second factor is closeness to roads. Both papers implement a cost parameter based on buildings, requiring planned paths to avoid buildings. Nighttime lights are not available at resolutions where avoiding buildings makes sense and was therefore discarded. Further, both papers suggest a parameter based on soil type or similar to incorporate the risk of landslides and a parameter for protected areas.

2.3.2 Documenting Power Lines

A broad literature review suggested two main methods for documenting power lines. The deciding factor between the categories was the dependency on human interaction. Traditional mapping methods are time-consuming and manual processes that require significant resources. Modern cartography research has focused on creating tools and applications simplifying the mapping process [31], [32].

[2]This distinction is based on the ANSI standard https://electrical-engineering-portal.com/ansi-standard-c84-1

On the other hand, simultaneous advancements in machine learning and remote sensing have presented exciting research areas, such as detecting man-made objects from outer space. Support vector machines and neural networks have produced extensive and accurate spatial data sets of street networks [33], [34]. The resolution of remote sensing has previously limited the detection of man-made objects to large objects [35]. Newer satellites and sensors can provide data at centimetre resolution, reducing the detectable object size requirements. An object detection approach for high voltage pylons was implemented in [32]. The project applied machine learning helping mapping professionals efficiently document transmission lines in Nigeria, Zambia, and Pakistan. Object detection algorithms were trained to suggest areas of interest, likely containing a pylon. The area-of-interest proposed by the algorithm was subsequently checked by a human who pinpointed the pylon's position. The approach optimized the transmission line mapping process. We considered the implemented methodology unsuitable for distribution grid inference due to different prerequisites. Differences between transmission grids and distribution grids include the size of structures supporting lines and the use of underground cables. Detecting giant steel pylons is significantly easier than detecting a wooden pole, and underground cables are naturally undetectable at the surface.

In [11], Rivera created OpenGridMap. OpenGridMap was a platform providing data sufficient for simulation studies. The project was discontinued, but the proposed methodologies for power line inference are still highly relevant. The initial paper provided an application for efficiently mapping electrical infrastructure. The application gathered location, images and labels[3] from crowd-sourcing communities. A professional validated mapped features before being accepted to the platform, thus guaranteeing correct labels and accurate data. The authors proposed various methods for distribution grid inference utilizing crowd-sourced data. The authors do not perform any quantitative experiments but provide figures of inferred power lines. Based on visual inspection, the most promising approach applied Dijkstra's algorithm to connect buildings, travelling along streets and using transformers as the source. The algorithm was implemented at a neighbourhood-level scale.

A variation of the power line inference methods proposed in OpenGridMap was further studied in [36]. In 2019, a team at Facebook Engineering proposed a modified version of Dijkstra's algorithm using nighttime lights as targets [24]. Their work was further explored by Arderne [36] who created an open-source application, *Gridfinder*, for power line prediction using the modified version of Dijkstra's algorithm. The methodology differs from other related research in

[3]I.e. Substation, transformer

terms of providing a global-scale method. The study acquired official distribution grids in 14 selected countries for validation. The countries represented a wide range of electricity access rates and income levels to indicate results' quality globally. The team introduced a minimum Intersection-over-Union (IoU) score of 0.5 for a satisfactory prediction. Predicted power lines were validated at decreasing resolutions until the IoU-criteria was met. The highest validation resolution was 500m, which was the resolution of night lights provided by VIIRS/DNB, and the lowest was 100 km. The average country exceeded the 0.5 IoU requirement at a resolution of 15 km. *Gridfinder* incorporates documented power lines in OpenStreetMap through a cost matrix, where all cells containing power lines are explored without additional cost. Dijkstra is a greedy algorithm, exploring the lowest costs first. Zeroing OSM power lines' costs implies that they will be a part of the predicted distribution grid.

Chapter 3

Methodology

This chapter presents the methods and processes used in the project. Firstly, the preprocessing and methods necessary for creating the targets raster and cost raster are presented. Secondly, the chapter proposes a configuration of costs and filters based on the experiments. Lastly, the path prediction method, Dijkstra's, is reviewed.

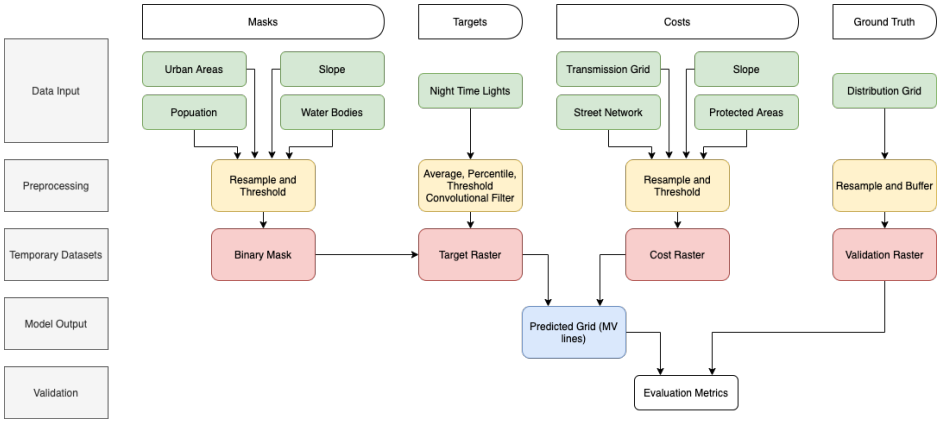


Figure 3.1: Simplified overview of the project

Figure 3.1 illustrates the data flow and order of operation in the project. Green entities represent input data, and input data are split into four categories. Masks

are data layers used to mask and remove targets. The shortest path algorithm will connect targets and uses costs data layers as guidance when predicting distribution grids. The last category of input data is the ground truth. A yellow entity indicates preprocessing and contains keywords for required steps. Red squares indicate datasets after preprocessing of the input data. Temporary datasets (red entities) are used as masks, model inputs or for validation. The blue entity illustrates the output of Dijkstra's shortest paths. Finally, the predicted grid is validated against the validation raster, which produces evaluation metrics. The figure fails to illustrate the iterative process where improvements in the evaluation metrics are used as the basis for the inclusion or exclusion of costs and masks.

3.1 Targets Raster

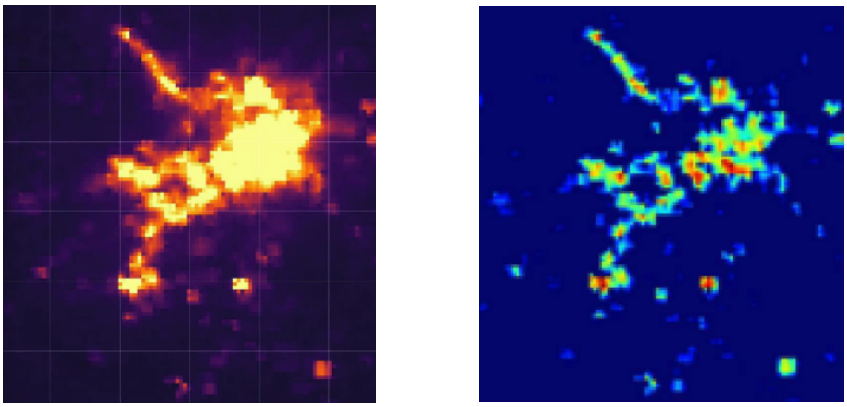
The targets raster is a binary raster of electrified settlements. A target cell is the representation of an area assumed connected to the grid. A target cell can also be described as an electrified settlement, a pixel of the nighttime light raster, averaged over 12 months, emitting stable light radiance. Dijkstra's primary goal is to connect all targets. Lines connecting targets make up the predicted layout of the distribution grid in a country. This section introduces the intended logic of filters and the processing of targets. Nighttime lights were assumed to be an appropriate consumption proxy, supported by the promising electricity access results. We chose VIIRS/DNB as the nighttime light data source due to the vast amount of research and preprocessed composites, facilitating power line prediction focus.

3.1.1 Preprocessing Night Time Lights

VIIRS/DNB captures anthropogenic and natural light sources. When applying night lights to model electrical grids, it is crucial to filter out natural lights, as they are not connected to any electrical grid. This study averages 12 months of radiance values, the entire year of 2020. Averaging composites limits the impact of ephemeral lights. Examples of ephemeral lights are forest fires and gas flares. A fire lasting for a week would inflate radiance values in that month's composite but would be negligible after 12-month averaging.

Further, a convolutional filter is applied to the averaged night lights raster. Radiance values of raster cells are dependent on the intensity of neighbouring cells; adjacent cells to a bright cell are also bright. The convolutional filter's intended

purpose is to concentrate light at the source. The convolutional filter is a counterpart to saturated target areas. Urban areas and their surroundings tend to all pass the binary threshold, resulting in a large cluster of target points. Convolution concentrates radiance values in the centre of a high-radiance cluster.



(a) Before convolution (b) After convolution

Figure 3.2: Impact of convolution filter visualized

Figure 3.2 shows the result of applying the convolutional filter. Pixel values in the left image [3.2a] are emitted radiance, where bright yellow indicates high radiance. Dark blue in the right image is pixels without light radiance. The filter and images were created by Facebook Engineering [24].

The final step of preprocessing night lights applied a binary threshold. The threshold created a binary targets raster, where 1-valued cells represent a spot connected to the grid. For example, the red pixels in Figure 3.2b were most likely above the threshold and classified as target points.

3.1.2 External Masks

This paper proposes four different data layers for the removal of target points. A data layer containing a country's water bodies was applied to remove target points on the water. Possible lighting sources on the water are fishing boats, light pollution from a nearby city or lighthouses. Although anthropogenic, these targets are unlikely connected to the distribution grid. The water mask removes yellow cells on water. Two different population thresholds were examined; pop-

ulation greater than zero and greater than ten. The intended logic of a population threshold mask is that documented human settlements are more likely to be connected to the electrical grid. The criteria require electrified cells to have documented population. Thus, removing all target points with no documented population. The population layer is proposed as a mask, not a cost, considering that large parts of the Sub-Saharan population live without power.

The remaining two masks are a slope layer and an urban areas layer. The slope mask removes all electrified settlements where the slope is above 30%. The filter eliminates targets in hillsides or other places where electrified settlements were deemed unlikely. Lastly, a mask of urban areas intends to remove all target points where no buildings or settlements are documented. Urban areas are a symbol of human presence and activity and represent points of electricity consumption. The expected downside of an urban-area mask is the removal of target points in rural areas.

3.2 Cost Raster

The secondary raster in this project is the cost raster. Its task is to guide Dijkstra's Algorithm by penalizing or rewarding specific paths. The optimal cost raster imposes constraints similar to those considered in a real-world grid design. For example, the shortest path algorithm is rewarded (less penalized) for choosing a road path, simplifying maintenance access. A cost raster is a model of real-world grid design and would ideally enforce all grid design constraints.

$$C = min(R, L) \tag{3.1}$$

Two data layers comprise the cost raster in [36], OSM power lines and OSM Roads. The implementation differentiates between the various road types (Table 4.2) and all power lines are zero-cost. If an area contains multiple costs (eq. 3.1), the minimum is assigned. This master thesis implemented additional cost data layers, which required reformulation of cost calculation.

$$C = (\alpha R + \beta S + \gamma P) \circ L \tag{3.2}$$

In equation 3.2, C denotes the cost raster, R the roads, S the slope, P protected areas, and L is the HV-lines raster, and α, β, γ are scalars. The reformulation facilitates the combination of multiple cost layers. A cell's cost with a high slope in a protected area is now due to both parameters. The presence of HV-lines is still the dominant factor. The HV-lines raster is constructed such that 0-values

indicate HV-lines, and all other cells are valued 1. The cost raster is the element-wise product of the HV-lines raster and summed cost of R, S, P. A cell's cost is zeroed if HV-lines are present but remains unchanged otherwise.

3.2.1 Continued Cost Data Layers

The transmission grid plays the role of the trunk in the electrical grid tree, while distribution grids represent branches. HV-lines are incorporated in the cost raster as distribution grids are always connected to the transmission grid. A zero cost cell will consistently be expanded by Dijkstra, ensuring free travel along with a nation's transmission network. In addition, the street network in a country is implemented to reward travel in cells containing roads. Roads are a crucial factor in grid design as they ease maintenance access and are designed to minimize travel time between human activity hotspots. Further, various road categories have different costs to take road size into account. Street network's impact in distribution grid prediction or planning is widely recognized [11], [25], [30], [37] also found improvements when differentiating costs by road type.

3.2.2 New Costs Data Layers

Protected areas and national parks are introduced as a new cost metric with the motivation to penalize power lines in protected areas. Norwegian authorities require additional research, impact assessments, and permissions if intervening with national parks[38]. The same logic was assumed relevant in all European countries. A wide range of protected area categories exists, often with different levels of regulation. The variations in regulations across countries and protected area types complicate the differentiation of costs. This study generalizes and considers all protected areas the same. Finally, the same slope raster introduced as a target mask is also proposed as a cost raster. Power lines are designed to avoid touching the ground or underlying vegetation at all times. The intended logic of slopes as a cost parameter is to penalize traversals where an increased number of pylons are required to cope with rapid elevation changes.

3.3 Optimizing the Configuration

In order to find the optimal configuration of masks and costs scalars, new features will be validated in a step by step process. The first goal is to implement a baseline

model with parameters similar to what has been tested in previous studies. The second goal is to improve on the state of the art method. The experiments will firstly determine the optimal target raster. Modifiable parameters in the target raster are the radiance threshold and different masks. The best performing radiance threshold will be the threshold used in further optimization. Masks will be tested similarly, meaning that a mask is kept if it improves the results. All changes are validated in all countries and only kept if it improves the average result. A downside of this step-by-step process is that it will not detect synergies between all combinations of masks. However, the masks proposed above remove different targets and performance loss of not testing all possible combinations was considered negligible.

The cost optimization process aims to find the best combination of scalars. Equation 3.2 contains three tunable parameters; the roads scalar, the protected areas scalar, and the slope scalar. All possible combinations of scalars, 0, 1, 5, 10, 100, will be tested for the protected areas and slope layer. By keeping the roads scalar constant, improvements in validation metrics can be attributed to the proposed new cost data layers.

3.4 Validation Process

The following sections present evaluation metrics, how predictions were validated, and the limitations of the validation method. To properly assess the quality of a prediction, several evaluation metrics were used. The predictions of the algorithm were validated in a pixel-by-pixel manner. A positive prediction describes a pixel where the algorithm outputted grid, while a negative prediction means no-grid. The main components of all evaluation metrics are the classifications of the pixel predictions. A **True Positive (TP)** is the event when both the prediction and the ground truth agree on the presence of grid in a pixel. **False Positive (FP)** if the prediction is positive, but no power line is documented in the ground truth dataset. **False Negative (FN)** if no-grid is predicted, but the ground truth pixel contains power lines. Lastly, **True Negative (TN)** if both ground truth and the prediction is negative. The guess classifications presented above can be further combined into evaluation metrics.

Precision is a validation metrics that represents the ratio of TP over the total of positive predictions. A low precision score is the result of a large amount of

false positives, implying an overestimation of positives.

$$Precision = \frac{TP}{TP + FP} \tag{3.3}$$

Recall measures the ratio of found positives overall existing positives. A high recall score suggests that the algorithm was successful in finding/recalling most of all existing positives in the ground truth dataset. A positive prediction in all pixels will achieve a perfect recall score. As precision penalizes the presence of false positives, the two metrics are often used in parallel.

$$Recall = \frac{TP}{TP + FN} \tag{3.4}$$

F1 Score is a metric that combines Precision and Recall. The combination eliminates the shortcomings of individual metrics. A high precision and high recall results in a high F1 score.

$$F1 = 2\frac{Precision * Recall}{Precision + Recall} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{3.5}$$

Accuracy Accuracy is a measure of a prediction's overall correctness. The metric is commonly used, but the descriptive property is reduced when applied to an unbalanced dataset. This thesis does not apply accuracy as a validation metric due to the significant imbalance in the presence of grid versus no-grid. The average European country consists of 98% no-grid pixels. In the average case, an all no-grid prediction achieves an accuracy score of 98% while completely failing to predict the layout of the distribution grid. Due to the imbalance between positives and negatives in this application, accuracy was not measured.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.6}$$

Intersection-over-Union Intersection-over-Union (IOU) is a validation metric introduced to fill the gap of accuracy in imbalanced datasets. The metric does not include TNs, thus avoiding score-inflation. IOU is applied in object detection tasks and a score above 50% or 0.5 is considered satisfactory by industry standards.

$$IOU = \frac{TP}{TP + FP + FN} \tag{3.7}$$

3.4.1 Buffering of Ground Truth

When validating the predicted power lines, it is interesting if a prediction is right or wrong and how wrong it is. A prediction parallel and close to a real-world grid line is an interesting finding and not directly wrong. In order to reward the algorithm for selecting a path close to a real-world power line, the ground truth vector data set is buffered by 1000 meters before being rasterized. The buffering will classify predictions within a kilometre of actual power lines as true positives.

The logic behind the implementation of the buffer is related to the early screening phase of a project. Power producers have expressed a need to discard completely unrealistic sites, for example, where the closest infrastructure is many kilometres away. Validation without the buffer would set higher requirements for the predictions in terms of accuracy.

3.4.2 Processing Predictions

High voltage lines are included in the cost world as zero cost. The method will predict power lines in all zero-cost cells. Thus, HV-line predictions can not be counted as true positive. The predicted power lines raster will be element-wise multiplied with the high voltage lines raster. High voltage lines are zero-valued and will remove all predictions where they are present. The process avoids artificially good results, where validating rewards recreation of the input data.

Chapter 4

Experiments and Results

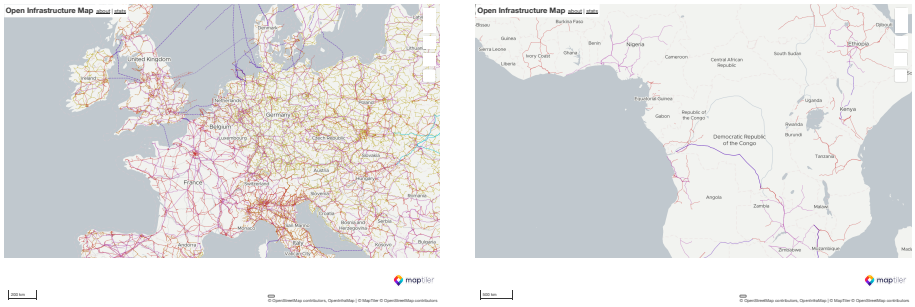
The following chapter is comprised of two main sections. The first section provides all steps necessary to reproduce the experiments carried out in this thesis. We introduce the datasets, where and how to download them, and the required preprocessing for a dataset to be compatible with the methods. The second section presents the experimental results. Results are presented and analyzed in two different manners. The first analysis is a comparative study where improvements are measured against a baseline model, and the second presents analysis results through case studies. The case studies illustrate the pros and cons of the implemented methodology.

4.1 Experiments

The conducted experiments inferred distribution grids in 34 European countries. Each country was predicted on a stand-alone basis, thus missing potential synergies across national borders. These implications would be detrimental to transmission line predictions but were considered negligible for distribution grids. The experiments were computationally costly but were conducted using parallel programming. The implementation facilitated efficient run-through of approximately 30 parameter configurations. Each configuration ran for all 34 different countries.

4.1.1 Selection of Study Area

Europe was selected as the region of interest due to their leading position within open-source infrastructure data. The validation process of this project relied solely on open-source data, and the quality and completeness of that data are of high importance. Figure 4.1 illustrates the vast completeness differences in OpenStreetMap power line data when comparing Central Europe to Sub-Saharan Africa. The visualization was provided by www.openinframap.com[39]. Despite lacking power line data, Sub-Saharan African countries have substantial datasets with street networks [40]. Research Goal 2 was to determine the reliability of grid predictions, thus requiring validation data. The initial selection included all European countries, but challenges concerning size and data availability excluded some countries [Table 4.1].



(a) OSM power line data in Europe (b) OSM power line data in Africa

Figure 4.1: Comparison of the documented power lines in Europe and Africa

Admin Boundaries [41] is a dataset with information about a country's borders. Natural Earth Data provides global coverage, but only European boundaries were downloaded. The minimum mapping unit or the resolution of Admin Boundaries is 10 meters. Modifications to the out-of-the-box admin boundaries are due to complications with size, outlying islands, or location. Table 4.1 lists countries and the modifications made.

Country	Modifcation	Cause
France	Trimmed	Outlying Areas (French Guiana)
Portugal	Trimmed	Outlying Areas (Azores, Madeira)
Netherland	Trimmed	Outlying Areas (Antilles)
Norway	Dropped	N of slope-dataset
Sweden	Dropped	N of slope-dataset
Finland	Dropped	N of slope-dataset
Island	Dropped	N of slope-dataset
Vatican	Dropped	Size (Too Small)
Monaco	Dropped	Size (Too Small)
Liechtenstein	Dropped	Size (Too Small)
Andorra	Dropped	Size (Too Small)
Russia	Dropped	Size (Memory Limitations)

Table 4.1: Modified countries

The primary function of the admin boundary file is the clipping of nighttime lights and other datasets to each specific country. In addition, the admin boundary file functions as an administrator of processes. *Gridfinder* runs processes on a country-by-country basis and uses the list of countries in the admin boundary file as a list of tasks to be completed.

4.1.2 Datasets Applied for Grid Layout Prediction

A wide range of datasets from multiple sources was used to arrive at the final product of estimated grid layouts all over Europe. This section describes the datasets' attributes, source, and preprocessing required. Firstly, the nighttime lights are presented. The datasets used in cost calculations, street networks, transmission grids, slope, and protected areas are covered. Finally, the section covers mask datasets and the ground truth; population, urban areas, water bodies and distribution grids.

4.1.3 VIIRS Night Time Lights

Night Time Lights (NTL) can be downloaded from the Earth Observation Group [42]. NTL are provided at a spatial resolution of approximately 450 meters at the equator. The EOG provides daily, monthly, and annual composites of night lights. This experiment uses all monthly composites of 2020. The final NTL target

raster is based on 12 months of data, the 70th percentile of each grid cell and a threshold of $0.2\frac{W}{m^2 sr}$. The threshold was chosen based on anecdotal experiments. These steps reduce the presence of ephemeral and natural lights. This raster was compatible with Dijkstra's, but a convolutional filter and masks were applied to increase the quality of results. The NTL raster was the fundamental matrix, and all other datasets were resampled to match the resolution of this raster.

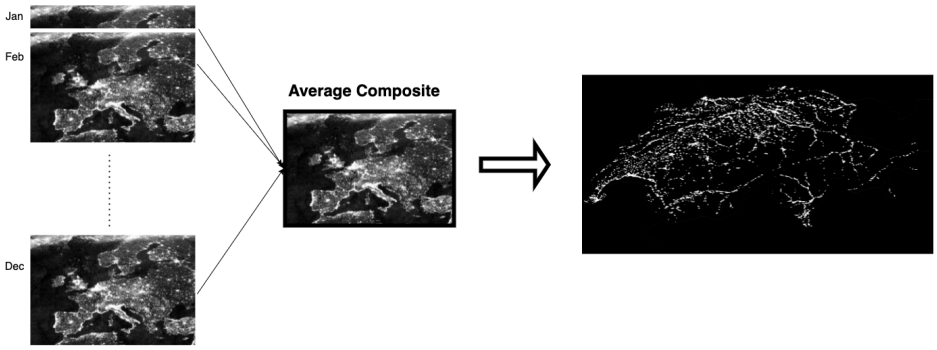


Figure 4.2: Process from raw data to targets with night time lights

Figure 4.2 illustrates the preprocessing of night lights. The rightmost image in Figure 4.2 displays a targets raster in Switzerland.

4.1.4 Costs

OpenStreetMap OSM data [43] was downloaded for all of Europe in a single file. Geofabrik[44], a German OSM consultancy firm, provides downloadable osm.pbf files containing all available OSM data by region. This study downloaded the European file and later clipped the data to individual countries using poly-files. Polyfiles are sequences of points describing the administrative boundary of a region or country. Using the processing tool osmfilter[45], the OSM data of interest, roads and power lines were extracted from the pbf files and converted to geopackages.

Roads After the abovementioned steps, road networks were available as geo-packages for all European countries. Rasterization of the street networks was necessary as geo-packages are vector files. The conversion was done using rasterio's rasterize [46], where the night light raster was used as target resolution. OSM road network data contains a high-level categorization of road types, ran-

ging from service roads to motorways. This categorization was used as the basis for the raster burn-in values shown in Table 4.2. The lowest cost is assigned if a pixel contains multiple road categories.

Category	**Costs**
Motorway	1/10
Trunk	1/9
Primary	1/8
Secondary	1/7
Tertiary	1/6
Unclassified	1/5
Residential	1/4
Service	1/3

Table 4.2: Road categories and costs

Power Lines Power lines go through the same conversion process as the street network [43]. OSM has labelled data with a "power" tag, facilitating for power line filtering. Data labelled with the key-value pairs power:line and power:cable extracted layout and voltages. The power line dataset was then split into costs and validation depending on the voltage level of the line or cable. HV lines made a cost layer while MV lines created the validation set. In Figure **??**, HV lines are COLOR, and MV lines are COLOR. The power line cost dataset for a country is valued at 0 when a raster-cell contains a transmission grid and one elsewhere.
Slope A slope dataset was created using Shuttle Radar Topography Mission's (SRTM) Digital Elevation Model (DEM)[47]. DEM tiles are available for download in two sizes, 5 x 5 degrees and 30 x 30 degrees. Tiles N30W030, N30E000 and N30E030 (Figure 4.3) provide a near-complete coverage of European countries. The data has a spatial resolution of 3 arc-seconds, approximately 90m. The slope dataset was created using the built-in DEM to slope method in QGIS. This method requires a scale factor, which was calculated using the average latitude of the region.

$$s = 111320 * cos(Lat * \frac{\pi}{180^\circ}) = 78715 \tag{4.1}$$

The scale factor (4.1) will accurately convert from DEM to slope in areas close to 45^{th} parallel north, but errors increase the further away from the 45^{th} parallel they get. Slope-costs are derived from a rough categorization, displayed in Table 4.3. While errors in slope calculations will lead to the misclassification of some raster cells, the impact was considered negligible. The slope raster was resampled by averaging to match the spatial resolution of nighttime lights and clipped to individual countries.

Degrees	Costs
$30 <$	3
$20 - 30$	2
< 20	1

Table 4.3: Slope costs

A downsampled slope raster cell's value is the basis of the assigned cost. The cost of slope values are found in Table 4.3. The configuration of slope costs is based on anecdotal experiments. NASA does not provide SRTM DEM data above 60 degrees north. The lacking coverage forced removal of northern European countries [Table 4.1].

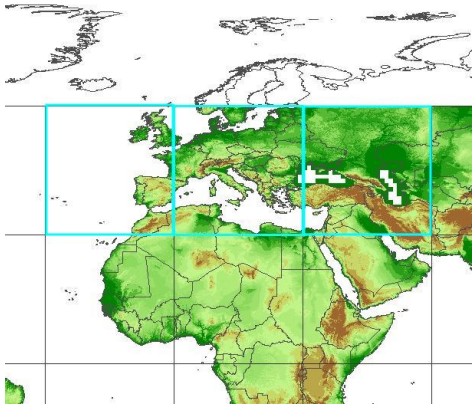


Figure 4.3: Downloaded DEM data in Europe.

Protected Areas The protected areas dataset was downloaded from Protected Planet[48], a platform with extensive geographical data regarding terrestrial and marine protected areas. The exact dataset used for generating a cost layer is the World Database on Protected Areas (WDPA). WDPA is a global dataset. Preprocessing of WDPA consisted of rasterization and resampling to night light's resolution before clipping to individual countries. Cells within a protected area are assigned a value of 1; all other cells are valued at 0.

4.1.5 Filters

Population Population data was downloaded from the Global Human Settlement Layer (GHSL)[49]. GHLS provides data related to human presence on the planet through data sets such as population and built-up areas. The spatial resolution of the population data is nine arc-seconds, approximately 270m, and covers the whole globe. The population raster was resampled by averaging to NTL's resolution. A binary threshold was applied to the resampled population layer to create a mask. Two masks were created, greater than zero and greater than ten.

Land Cover The 2019 land cover classification dataset is provided by the Climate Change Initiative (CCI)[50]. The raster values inform about the type of land cover based on the Land Cover Classification System (LCCS)[51]. The raster comes with a spatial resolution of 300m. Therefore, resampling was necessary to match the night lights raster. LCCS resampling was done using the median, as values are categorical. Two masks were extracted from the LCCS dataset; an urban area mask and a water bodies mask. Urban areas are denoted by a cell value of 190 and water bodies by 210.

Slope The same slope dataset used for costs was also applied to create a slope mask. The difference in processing from costs to mask is the burn-in values assigned. A slope mask cell was valued one if the slope was less than 30 degrees and 0 if above, thus removing all targets in steep hills.

4.2 Results

The following section will present two sets of results. Each set of results is linked to a specific research goal. Firstly, we introduce the baseline model, as implemented in [36]. Further, we validate the proposed improvements from the Methodology chapter (3). The suggested improvements include new masks and new cost layers. Model iteration results are analyzed comparatively. This aspect is linked to improving on the state of the art methods.[1] Secondly, we present maps and key metrics from two countries, Switzerland and Spain, and an overview from the complete data set. Switzerland is an outlier in terms of outstanding performance, while Spain represents the average country and results. Their metrics are reviewed in a general perspective and will be used to answer Research Goal 2.

[1] Research Goal 1 [Sec 1.2]

4.2.1 Baseline Model and Suggested Improvements

Model iteration results are presented in two categories; masks and costs. Results presented describe the average results of all countries unless stated otherwise. The baseline model uses a brightness value threshold of 0.1 and road networks as costs. Complete tables of results for each model can be found in Appendix .1.

Masks

Table 4.4 shows the obtained evaluation metrics for different model configurations. All masks removed target points from the targets raster, which limits the chance of increased recall rates with additional masks. Additional masks decreased recall rates in 84% of the individual country simulations.

	Precision	**Recall**	**IoU**	**F1**
$Baseline$	0.303	**0.868**	0.279	0.430
$Water_{mask}$	0.303	0.868	0.279	0.430
$Water_{mask} + Pop_0$	0.311	0.844	**0.282**	**0.433**
$Water_{mask} + Pop_{10}$	0.318	0.791	0.275	0.425
$Water_{mask} + Pop_0 + Slope_{mask}$	0.310	0.840	0.280	0.431
$Water_{mask} + Pop_0 + Urban_{mask}$	**0.418**	0.447	0.235	0.373

Table 4.4: Mask configuration results

The water mask shows no changes after rounded to three decimals but produced slightly higher results than the baseline model. The two different population masks produced different results, where Pop_{10} improved precision twice as much as Pop_0. However, Pop_0 achieved higher metrics in the remaining three categories. We expected the Pop_{10} mask to provide lower recall scores as the mask removes the same target points as Pop_1, and many more. Pop_{10} achieved overall lower results than the baseline model. The next mask implemented was $Slope_{mask}$. $Slope_{mask}$ performed worse than its predecessor, Pop_0, by all metrics. Finally, target points were filtered to by urban areas. The urban filter significantly increased precision but almost halved recall rates. Pop_{10} and $Urban_{mask}$ obtained the two best precision scores, which also removed the most target points. Both masks removed noise from the target raster but removed many valid targets in the process.

Costs

Table 4.5 shows the results of various cost configurations. $PA_{cost=100}$ means that the scalar of the cost layer, Protected Areas, is set to 100.[2] The cost experiments used target rasters filtered by $Water_{mask} + Pop_0$, and their results should be compared to the scores in row two. The cost raster with protected areas costs of 1 and slope penalization multiplied by 5 achieved the best scores.

	Precision	**Recall**	**IoU**	**F1**
Baseline	0.303	**0.868**	0.279	0.430
$Water_{mask} + Pop_0$	0.311	0.844	0.282	0.433
$PA_{cost=1} + Slope_{cost=0}$	0.320	0.826	0.288	0.441
$PA_{cost=0} + Slope_{cost=1}$	0.327	0.813	0.291	0.445
$PA_{cost=100} + Slope_{cost=100}$	0.328	0.806	0.290	0.444
$PA_{cost=1} + Slope_{cost=5}$	**0.330**	0.807	**0.292**	**0.446**

Table 4.5: Cost configuration results

Changing costs had little impact on results, as can be seen if comparing the most extreme cost configuration of $PA_{cost=100}+Slope_{cost=100}$ with the best performing configuration. Regardless, the introduction of protected areas and slopes as a cost layer improved the overall results.

4.2.2 General Results and Case Studies

While the first section focused on comparing different configurations, this section focuses on the overall quality of the results. Table 4.6 contains the metrics obtained for Switzerland, Spain, and the average results. Presented results and maps are created with mask $Water_{mask} + Pop_0$ and cost configuration $PA_{cost=1} + Slope_{cost=5}$. Three maps were produced for each country, the first illustrating the cost raster, the second presents target points and predicted power lines, and the last contains predicted power lines and buffered ground truth.

[2]Referring to scalars in cost equation 3.2

	Precision	**Recall**	**IoU**	**F1**
Average	0.3281	0.7980	0.2889	0.4426
Switzerland	0.5747	0.8989	0.5398	0.7011
Spain	0.2905	0.7678	0.2670	0.4215

Table 4.6: Evaluation metrics in Switzerland and Spain

Switzerland

Measured by F1 score, Switzerland performed five percentage points better than the second-best and 35 percentage points above average. Figure 4.4 illustrates the extraordinary nature in Switzerland, with the Swiss Alps dominating the south. The large red areas are a result of steep hills, greatly increasing the cost of travel. Although steep hills increase the path cost, they are often accompanied by valleys. Valleys are suitable for roads, power lines and human settlements. In the enlarged focus area in Figure 4.4,. The juxtaposition of steep hills and valleys reduces the freedom for Dijkstra's and practically enforces a path.

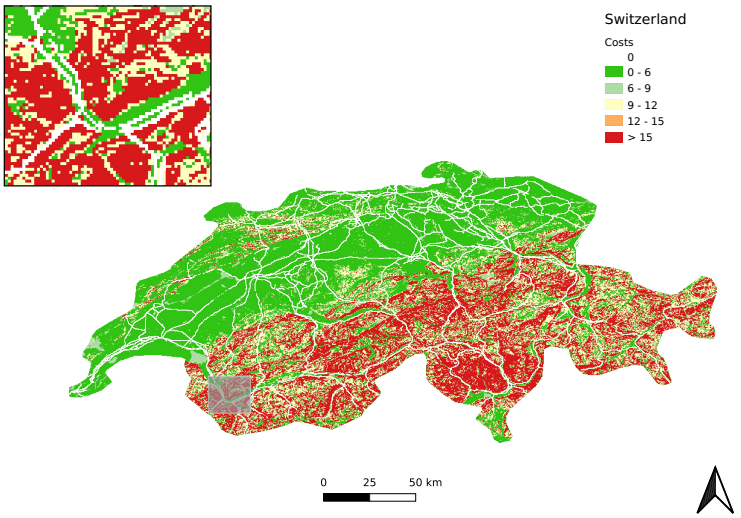


Figure 4.4: Cost raster in Switzerland

Figure 4.5 illustrates the target raster and predicted power lines combined. All yellow dots and areas are connected by black lines representing electrified set-

tlements and the predicted grid. The map illustrates how no yellow pixel, an electrified settlement, is left unconnected to the grid. This feature is desirable in countries with a single electricity grid but undesirable in countries with significant usage of off-grid power plants. Switzerland and most European countries have a single national high voltage network functioning as a stem for the distribution grid tree. The inhabitable area in the south of Switzerland is almost exclusively in the valleys. Target points fall in natural lines, which help Dijkstra's algorithm finding the correct path. If we view the enlarged areas in figures 4.4 and 4.5), we can see that the cost world succeeds in facilitating real-world grid design. The predicted power lines lie in the green lane of the cost world. The enlarged area is the same in both maps, the city of Martigny and its surroundings.

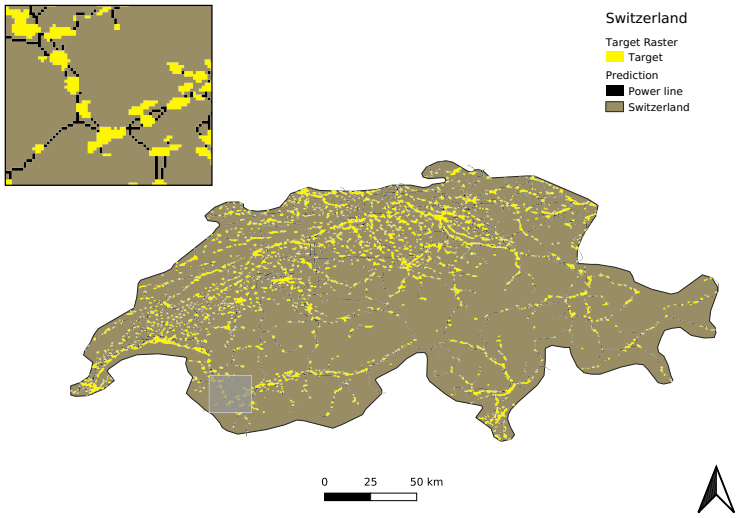


Figure 4.5: Predicted power lines in Switzerland

The last map combines the prediction with the actual power lines in Switzerland. In Figure 4.6 the thick green lines represent the ground truth, and the black lines are the predicted power lines. The density of power lines in the northern areas of Switzerland creates an overweight of valid prediction layouts. The map illustrates complications of buffering ground truth in countries with near-complete documentation of power lines. The focus area in the top left corner of the figure attempts to visualize a common pitfall of the method. The enlarged map illustrates how to model predicted power lines connecting target points from the west, while the ground truth connected the central area from the east.

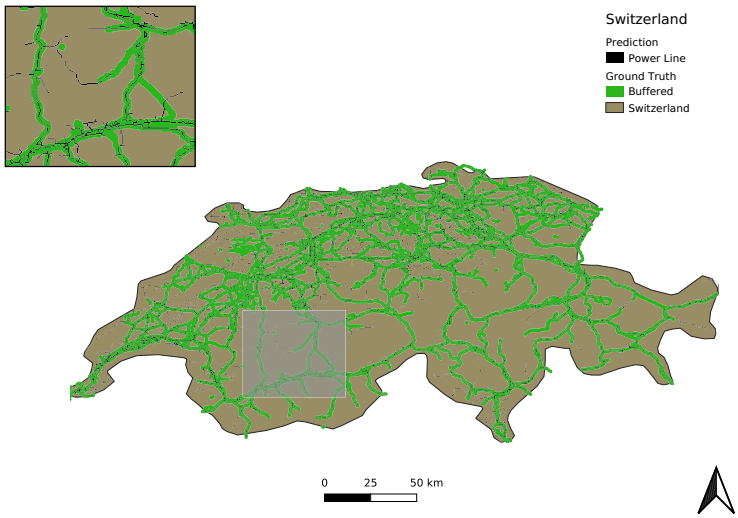


Figure 4.6: Predicted power lines and buffered ground truth in Switzerland

Spain

The evaluation metrics obtained in Spain were most similar, measured by absolute difference, to the average results. The low IoU metrics for Spain and the average country explains the results in a more general perspective. Object detection standards consider IoU scores above 0.5 satisfactory, meaning a minimum 50% overlap between the predictions and ground truth. The predicted power lines in Spain scored 0.2670, far below industry standards. The metrics alone suggest that the methodology used is unfit for its purpose, but some of the deficiencies can be explained by external parameters. The produced maps aim to illustrate errors explainable by the methodology as well as external parameters.

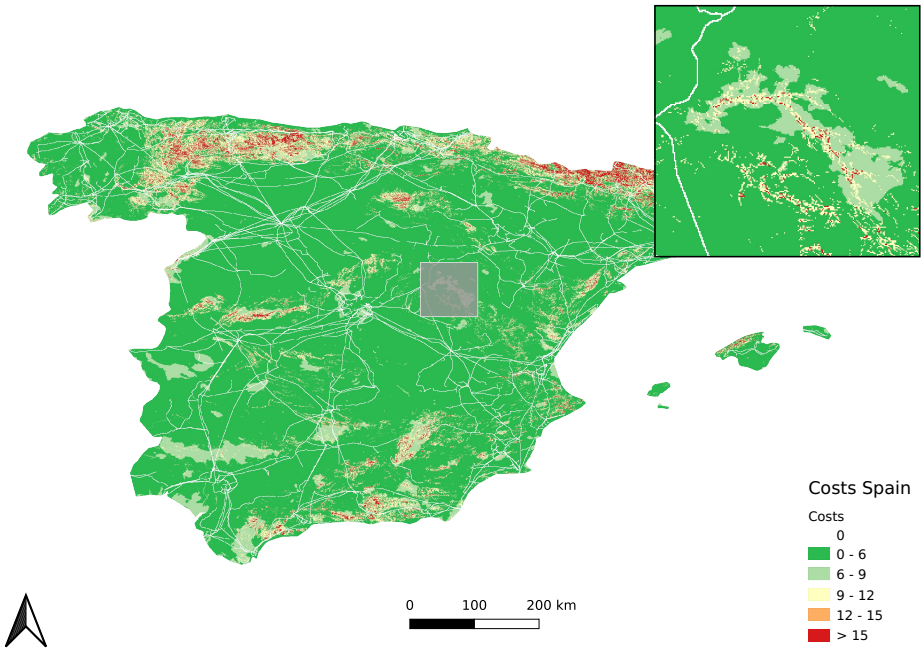


Figure 4.7: Cost raster in Spain

The cost raster in Spain (Figure 4.7) differs greatly from what we observed in Switzerland (Figure 4.4). Spain is a flatter country, and terrain proposed fewer limitations for the path prediction algorithm. Fewer restrictions reduced the likelihood of the algorithm accurately predicting the layout of the distribution grid and were likely the case in most countries. The roads cost layer, which was included in the baseline model, visualized in the appendix (Figure 1), was vital for pathfinding in large similar-cost areas (large green areas in 4.7). The enlarged area of the cost figure shows a wide range of costs; white lines indicate high voltage power lines. Green is a flat area, and light green is a flat, protected area. Yellow and red pixels indicate rough terrain, where red is the steepest or roughest.

Figure 4.8 shows the predicted power lines in combination with ground truth. The enlarged area illustrates the challenge with this study, the completeness of the validation data set. Small areas of documented power lines appeared disconnected from the nearby grid. The methods used assumed that all lights were connected, thus creating a fully connected grid, thus incorrectly predicting

grid lines where there were none. In addition, the focused area illustrates the difference in the level of detail. Night lights identified electricity consumption hotspots at a higher level of detail than what was documented in open source power line data. The level of detail mismatch was found in the majority of inspected countries; Switzerland was an exception. The figure also illustrates the differences in coverage in different regions of the country. The high density of green pixels in the northwest of Spain results from a more complete ground truth data set. OpenStreetMap mapping processes are often local initiatives, which can explain coverage differences.

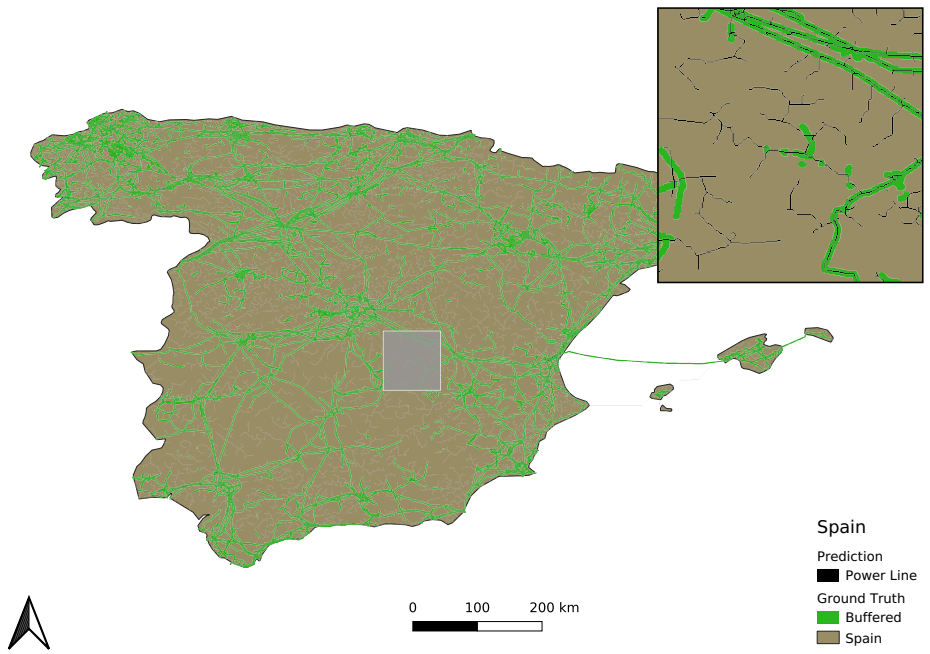


Figure 4.8: Predicted power lines and buffered ground truth in Spain

Palma, Ibiza, and Menorca east of the mainland created a new challenge for Dijkstra's. The predicted path left the mainland at Cap Martí, the easternmost point in the area, disregarding the inadequate infrastructure. In addition, the path's first destination is Ibiza (western island), then Palma (middle island), before arriving at Menorca (eastern island). The prediction illustrates the failure to incorporate consumption loads. Palma, the central island, is more prominent in size and population and requires significantly more power than the neighbouring

islands. The actual power lines leave the mainland further north near Valencia, a large city with adequate infrastructure, and goes directly to Palma.

4.2.3 General Observations

Cost methodology Figure 4.9 illustrates how the algorithm predicted power line paths near a protected area. The area is the same as the focus-area in Figure 4.7. The path tends to avoid crossing the protected areas, especially where steep hills separate target points. The illustration validates the intended logic of introducing the slope data layer and protected areas data layer as costs. While the methodology worked as planned, there was no ground truth data in the selected area to validated the predictions' correctness.

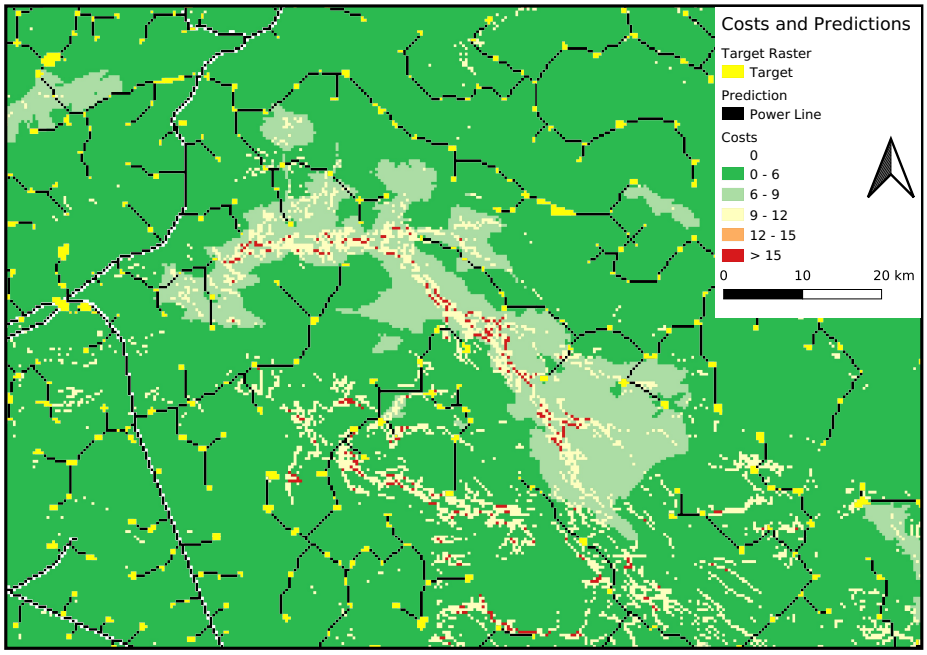


Figure 4.9: Predictions around a protected area

Repeated transmission line branching Figure 4.10 illustrates an undesirable result of method design, present because high voltage lines have zero cost. All distribution grids are connected to the transmission system, but real-world

transmission grids are not connectable everywhere. The algorithm predicted small MV-branches along 400 kV transmission lines, which would require a power substation and transformer. It would not be cost-efficient to build a power substation every 5 kilometres as the predicted paths would have required if they were correct. Thus, the methodology fails to take connection limitations into account. A different solution to power line prediction using substations as nodes could solve this problem.

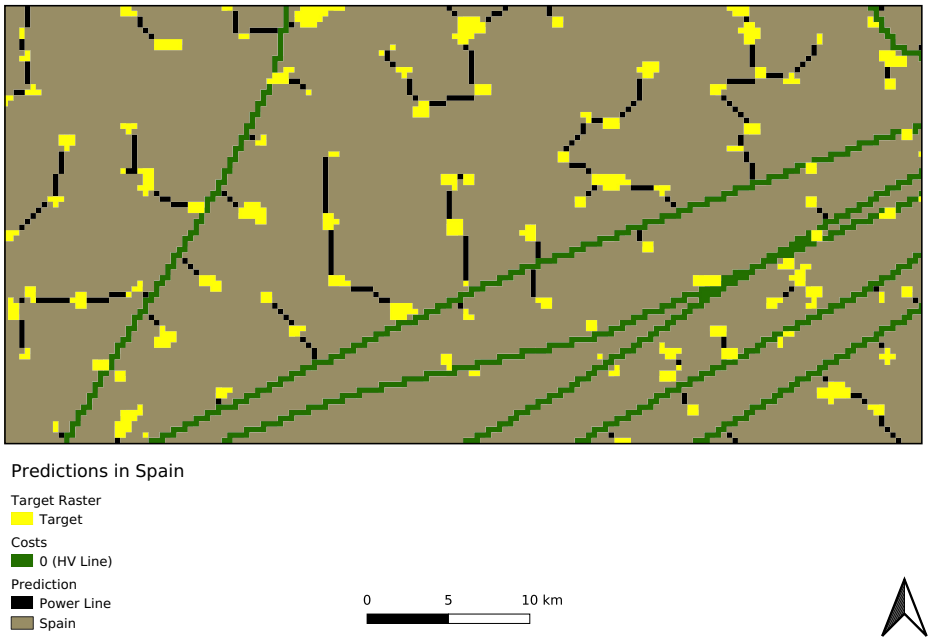


Figure 4.10: Transmission line branching

Parallel Grid Lines Power producers often build parallel grid lines for load and redundancy reasons. Redundancy does not make sense from a minimization of costs perspective, and power loads are not incorporated in the cost world. Thus Dijkstra's fails to incorporate these aspects of grid design. However, parallel grid lines are most commonly used on high voltage networks, which are zero-cost and automatically expanded by Dijkstra. In addition, the rasterization of vector power lines downsamples the resolution to 450 m. Power lines less than 450 m apart will be classified as a single power line. Figure 4.11 illustrates the two cases. The left image shows an area east of Valladolid, Spain, with parallel

high voltage lines at zero cost. The right image shows how parallel gridlines in Switzerland were combined under rasterization; green pixels are the black lines after rasterization.

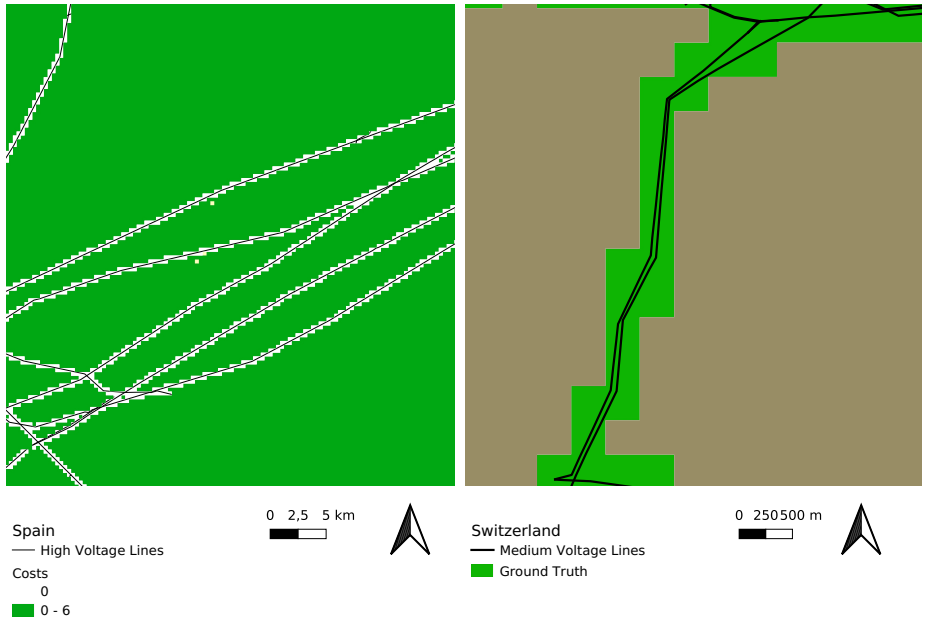


Figure 4.11: Parallel power lines

Chapter 5

Discussion

This chapter discusses the results in a more general context. Firstly, we discuss the overall quality of the results and how they compare to similar studies. The second section discusses OpenStreetMap power line data quality and whether it was suited for being used for validation. Continuing, we discuss the lack of formal requirements for the predictions.

5.1 Results

The results and metrics obtained were far below those presented in [36]. This thesis' implemented validation method expected lower metrics, as there was different usage of OpenStreetMap data. Their paper validated predicted grids in selected countries against official documentation and used all OpenStreetMap power lines as zero-cost. Comparatively, this thesis used only transmission lines as zero-cost. We argue that their results were too dependent on the quality of OpenStreetMap power line data in the specific country, and the validation process failed to explain the quality of night light grid prediction.

The results in Switzerland were far better than expected, but the performance was unmatched by other countries. Two key factors can explain the great results; the completeness of validation data and extreme terrain. The northern regions of Switzerland had well-documented power lines, and the buffering of ground truth created a forgiving world to predict in. Further, the extreme nature in the south of Switzerland concentrates target points in the valleys. The valleys were

easily identifiable in the cost world, and Dijkstra's pathfinding was simplified. A more representable set of results were obtained in Spain. The average results are lacking in the context of object detection standards. The methodology fails to consider several aspects of grid design, mainly the absence of transformers and consumption loads. Although the pathfinding algorithm has limitations, the selection of ground truth data sets significantly impacted the results. Populated areas all over Europe were without documented power lines.

Most of the existing literature on power line layout focused on planning, and an experiment's correctness was rarely quantifiable. Placing the results in a greater context is difficult, but qualitative assessments of predictions highlighted fundamental flaws in the applied methodology. The object detection results are not directly comparable to our results. However, their efforts suggest that high voltage power line documentation is a more significant challenge than assumed in this thesis. Usage of high voltage lines as costs is a crucial aspect of the methodology. We assumed that high voltage lines were available in most countries, which is correct for Europe, but the assumption will not necessarily hold in Sub-Saharan Africa.

5.2 OpenStreetMap Power Line Data Quality

Research Goal 2 focused on prediction quality. Quality and accuracy were important as predicted distribution grids could be applied in practical applications. Validation in Sub-Saharan Africa was avoided due to the lack of documented power lines. Documentation of high voltage networks in Europe is near complete [39], but distribution grid or medium voltage networks are under-documented everywhere. Figure 5.1 illustrates the difference in the level of detail between the predicted power lines and the buffer ground truth data set. The left image show predicted power lines and buffered ground truth, and the right image shows predicted power lines and population count; both images cover the outskirts of Madrid, Spain.

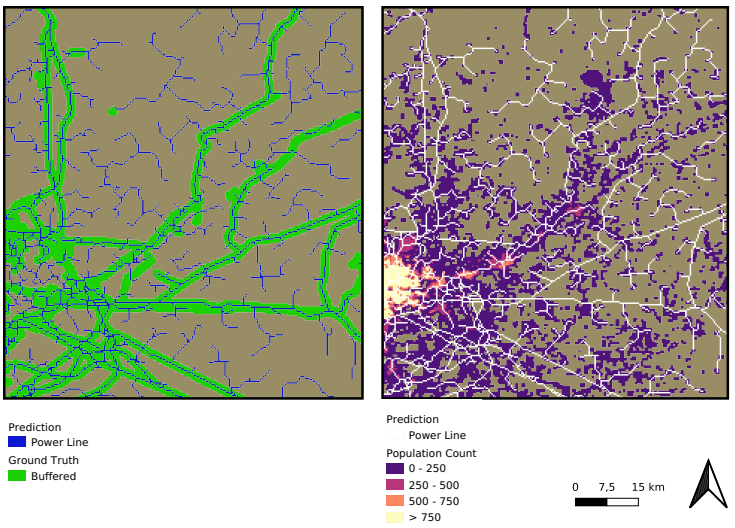


Figure 5.1: Distribution grid completeness example. Madrid, Spain

The bright area in the bottom left corner is Madrid, which is a high population density area. The purple spots east of Madrid are populated areas. As we can see from the figure, there is a large discrepancy between the ground truth in the left image and the purple areas in the right image. In 2020 a 100% of the population in Spain had access to power[52], which strengthens the case of an incomplete ground truth data set. Building on this, the observed values in the results section, tables 4.5 and 4.6, all have high recall and low precision values. Two things could explain low precision; Over-prediction of positives or incomplete validation data. We can conclude that low precision can be partially explained by incomplete data set.

While a more complete data set would likely result in higher precision, we expect decreased recall rates. In an attempt to measure the impact of incomplete data sets, we look towards Great Britain and Germany, which are global leaders in open source geospatial data. The table below, Table 5.1, presents evaluation metrics for Switzerland, Spain, Germany, Great Britain, and the average country. As we can see from the table, Germany and Great Britain perform comparably to Spain and the average country.

	Precision	**Recall**	**IoU**	**F1**
Average	0.3281	0.7980	0.2889	0.4426
Switzerland	0.5747	0.8989	0.5398	0.7011
Spain	0.2905	0.7678	0.2670	0.4215
Germany	0.3289	0.9282	0.3208	0.4857
Great Britain	0.2901	0.7079	0.2591	0.4116

Table 5.1: Evaluation metrics for the average country, Switzerland, Spain, Germany and Great Britain.

Even though Germany and Great Britain are at the forefront of open-source geographical data, we cannot conclude that their power line data is complete. The validation data incompleteness implications are complicated to quantify. The uncertainty could be avoided by re-running simulations on guaranteed complete data sets. The entire ground truth data set is visualized at www.openinframap.com[39]. The study failed to obtain a validation data set suitable for the resolution of nighttime lights.

5.3 Method Limitations

The methodology requires high-quality data in multiple aspects. Primarily, the method depends on night lights being an accurate proxy for electricity consumption. Further, seven data layers were implemented either as masks or costs. The method relies on both the quality of the data and the correctness of the logical implementation. The target raster, Dijkstra's limitations, and cost layers are discussed to understand the results better.

5.3.1 Target Raster

The target raster is a binary representation of all pixels assumed connected to the distribution grid. Dijkstra's connects all target points and does not stop until all points have been linked together. This feature implies that the quality of predictions by Dijkstra is dependent on the input data set quality. There are two main challenges concerning the quality of target points. The primary challenge is that pixels are classified based on radiance intensity and not radiance source. Several preprocessing steps were implemented to limit the impact of natural lights, but there is no guarantee of removing all-natural sources. A secondary challenge arises when predicting power lines in developing countries. A report presented in

2018 by Dalberg Advisors and Lighting Global suggested that the current situation and future needs depend on off-grid solar power systems[53]. Dijkstra's algorithm assumes a single grid in each country and connects off-grid settlements to the main grid.

Possible solutions to the off-grid challenge could be to remove target points through preprocessing or avoid connecting them to the main grid. The first solution would require an extensive data set of off-grid settlements and identify unique attributes of these. The second approach could use a maximum cost for cut-off. This approach assumes the off-grid systems are in regions far away from the central grid, and the aggregated cost of connecting a target to the grid would be considered too high.

5.3.2 Cost Layers

The main concern with new cost layers was the lacking impact they had on the results. The cost layers marginally improved results, and extreme scalars had less impact than we expected. To illustrate, Table 5.2 shows that the difference in evaluation metrics between two extreme scalar configurations is almost non-existent.

	Precision	**Recall**	**IoU**	**F1**
$PA_{cost=0} + Slope_{cost=1}$	0.327	0.813	0.291	0.445
$PA_{cost=100} + Slope_{cost=100}$	0.328	0.806	0.290	0.444
Differences	0.001	0.007	0.001	0.001

Table 5.2: Impact of scalars

The marginal differences suggest that new cost layers have little impact on results, and significant methodology improvements are likely made elsewhere. The cost configuration fine-tuning is more impactful after the fundamentals of the process are in place, such as a proper ground-truth dataset.

5.3.3 Validation

The incompleteness of the ground truth data set remains the primary source of uncertainty in this thesis, but multiple other factors could significantly impact results. Power producers expressed that an overview of the local power infrastructure was crucial in the early screening phase of the site selection processes.

However, the expressed importance was not quantified in terms of a minimum resolution or accuracy. Their requirements remained on an abstract level, which complicated the assessment of evaluation metrics. Modifiable validation parameters are, for instance, spatial resolution and ground truth buffer size. Firstly, the prediction and ground truth resolutions were set to be the same as the primary data source. A decreased resolution could improve metrics [36] but was not applied in this thesis. Alternatively, we could have further increased the ground truth buffer. The buffer of 1 km was arbitrary and implied that predictions were classified as correct when within a kilometre from the actual grid. While logical arguments back the resolution and buffer, they are not linked to industry-specific requirements.

Chapter 6

Conclusion

This thesis inferred power lines based on nighttime lights using a modified version of Dijkstra's algorithm. The goal was to create a complete and reliable electricity infrastructure data set in Sub-Saharan Africa by improving the state of the art methods and validating predictions in a wide range of countries. This data set would facilitate the initiation of new energy projects in the region. The results showed promise in countries where electricity consumption strictly followed the limitations imposed by cost world, such as the rough terrain in Switzerland. However, when left with multiple suitable choices for path expansion, Dijkstra's algorithm could not accurately predict the power line layout. The performance difference between the average and best-performing countries was approximately 30% measured by IoU and F1 score. The average IoU score of 0.29 fall short of industry standards, and predictions were neither reliable nor accurate. Regardless of the failure to deliver the thesis' purpose, efforts must be made to increase the access rates in Sub-Saharan Africa.

A crucial finding was that changing the cost configuration and introducing new layers had a minimal impact on the results. The minimum and maximum values for cost weights produced near-identical results, suggesting that improvements are more likely achieved elsewhere. The target raster preprocessing had more impact on results. Successful masks, such as the population mask, removed more noise than valid targets and improved the baseline model. Overall imprecise results could also suggest that the methodology is not suited for its purpose. Energy transmission lines seldom have a single correct layout. Multiple paths might be suitable, and the deciding factor can be an experts opinion or parameters not accounted for in this implementation. Pathfinding for planning purposes and

documentation purposes differ significantly in requirements. The vast amount of relevant literature on planning versus documentation substantiate arguments of documentation being an overly ambitious application.

Future Work Qualitative analysis identified multiple shortcomings of the implemented methodology. The study's main drawbacks were the incompleteness of the validation data set and the impact zero-cost transmission lines had on the predicted distribution grid. First, a more complete data set would have significantly decreased the uncertainty of predictions. Current evaluation metrics offer little explanatory value, and persistently low precision values can be attributed to the systematic error. Further, we observed the undesired branching effects transmission lines had when implemented at zero cost. We suggest a regional approach for distribution grid inference. Inference at a regional scale would not include transmission lines as part of the cost world, as transmission lines transport power across regions. We suggest a combination of the Voronoi diagram substation approach proposed in [11] and what is implemented in this paper for further study. The approach uses substations with high to medium voltage transformers as start nodes and night lights as targets. The method should, at a minimum, solve the challenges caused by transmission lines.

New layers could also easily be implemented in the solution. The urban areas filter significantly increased precision but removed too many valid targets resulting in a decrease in overall metrics. A similar mask could be based on buildings. For a pixel to qualify as an urban area, the majority needs to be a built-up area. Implementing a new mask filtering on building presence would suggest a more lenient requirement than urban areas. The desired result would be a similar increase in precision and less decline in the recall score.

Nonetheless, it is essential to remember that even substantial improvements in evaluation metrics will not guarantee usable results. We are uncertain if the combination of nighttime lights and Dijkstra's algorithm will ever be capable of accurately documenting power lines.

Bibliography

[1] J. V. Henderson, A. Storeygard and D. N. Weil, 'Measuring economic growth from outer space', 2012, ISSN: 00028282. DOI: 10.1257/aer.102.2.994.

[2] D. Gurara, V. Klyuev, N. Mwase and A. F. Presbitero, 'Trends and challenges in infrastructure investment in developing countries', *International Development Policy*, vol. 10.1, 2018. DOI: https://doi.org/10.4000/poldev.2802. [Online]. Available: http://journals.openedition.org/poldev/2802.

[3] M. Du, L. Wang, S. Zou and C. Shi, 'Modeling the census tract level housing vacancy rate with the Jilin1-03 satellite and other geospatial data', *Remote Sensing*, vol. 10, no. 12, 2018, ISSN: 20724292. DOI: 10.3390/rs10121920.

[4] Z. Zhu, Y. Zhou, K. C. Seto, E. C. Stokes, C. Deng, S. T. Pickett and H. Taubenböck, 'Understanding an urbanizing planet: Strategic directions for remote sensing', *Remote Sensing of Environment*, vol. 228, pp. 164–182, 2019, ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2019.04.020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425719301658.

[5] F. Bickenbach, E. Bode, P. Nunnenkamp and M. Söder, 'Night lights and regional GDP', *Review of World Economics*, vol. 152, no. 2, pp. 425–447, May 2016, ISSN: 16102886. DOI: 10.1007/s10290-016-0246-0.

[6] X. Li, S. Liu, M. Jendryke, D. Li and C. Wu, 'Night-time light dynamics during the Iraqi CivilWar', *Remote Sensing*, vol. 10, no. 6, pp. 1–19, 2018, ISSN: 20724292. DOI: 10.3390/rs10060858.

[7] X. Zhao, B. Yu, Y. Liu, S. Yao, T. Lian, L. Chen, C. Yang, Z. Chen and J. Wu, 'NPP-VIIRS DNB daily data in natural disaster assessment: Evidence from selected case studies', *Remote Sensing*, vol. 10, no. 10, pp. 1–25, 2018, ISSN: 20724292. DOI: 10.3390/rs10101526.

[8] C. N. Doll and S. Pachauri, 'Estimating rural populations without access to electricity in developing countries through night-time light satellite imagery', *Energy Policy*, vol. 38, no. 10, pp. 5661–5670, Oct. 2010, ISSN: 03014215. DOI: 10.1016/j.enpol.2010.05.014.

[9] E. Dugoua, R. Kennedy and J. Urpelainen, 'Satellite data for the social sciences: Measuring rural electrification with night-time lights', *International Journal of Remote Sensing*, vol. 39, no. 9, pp. 2690–2701, May 2018, ISSN: 13665901. DOI: 10.1080/01431161.2017.1420936. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01431161.2017.1420936.

[10] G. Falchetta, S. Pachauri, S. Parkinson and E. Byers, 'A high-resolution gridded dataset to assess electrification in sub-Saharan Africa', *Scientific Data*, vol. 6, no. 1, pp. 1–9, Dec. 2019, ISSN: 20524463. DOI: 10.1038/s41597-019-0122-6. [Online]. Available: https://doi.org/10.1038/s41597-019-0122-6.

[11] J. Rivera, C. Goebel, D. Sardari and H.-A. Jacobsen, 'OpenGridMap: An Open Platform for Inferring Power Grids with Crowdsourced Data', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9424, pp. 179–191, Nov. 2015. DOI: 10.1007/978-3-319-25876-8_15. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-25876-8_15.

[12] N. Levin, C. C. Kyba, Q. Zhang, A. Sánchez de Miguel, M. O. Román, X. Li, B. A. Portnov, A. L. Molthan, A. Jechow, S. D. Miller, Z. Wang, R. M. Shrestha and C. D. Elvidge, 'Remote sensing of night lights: A review and an outlook for the future', *Remote Sensing of Environment*, vol. 237, no. December 2019, 2020, ISSN: 00344257. DOI: 10.1016/j.rse.2019.111443.

[13] Y. Katz and N. Levin, 'Quantifying urban light pollution — a comparison between field measurements and eros-b imagery', *Remote Sensing of Environment*, vol. 177, pp. 65–77, 2016, ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2016.02.017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425716300451.

[14] N. Levin and S. Phinn, 'Illuminating the capabilities of landsat 8 for mapping night lights', *Remote Sensing of Environment*, vol. 182, pp. 27–38, 2016, ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2016.04.021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425716301857.

[15] W. Univeristy. (2021). 'Portal for downloading night light imagery captured by luojia1-01', [Online]. Available: http://59.175.109.173:8888/app/login.html (visited on 15th Aug. 2021).

[16] C. D. Elvidge, K. Baugh, M. Zhizhin, F. C. Hsu and T. Ghosh, 'VIIRS night-time lights', *International Journal of Remote Sensing*, 2017, ISSN: 13665901. DOI: 10.1080/01431161.2017.1342050.

[17] N. A. Kramarova, E. R. Nash, P. A. Newman, P. K. Bhartia, R. D. McPeters, D. F. Rault, C. J. Seftor, P. Q. Xu and G. J. Labow, 'Measuring the Antarctic ozone hole with the new ozone mapping and profiler suite (OMPS)', *Atmospheric Chemistry and Physics*, vol. 14, no. 5, pp. 2353–2361, Mar. 2014, ISSN: 16807324. DOI: 10.5194/acp-14-2353-2014.

[18] N. Levin, C. C. Kyba and Q. Zhang, 'Remote sensing of night lights-beyond DMSP', *Remote Sensing*, vol. 11, no. 12, pp. 1–7, 2019, ISSN: 20724292. DOI: 10.3390/rs11121472.

[19] M. O. Román, Z. Wang, Q. Sun, V. Kalb, S. D. Miller, A. Molthan, L. Schultz, J. Bell, E. C. Stokes, B. Pandey, K. C. Seto, D. Hall, T. Oda, R. E. Wolfe, G. Lin, N. Golpayegani, S. Devadiga, C. Davidson, S. Sarkar, C. Praderas, J. Schmaltz, R. Boller, J. Stevens, O. M. Ramos González, E. Padilla, J. Alonso, Y. Detrés, R. Armstrong, I. Miranda, Y. Conte, N. Marrero, K. MacManus, T. Esch and E. J. Masuoka, 'NASA's Black Marble nighttime lights product suite', *Remote Sensing of Environment*, vol. 210, pp. 113–143, Jun. 2018, ISSN: 00344257. DOI: 10.1016/j.rse.2018.03.017.

[20] S. D. Miller, W. Straka, S. P. Mills, C. D. Elvidge, T. F. Lee, J. Solbrig, A. Walther, A. K. Heidinger and S. C. Weiss, 'Illuminating the capabilities of the suomi national polar-orbiting partnership (npp) visible infrared imaging radiometer suite (viirs) day/night band', *Remote Sensing*, vol. 5, no. 12, pp. 6717–6766, 2013, ISSN: 2072-4292. DOI: 10.3390/rs5126717. [Online]. Available: https://www.mdpi.com/2072-4292/5/12/6717.

[21] C. Ruan, J. Luo and Y. Wu, 'Map navigation system based on optimal Dijkstra algorithm', in *CCIS 2014 - Proceedings of 2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems*, Institute of Electrical and Electronics Engineers Inc., 2014, pp. 559–564, ISBN: 9781479947201. DOI: 10.1109/CCIS.2014.7175798.

[22] B. Musznicki, M. Tomczak and P. Zwierzykowski, 'Dijkstra-based localized multicast routing in wireless sensor networks', in *Proceedings of the 2012 8th International Symposium on Communication Systems, Networks and Digital Signal Processing, CSNDSP 2012*, 2012, ISBN: 9781457714733. DOI: 10.1109/CSNDSP.2012.6292692.

[23] E. W. Dijkstra, 'A Note on Two Problems in Connexion with Graphs', pp. 269–296, 1959. DOI: 10.1007/BF01386390.

[24] A. L. Dimitry Gershenson Brandon Rohrer. (2021). 'A new predictive model for more accurate electrical grid mapping', [Online]. Available: https://engineering.fb.com/2019/01/25/connectivity/electrical-grid-mapping/ (visited on 19th Aug. 2021).

[25] Y. Zinchenko, H. Song and W. Rosehart, 'Optimal Transmission Network Topology for Resilient Power Supply', *Lecture Notes in Business Information Processing*, vol. 262, pp. 138–150, Jun. 2016. DOI: 10.1007/978-3-319-73758-4_10. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-73758-4_10.

[26] E. Garcia-Garrido, L. A. Fernandez-Jimenez, M. Mendoza-Villena, P. M. Lara-Santillan, P. J. Zorzano-Santamaria, E. Zorzano-Alba and A. Falces, 'Electric power distribution planning tool based on geographic information systems and evolutionary algorithms', in *2017 European Conference on Electrical Engineering and Computer Science (EECS)*, 2017, pp. 396–401. DOI: 10.1109/EECS.2017.79.

[27] C. Monteiro, I. J. Ramírez-Rosado, V. Miranda, P. J. Zorzano-Santamaría, E. García-Garrido and L. A. Fernández-Jiménez, 'GIS spatial analysis applied to electric line routing optimization', *IEEE Transactions on Power Delivery*, vol. 20, no. 2 I, pp. 934–942, Apr. 2005. DOI: 10.1109/TPWRD.2004.839724.

[28] C. Monteiro, V. Miranda, I. Ramirez-Rosado, P. Zorzano-Santamaria, E. Garcia-Garrido and L. Fernandez-Jimenez, 'Compromise seeking for power line path selection based on economic and environmental corridors', *IEEE Transactions on Power Systems*, vol. 20, no. 3, pp. 1422–1430, 2005. DOI: 10.1109/TPWRS.2005.852149.

[29] S. Demircan, M. Aydin and S. S. Durduran, 'Finding optimum route of electrical energy transmission line using multi-criteria with Q-learning', *Expert Systems with Applications*, vol. 38, no. 4, pp. 3477–3482, Apr. 2011, ISSN: 0957-4174. DOI: 10.1016/J.ESWA.2010.08.135.

[30] V. Yildirim and R. Nisanci, 'Developing a geospatial model for power transmission line routing in turkey', May 2010.

[31] J. Rivera, J. Leimhofer and H.-A. Jacobsen, 'Opengridmap: Towards automatic power grid simulation model generation from crowdsourced data', *Computer Science - Research and Development*, vol. 32, no. 1, pp. 13–23, Mar. 2017, ISSN: 1865-2042. DOI: 10.1007/s00450-016-0317-4. [Online]. Available: https://doi.org/10.1007/s00450-016-0317-4.

[32] D. Seed. (2018). 'Machine learning for high resolution high voltage grid mapping: Pilot project for nigeria, zambia, pakistan.', [Online]. Available: https://development-data-hub-s3-public.s3.amazonaws.com/ddhfiles/148149/report_machine-learning-for-hv-grid-mapping_wbg-esmap_june2018_0.pdf (visited on 10th Aug. 2021).

[33] C. L. Kuo and M. H. Tsai, 'Road characteristics detection based on joint convolutional neural networks with adaptive squares', *ISPRS International Journal of Geo-Information*, vol. 10, no. 6, Jun. 2021. DOI: 10.3390/IJGI10060377.

[34] M. Mokhtarzade and M. J. Zoej, 'Road detection from high-resolution satellite images using artificial neural networks', *International Journal of Applied Earth Observation and Geoinformation*, vol. 9, no. 1, pp. 32–40, Feb. 2007, ISSN: 0303-2434. DOI: 10.1016/J.JAG.2006.05.001.

[35] 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, ISSN: 0924-2716. DOI: 10.1016/J.ISPRSJPRS.2007.05.011.

[36] C. Arderne, C. Zorn, C. Nicolas and E. E. Koks, 'Predictive mapping of the global power system using open data', *Scientific Data*, vol. 7, no. 1, pp. 1–12, Dec. 2020, ISSN: 20524463. DOI: 10.1038/s41597-019-0347-4. [Online]. Available: www.nature.com/scientificdata.

[37] M. E. Mathiesen, 'Electrical Grid Prediction in sub-Saharan Africa', 2021.

[38] Statsforvalteren. (2021). 'Hva er lovlig og hva er forbudt i en nasjonalpark', [Online]. Available: https://www.statsforvalteren.no/nb/Nordland/Miljo-og-klima/Verneomrader/Verneplanprosesser/Lofotodden-nasjonalpark/Hva-er-lovlig-og-hva-er-forbudt-i-en-nasjonalpark/ (visited on 13th Aug. 2021).

[39] R. Garret. (2021). 'Openinframap', [Online]. Available: www.openinframap.com (visited on 19th Aug. 2021).

[40] C. Barrington-Leigh and A. Millard-Ball, 'The world's user-generated road map is more than 80% complete', *PLoS ONE*, vol. 12, no. 8, e0180698, Aug. 2017, ISSN: 19326203. DOI: 10.1371/journal.pone.0180698. [Online]. Available: https://doi.org/10.1371/journal.pone.0180698.

[41] N. Earth. (2021). '10 m cultural vectors dataset', [Online]. Available: https://www.naturalearthdata.com/downloads/10m-cultural-vectors/ (visited on 12th May 2021).

[42] E. O. Group. (2021). 'Night lights dataset', [Online]. Available: https://eogdata.mines.edu/products/vnl/ (visited on 30th Apr. 2021).

[43] O. Foundation, *Licence — openstreetmap foundation,* [Online; accessed 19-august-2021], 2021. [Online]. Available: https://wiki.osmfoundation.org/w/index.php?title=Licence&oldid=8605.

[44] Geofabrik. (2021). 'Geofabrik openstreetmap download', [Online]. Available: https://download.geofabrik.de/europe.html (visited on 30th May 2021).

[45] OpenStreetMap. (2021). 'Osmfilter', [Online]. Available: https://wiki.openstreetmap.org/wiki/Osmfilter (visited on 15th May 2021).

[46] Rasterio. (2021). 'Rasterio resample', [Online]. Available: https://rasterio.readthedocs.io/en/latest/api/rasterio.features.html (visited on 15th May 2021).

[47] SRTM. (2021). 'Srtm digital elevation model dataset', [Online]. Available: https://srtm.csi.cgiar.org/srtmdata/ (visited on 19th May 2021).

[48] P. Planet. (2021). 'World database protected areas dataset', [Online]. Available: https://www.protectedplanet.net/en/thematic-areas/wdpa?tab=WDPA (visited on 15th Jun. 2021).

[49] GHSL. (2021). 'Global human settlement layer population dataset', [Online]. Available: https://ghsl.jrc.ec.europa.eu/index.php%20pop (visited on 22nd May 2021).

[50] ESA and CCI. (2021). 'Land cover map 2019', [Online]. Available: http://maps.elie.ucl.ac.be/CCI/viewer/index.php (visited on 22nd May 2021).

[51] Food and A. O. of the United Nations. (2021). 'Land cover classification system (lccs)', [Online]. Available: http://www.fao.org/land-water/land/land-governance/land-resources-planning-toolbox/category/details/en/c/1036361/ (visited on 30th May 2021).

[52] T. Economics. (2021). 'Electricity access spain', [Online]. Available: https://tradingeconomics.com/spain/access-to-electricity-percent-of-population-wb-data.html (visited on 16th Aug. 2021).

[53] D. Advisors and L. Global. (2018). 'Off-grid solar market trends report', [Online]. Available: https://www.lightingglobal.org/wp-content/uploads/2018/02/2018_Off_Grid_Solar_Market_Trends_Report_Summary.pdf (visited on 15th Aug. 2021).

Appendix

.1 Code and Results

Algorithm 1 contains pseudocode for the modified version of Dijkstra's algorithm. A Python implementation of the pseudocode can be found at https://github.com/carderne/gridfinder/blob/master/gridfinder/gridfinder.py. All necessary code to re-

Algorithm 1 Modified Version of Dijkstra's Algorithm

```
T = Targets
C = Costs
s = Start
function PATHFINDER(T, C, s)
    distance[s] = 0
    Q = PriorityQueue by distance
    Push S to Q
    while Q is not empty do
        current = Q.pop()
        for next in current.neighbours do
            if n is target then
                zero distances between current and n
            else
                Calculate updated distance
                if next is unvisited then
                    dist_add = distance[current] + C[next]
                    distance[next] = min(dist_add, distance[next])
                    Push next to Q
                else
                    Push next to Q
                end if
            end if
        end for
    end while
    return distance
end function
```

create the experiments are provided in the Github Repository. Approximately 30 different parameter configurations were experimented with, each configuration produced results in 34 countries. The raw result data was considered too much to include in the appendix, but is available in the Github Repository.

.2 Roads

The maps visualizing the cost raster in the two studied countries did not visualize the road networks. Dijkstra's algorithm traveled cheaper along known roads. Costs were differentiated based on road category, categories are displayed in the legend of Map 1 while the associated costs are displayed in the legend of Map 2.

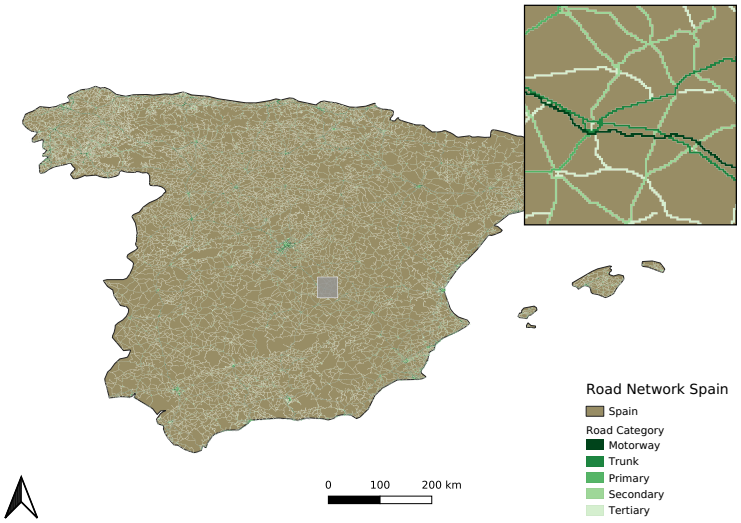


Figure 1: Road network Spain

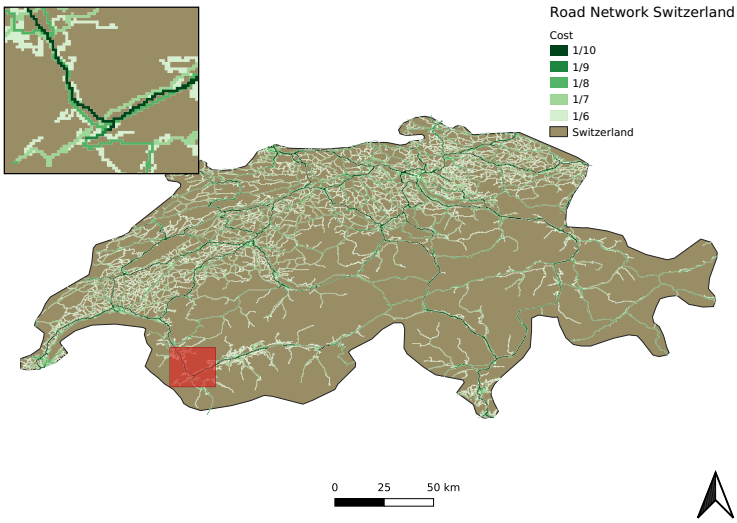
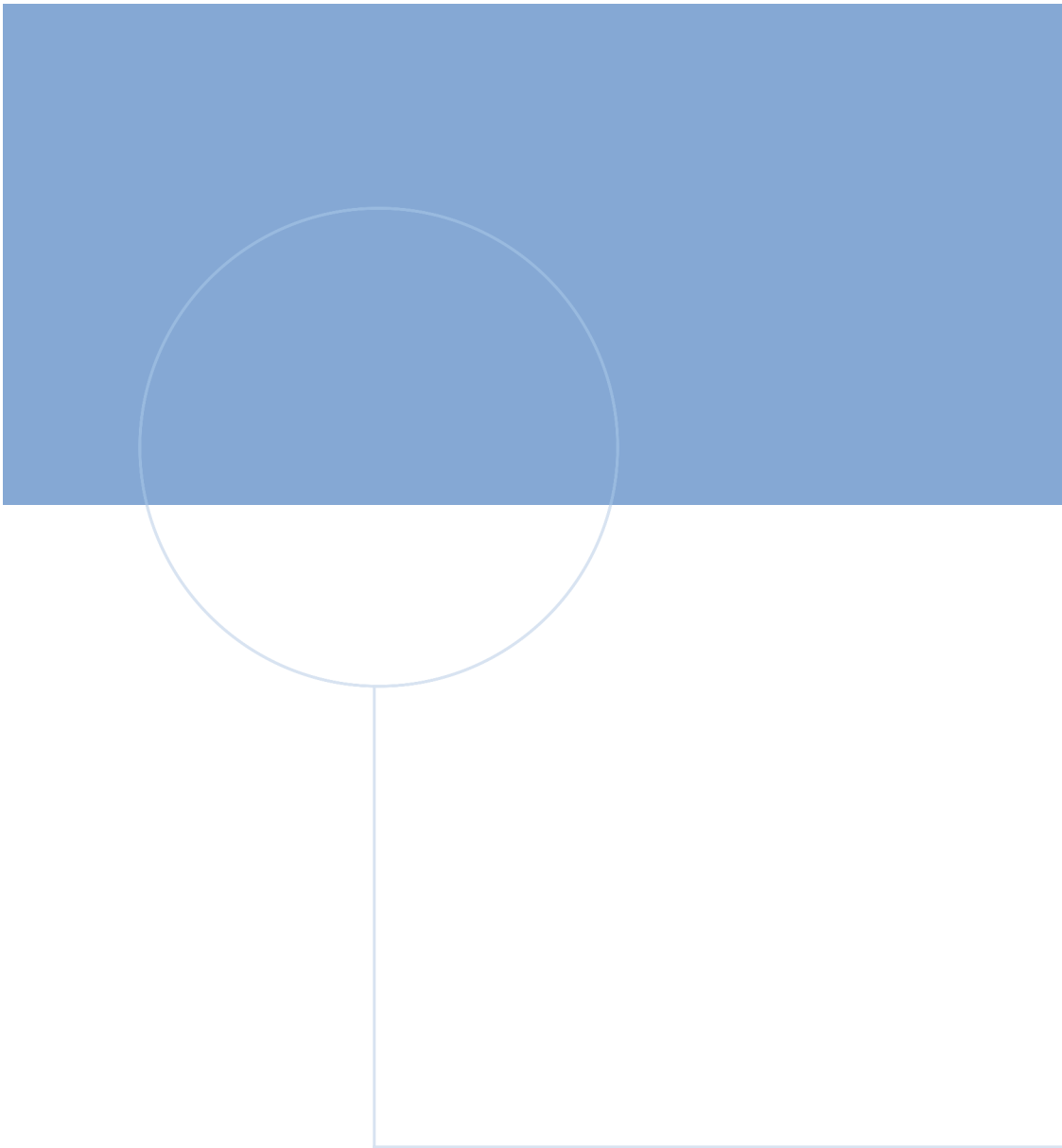


Figure 2: Road network Switzerland



NTNU

Kunnskap for en bedre verden