TÉCNICO LISBOA

Road lighting efficiency improvement through data processing and automated computation

Jakub Starczewski

Thesis to obtain the Master of Science Degree in

Energy Engineering and Management

Supervisors: Prof. Cláudia Martins Antunes
Prof. Sebastian Ernst

Examination Committee

Chairperson: Prof. Duarte de Mesquita e Sousa
Supervisor: Prof. Cláudia Martins Antunes
Member of the Committee: Prof. Francisco Afonso Severino Regateiro

October 2020

Acknowledgments

First and foremost, I would like to thank my supervisor PhD, Sebastian Ernst, for his immeasurable help. I can honestly say that I would not be able to complete this thesis without your guidance. Secondly, I would like to give sincere thanks to my co-supervisor Professor, Cláudia Antunes, for immense number of tips and help that I received from you. I would also like to thank my fiancée Kasia for her relentless support and motivation – Love you! The person who couldn't be missed here is Kasia's sister – Paulisia. Thank you for making our wedding possible! Next, I would like to thank my friend Laurka for her patience and for providing me with all the necessary information, which allowed me to submit the thesis on time. Also, a big thank you to Łukasz and Kozioł for their support and encouragement. You are the best! Last but not least, I would like to acknowledge my parents – Starczi Sr. and Dorota, and my brother – Starczi Jr., who were all a great support for me. You guys are my strength and everyday inspiration.

Abstract

Road lighting design requires extraction of crucial parameters i.e. lighting situations from the data (both road and lighting infrastructure) regarding the area intended for the retrofit. As the commercially available software is not capable of batch processing geospatial data, this process is currently done by hand, which inevitably leads to many simplifications and omissions, especially in large-scale projects. For that reason, an attempt to automatically derive the lighting situations from the official data concerning the area of Washington, D.C has been performed. The process assumed determination of spatial relationships between three sets of data (street lights, street segments, and roads) in order to obtain the luminaires and their corresponding road segments. The procedures, by means of which the best ascription candidate was chosen, were formed by initially reviewing the problematic cases manually and extending the algorithms accordingly, until the satisfactory outcomes were achieved. Therefore, a set of criteria has been determined to minimize the ambiguity of the assignments. As a result, 99% of street lights have been automatically assigned to the road segments. The results suggest that the lighting design performed with the help of the developed solution yields significantly more accurate outcomes in comparison with the other considered approaches. As the tool is versatile, it also allows to automatically estimate the possible power savings for any other area (assuming similar data) without the need for human assistance. This is especially helpful during the strategic planning phase of the lighting retrofit projects.

Keywords: street lighting, lighting design, energy efficiency, data processing, automated computation

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Road lighting design requires extraction of crucial parameters i.e. lighting situations (understood as a road segment with assigned luminaires) from the available data. As the commercially available software is not capable of batch processing, this process is currently done by hand, which leads to many simplifications and omissions, especially for projects covering entire cities. Therefore, in order to automate the process, the design procedure has to begin with determination of spatial relationships between the infrastructure inventory data (e.g. “what provides the light”) and data regarding the illuminated areas (e.g. “what is being lit”).

The primary goal of the thesis is to investigate the possibility of utilizing existing GIS datasets to develop automatic algorithms to handle the tedious task of data integration and data analysis in road and street lighting modernization projects. This will be done by developing a tool to process geospatial data in order to automatically determine spatial relationships between the datasets.

The solution will be based on a set of procedures build upon the examination of the analyzed collections of data, starting from a manual review of questionable results and ending with fully automatic processing. The tool will encompass the following functionality targets:

- Cleaning, conversion, and unification of the data,
- Automatic assignment of segment linestrings to the lit areas,
- Automatic assignment of lamps to the lit areas,
- Evaluation of the results of the automatic process and road segments quality analysis,
- Preparation of the data for photometric optimization.

The structure of the thesis is as follows. The following chapter covers the basic concepts related to road lighting and other approaches that are being applied to the problem. In Chapter 3, the main issue addressed in the thesis is formulated. Moreover, the methodology of the solution development is outlined. Then, in Chapter 4, a case study is thoroughly described and the results are evaluated. Finally, in Chapter 5, conclusions and future considerations are presented.

Chapter 2

Literature Review

2.1 Basic concepts for the domain

Although it is difficult to anticipate the final outcome of anthropogenic activity on climate change, the vast majority of academics from around the globe generally agree that it could be catastrophic. While the phenomenon is a composite of several factors, the use of energy obtained from fossil fuels is perceived as a major contributor. For this reason, renewable energy technologies are being researched and constantly improved. While they significantly help in tackling the issue, improving energy efficiency is often the cheapest and most immediate approach to reduce fossil fuel consumption. In addition to energy and GHG emissions savings, lowered demand provides further benefits to consumers and governments (e.g. reduced energy imports, which implies both cheaper energy and increased energy security). Energy-saving techniques demonstrate great potential for boosting economic growth in sectors that account for a substantial fraction of global electricity consumption. According to the report published in 2013 [1], lighting represents 19% of electricity usage globally; however, the figure regarding the European Union is several percentage points lower at 14%. The inequality in the aforementioned numbers results from the EU's obligation to implement green lighting solutions like solid-state lighting in place of legacy technologies.

Solid-state lighting (SSL) concerns the use of light-emitting diodes (LEDs) or organic light-emitting diodes (OLEDs) as a light source, instead of electric filaments or gas, which are used in incandescent lamps, and high-intensity discharge lamps respectively [2]. Despite the fact that the OLED technology is widely developed nowadays, it is not yet at the performance or cost-competitiveness level of LED [3], thus the main focus of the thesis will be put on the LED-based lighting.

The report of the U.S. Department of Energy [4] revealed that SSL have enormous capabilities of energy savings. It is projected that, by 2025, a transition to the aforementioned technology could save as much as 217 terawatt-hours, which is the predicted generation of electricity from wind turbines and more than ten times that from solar appliances in 2025. The number could be further extended to 300 TWh in

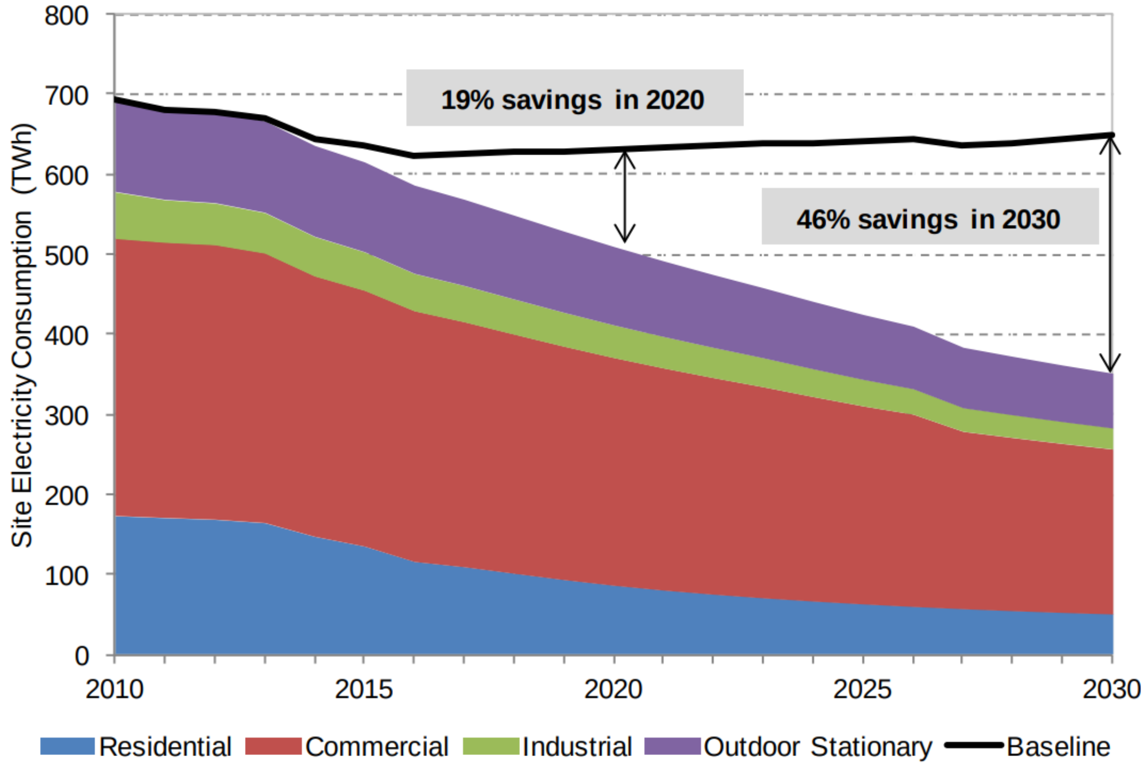

Figure 2.1: Forecasted U.S. Lighting Energy Consumption and Savings, 2010 to 2030 (source: [4])

2030, by means of inevitable growth of LED lighting market share, which would amount to around $30 billion in energy savings assuming current prices of energy. As the energy expenditure of 300 TWh is corresponding to the electrical output of about fifty 1,000 MW power plants annually, these would reduce GHG emissions by 210 million metric tons of carbon dioxide equivalent [4].

What is more, as the illuminance levels of the current lighting installations are usually much higher than required, they contribute to the intensification of the light pollution phenomenon. The anomaly is defined as any negative effect of redundant artificial light during night time. Besides the obvious concern, which is energy waste, light pollution influences several other areas of life. For example, it is responsible for disruption of diverse ecosystems (i.e. bird migration guided by light, restructured predator-prey interactions etc.) [5]. Furthermore, exposure to excessive light during the night adversely affects human health by suppressing the melatonin production and thus, the circadian rhythm, which is the process responsible for regulating our sleep-wake cycle, becomes disorganized [5]. As a result, health problems such as sleeping disorders, increased fatigue, stress, anxiety and frequency of headaches have been associated with the above circumstance. Light pollution has also been linked with the growth of obesity rates and risk of getting several types of cancer [5].

In conclusion, the urgency to reduce the environmental impact of fossil fuel combustion makes a transition to efficient lighting imperative and, therefore, improving its performance has the capacity to not only yield massive economic benefits, but also to become one of the most powerful short-term weapons to combat climate change and diminish the effects of light pollution.

2.1.1 Outdoor lighting

While the outdoor lighting sector is estimated to represent only 15% of total lighting electricity consumption (Fig. 2.1), it is one of the fastest-growing application areas for LED technology. The most important function for outdoor lighting is to enhance the visual conditions and security for drivers and other road users during night hours. The lighting system should operate in a cost-effective way and besides its essential role, it should also take into account both social and environmental aspects.

The reasoning behind the implementation of LED streetlights is straightforward and, from the investor's perspective, the benefits are twofold. First and foremost, LEDs energy consumption is on average halved in comparison with legacy high-intensity discharge (HID) luminaires like e.g. high-pressure sodium. Moreover, their significantly prolonged lifetimes result not only in avoided maintenance, but also less frequent replacements. Although in the past, these advantages had to be confronted with unquestionably greater initial costs for luminaires in this specific technology, nowadays, as the prices for LEDs deteriorated, they are now enough to substantiate the investment in LED streetlights in most cases (Fig. 2.2) [6].

According to the report prepared by Northeast Group in 2014, by 2025 the number of streetlights installed globally is going to add up to 352 million. It has also been estimated that the annual energy expenditure associated with powering them is, depending on energy prices, fluctuating between $23,9–42.5 billion. These numbers speak for themselves, and the message is,that even a small increase in energy efficiency can considerably boost the budget of the cities [7]. While outdoor lighting currently accounts for up to 40% of many municipalities' energy costs, it can be initially cut down by around 40% by application of quality LEDs instead of HID technologies. Light-emitting diodes are characterized by an exceptionally low onset time and photometric solid adaptability (luminous flux distribution shaping) [8], therefore the replacement opens up possibilities for the additional decrease in energy consumption; further savings of 30% can be achieved by utilizing these features. It can be achieved by means of lightning design optimization (optimal selection of each fixture type and its installation settings, which prevents from over-illumination) and introduction of advanced control of the light intensity, based on the input from sensors (dimming level matched to the actual state of surroundings) [9]. Retrofit strategies, however, will be covered in-depth in the subsequent sections. The latest report by Northeast Group (2019), based on the data from the most influential streetlight dealers, forecasts the behaviour of street lighting market from 2019 to 2028. The paper suggests that the LED market is developing rapidly and is predicted to experience steady growth over the next ten years (Fig. 2.2). Total investment in outdoor lighting is expected to reach around $50 billion in the researched time span. It is also mentioned that by the year 2028, due to over 200 million street lights to be installed worldwide, SSL technology will reach a penetration rate of 85%. More and more street lights will be networked and additional sensors will be attached to 2.8 million streetlights, which will not only make use for dimming control purposes, but also for many other smart cities applications.

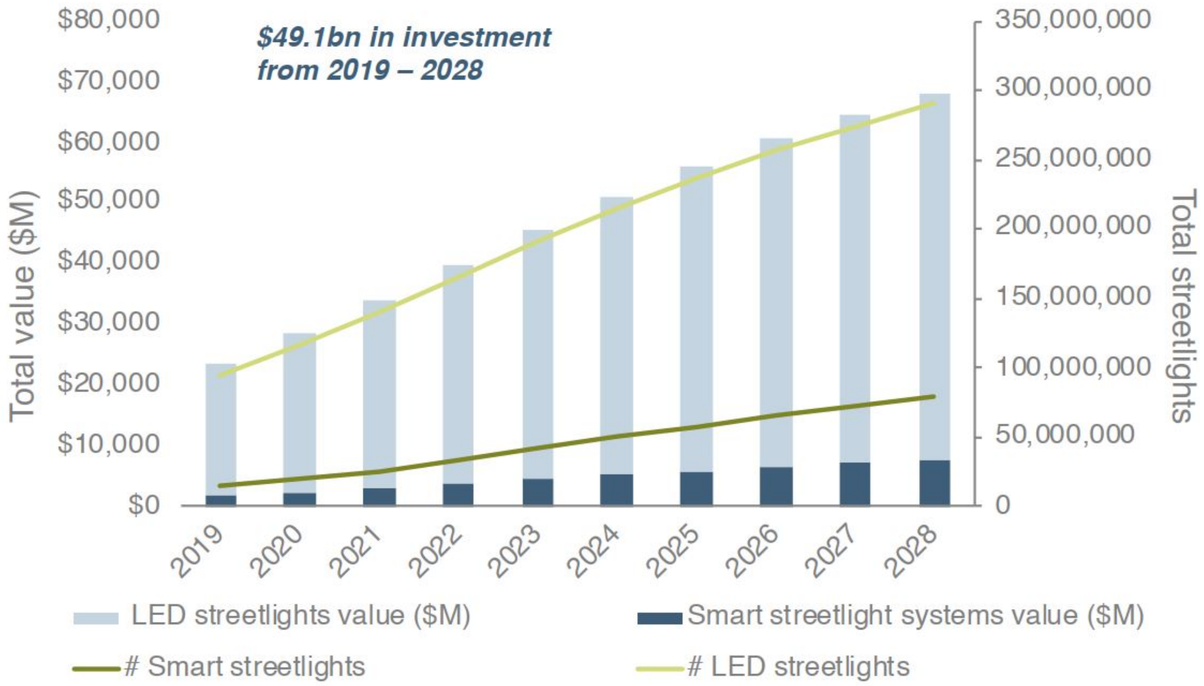

Figure 2.2: Aggregate investment in LED and smart streetlights technologies (source: [6])

Having taken a closer look at these figures it is now clear that LEDs will dominate the street lighting market in the years to come. As the scale of development is enormous, it is no longer the question of “if”, but rather of “when”.

2.1.2 Estimating investment potential based on the data

As stated in the previous subsection, investments in efficient outdoor lighting systems will result in immense energy savings, but the decision to upgrade the lighting infrastructure has to be carefully thought through and be made upon a thorough inference process. Fortunately, the investment potential can be found by collecting and analyzing specific sets of the data described in following subsections [10].

2.1.2.1 Data necessary for lighting class selection process

A lighting class is a group of lighting requirements concerning visual satisfaction of specific road users [11]. The distinction is made on the basis of the type of public space (e.g. roadway, intersection or sidewalk), thus there are several lighting class families:

- M for drivers of motorized vehicles on traffic routes or some residential roads (moderate to high driving speeds allowed),
- C mostly for motorized users on traffic routes, where the conflict areas (e.g. roundabouts, junctions) occur,
- P and HS for pedestrian and cyclists routes,

- other, supplementary groups like SC, where the main objective is to identify a certain entity (areas with inflated crime rates) or EV, where the visibility of vertical surfaces is necessary (toll booths).

The classes are enumerated accordingly to the required lighting level and, generally speaking, the lower the number of the class, the more light has to be supplied for a given area. For instance, the “M” classes range between M1 to M6 with the requirements of the former being the strictest. Moreover, each class has individual performance requirements assigned. The lighting criteria are based on the most relevant photometric parameter. In the case mentioned class M, its crucial lighting purpose is to guide the drivers along the road, therefore the luminance of the carriageway is the basis for the measurement. It is characterized by the average road surface luminance, the overall uniformity of the luminance and the longitudinal uniformity of the luminance. Figure 2.3 demonstrates how luminance requirements vary depending on the assigned lighting class. Apart from luminance, two other groups of photometric properties are distinguished for class M, namely the disability glare and the lighting of surroundings. They are described by the threshold increment and the surround ratio attributes respectively.

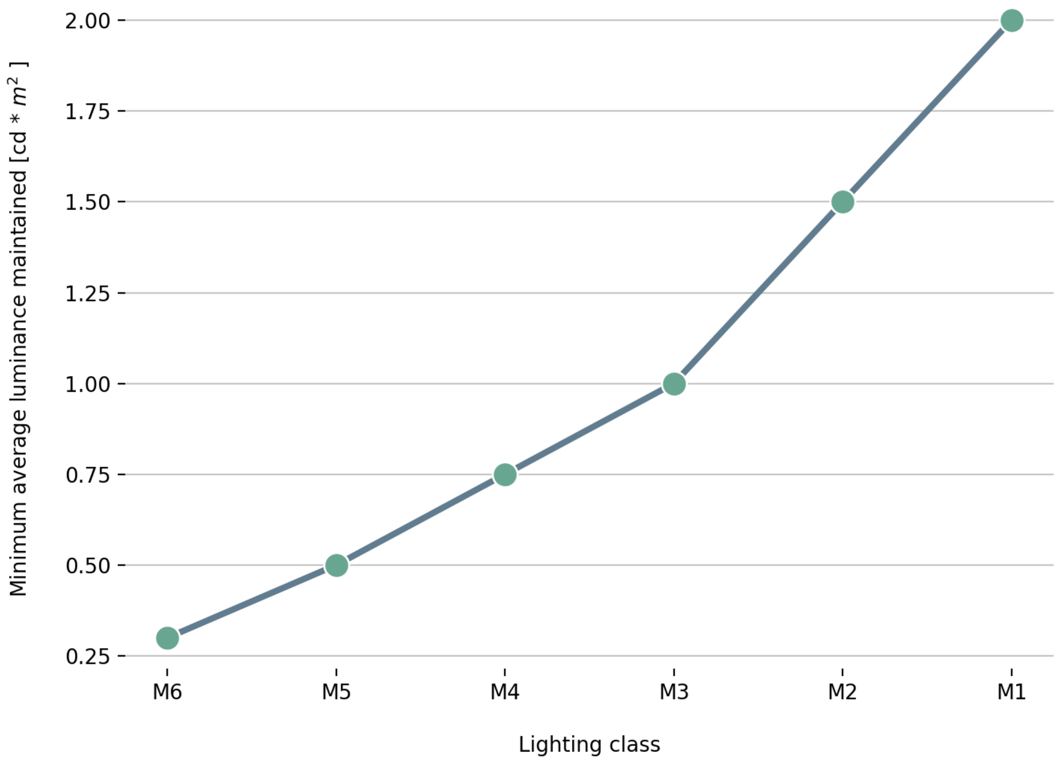

Figure 2.3: Luminance requirements depending on the assigned lighting class

On the other hand, horizontal illuminance is used to define both the C and P classes. This is because the visibility of other road users and obstacles that may appear on the roadway is much more important than navigating the user. All photometric parameters, which are the basis for specific lighting classes, are thoroughly described in the EN 13201-2 norm [11]. Selection of the lighting class is made in compliance with another part of aforementioned EU norm i.e. 13201-1:2014 [12], which serves as a guideline. Again, the criteria vary depending on the type of public space. According to the standard, eight following parameters are essential for carrying out the lighting class selection process in group “M”:

- speed limit,
- traffic volume,
- traffic composition (motorized only, non-motorized, mixed),
- carriageway separation,
- density of junctions,
- presence of parked vehicles,
- intensity of ambient light,
- navigational task difficulty.

2.1.2.2 Infrastructure and road configuration data

Infrastructure data is a precise log of the existing lighting points and their features used for calculations, including their height, number of luminaires, type of the luminaires, photometric solid (distribution of luminous flux), wattage and many more. Road configuration data, on the other hand, is a set of detailed information about the areas being lit including the street layout, its width etc. These datasets contain information which describes a particular location on the globe and therefore are referred to as geospatial data. Geospatial data is generally modelled as rasters or vectors. The former present the data as a pixel grid, which is why it is usually associated with images. Each raster element consists of an attribute value (for example a colour on the topographic map) together with its coordinates. Although collecting raster data is much cheaper most of the time, vector data is much more accurate and, as a result, better suited to represent the topographic shapes [13]. GIS vector data, similarly to the raster equivalent, use x-y coordinate pairs to portray the latitude and longitude of each record in the dataset. These coordinates are the basis for the formation of three main GIS data types, which are used to model objects in the real-world, i.e. points, linestrings and polygons. The first, self-explanatory data type is used to describe a discrete location on the surface of the earth. Moreover, by connecting point sequences, a linestring can be created. Finally, closed linestring features (the same coordinate pair in the beginning and the end of the point sequence) are considered polygons, which model road shapes in our particular case [14].

A coordinate reference system (CRS) is used to define the way in which locations on the three-dimensional earth’s shape are related to the same places, on a projected, 2D plane. The mentioned projection, which is a component of a CRS, defines the mathematical equation, basing on which a 2D map is developed from the 3D Earth’s surface. Lots of projections have been created thus far, however, it is certainly impossible to create a perfect two-dimensional representation of a 3D surface [15]. Coordinate reference systems can be divided into geographic (spherical) coordinate reference systems and projected (isometric) coordinate reference systems. The former uses latitude and longitude values (expressed in decimal degrees) to determine the position of the record on the map (with equator and prime meridian set as the references). Among the geographic coordinate reference systems, the WGS 84 is the most commonly used. Figure 2.4 presents the world map in Web Mercator coordinate system, which is WGS 84 projected into a square. The latter, however, as the name suggests, utilizes projected coordinates i.e. the distance from a set of reference points expressed in linear units (e.g. meters).

Using projected coordinates enables effective spatial calculations such as distance measurement. These coordinates, however, are only applicable to areas of a rather small extent. On the other hand, geographic coordinate systems can be used for analyses covering the entire globe, but the measurements are more complicated. To sum up, it is important to choose a projection and the CRS, which are suitable for the size and location of the examined area or the type of analysis to be performed, in order to maximise the accuracy of the further computations [15].

Figure 2.4: The map of the world in Web Mercator projection

2.1.3 How to make use of the data?

2.1.3.1 Design Phase

The data described in the previous section serves as a basis for the most important part of every street lighting modernization project – the design phase itself. There are several objectives that the street light system designer has to accomplish. First and foremost, outdoor lighting design has to ensure safety and comfort of moving in public spaces, by providing adequate illumination during night-time i.e. by selecting the appropriate fixture models and their parameters. The selection is being made by means of photometric computations. The draft has to comply with current standards while maintaining cost-effectiveness, which can be represented by any quantitative criterion like the reduction of energy consumption or investment expenditure. Finally, the aspect of aesthetics also has to be taken into account [16].

The first task for the designer is to prepare the wire-frame of the project. The proposed luminaires and their photometric properties are then verified by a lighting design software (like DIALux, Relux or Ulysse) and a three-dimensional visualization is being formulated subsequently. Eventually, the final outcome is being evaluated and should it not meet the designer's expectations or the norm's requirements, corrections are being made accordingly. The process then loops back to its initial stages, indicating the trial-and-error character of the described technique [10].

A great number of adjustable parameters together with an anthropocentric nature of the design process implies excessive workload and exposes a crucial issue with the schema presented above, namely, the complexity of calculations and multitude of combinations [17]. While the designer's expertise in the photometric design area may significantly decrease the search space, the data volume makes the identification of both optimal and requirement compliant solutions next to impossible. Luckily, data processing tools capable of solving the described problem are currently being developed [10], therefore the optimal configuration can be easier to obtain.

2.1.3.2 Retrofit scopes

Three feasible retrofit scopes based on lighting design paradigm can be distinguished:

- luminaire replacement,
- lighting design optimization,
- implementation of street lighting control.

Their applicability is dependent on numerous factors like modernization type or financial restrictions etc. [10].

Luminaire replacement

Simple replacement of luminaires is the most obvious modernization model. As stated before, using

LED light sources instead of high-pressure sodium or metal halide luminaires, results in 40% energy usage reduction on average. The process can be executed twofold: either on the basis of lighting design or a conversion chart, which is a document provided by street light dealers used to find the LEDs wattage corresponding to the power of the exact lamp to be replaced [18]. However, even if the identical distribution of luminous flux was assumed, the previous installation could still be over/under illuminated, thus either not fulfilling the legal requirements or reducing the benefits of replacement process alone. Therefore, using a conversion chart is not recommended in procuring the new lighting infrastructure. The replacement model assumes no additional adjustment to the installation parameters (e.g. fixture dimming).

Lighting design optimization

Another retrofit model, relying on the LED technology, assumes reducing the power consumption to the lowest possible level, while at the same time conforming to the requirements of the current lighting standards, by tuning the design variables. The adjustment is concluded by selecting the most optimal environment for each luminaire. The parameters to be modified are dependent on the modernization scope e.g. for a completely new lighting system, they might consist of the pole height, its overhang and setback as well as the photometric solid, together with its mounting angles, dimming and others. However, when a retrofit is considered, only some parameters will be altered, as variables like the position of the pole or its height are most likely unchangeable.

Two methodologies related to design optimization of outdoor lighting systems can be distinguished: the standard and the custom approach (Fig. 2.5). The former imposes simplifications on layouts of examined streets or roads, by assuming the road width to be constant and luminaires to be arranged in equal intervals. This procedure has been used as a default by all widely recognized and available lighting design applications [9]. By making such simplifications, the process itself is significantly shorter and it ensures compliance with the requirements imposed by standards. However, on the other hand, as both the road width and the luminaire spacing are considered as an average or maximum values, this methodology does not yield fully optimized results with regard to electricity consumption [19]. The over-illumination generated by applying the above scheme causes additional energy expenditures in effect.

The latter, commonly named the custom approach, was proposed by Sędziwy in 2015 [9]. It assumes that the lighting situation for which the photometric computations are being performed (i.e. street or bicycle path) is not treated as homogeneous anymore. As a result, neither the luminaire spacing nor road structure is considered constant. The road segment (computation field, which reflects specific lighting situation) is formed by creating a rectangle from the value of luminaire spacing and road width measured from the perspective of each street light along the appropriate part of the road, while the exact geographic coordinates serve as the foundation for positioning of lighting poles. An appropriate lighting class is then selected for each segment, however, one must take into account lighting requirements for neighbouring segments to prevent over-illumination. This is due to the fact that as both luminaires and

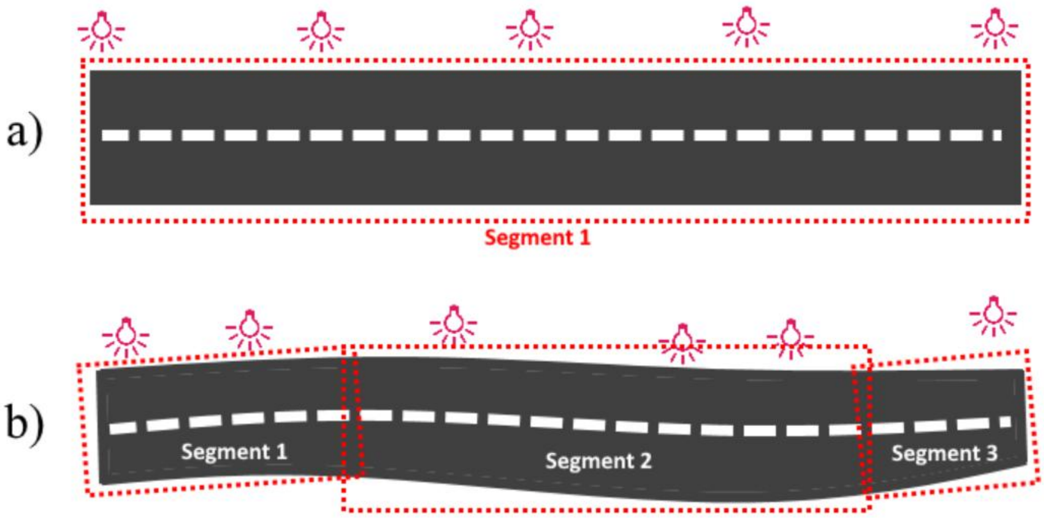

Figure 2.5: Standard(a) and custom(b) approach to photometric computations (source: [18])

roadways are now treated independently, a single lamp can illuminate several road segments. As the proposed methodology imposes examination of numerous segments of the road instead of its simplified representation, the computational complexity of the process is respectively greater, therefore, the process is usually supported by both methods based on artificial intelligence and time-efficient heuristics [8]. On the other hand, the design prepared by means of the above scheme minimizes the surplus lighting. Sędziwy in his article from 2015 [9], demonstrates that by considering the differences from homogeneity, the energy expenditure of lighting installations can be decreased by 15%. It has to be noted, however, that as the industrial practice suggests, averaged calculations are primarily used, while the greater emphasis is put on accurate modelling of road segments (high spacing regularity of street lights).

Control

Finally, lighting efficiency improvement can be achieved by the introduction of control systems. Here, the crucial role is played by the lighting class, which describes the necessary amount of light to be provided for each assessed road segment. Its selection is dependent on specifications described in 3.1; however, certain criteria (like for instance traffic volume) vary throughout the day. Therefore, in accordance with the CEN/TR 13201-1:2014 standard [12], it is possible to alter the lighting class proportionally to fluctuations in vehicle flow intensity, level of ambient light or weather conditions. A collection of lighting classes is, in fact, a set of luminaire configurations with differing luminous flux ratios, which make the generation of energy savings possible. There are of course several ways of implementing this methodology, however, the following strategies are generally considered:

- calendar-based control,
- statistics-based control,
- dynamic/adaptive control.

Calendar control

The first, and by far the simplest solution for outdoor lighting system management is based on the sunset and sunrise times at the given geographic position. Fixtures are then simply switched on and off at these times, while maintaining a single lighting configuration.

Statistics-based control

Although statistics-based lighting control uses the "calendar" approach to turn the power on and off, it also allows for luminous flux changes in compliance with so-called performance schedules. A performance schedule determines the light intensity levels for luminaires at specific hours, thus it is possible to temporarily assign less restrictive lighting class (i.e. M2 instead of M1) to a given area. An example performance schedule is presented in Fig.2.6. As many lighting control systems provide solutions which support this type of control technique, it is nowadays regarded as an industry-standard. Available lighting control systems including CityTouch [20], LightGrid [21] or Owlet [22], are equipped with a user-friendly interface, which allows to remotely determine the performance schedules [23]. They, however, are based on archival statistics of traffic volume, rather than the actual state of the environment.

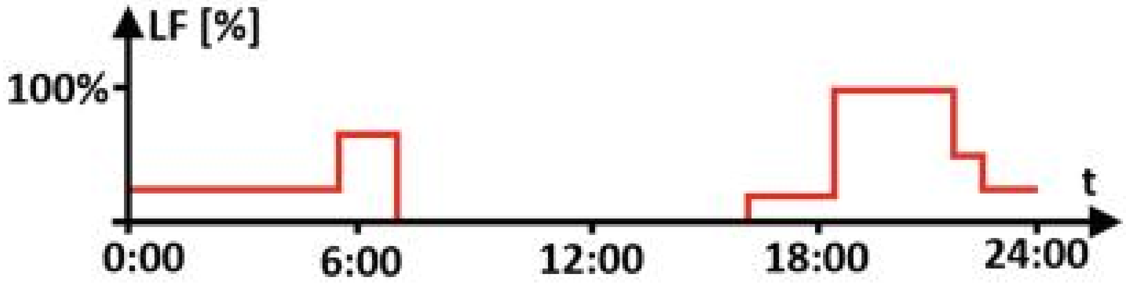

Figure 2.6: The sample performance schedule for a fixture. The horizontal axis corresponds to the hour of the day. The vertical one denotes luminous flux ratio (LF) of a fixture (source: [24])

While the solution is quite popular, it introduces serious issues related to the breach of requirements imposed by lighting standards, due to the unstable nature of traffic volume quantity. Therefore, exceeding the values based on historical and most of the time fragmentary data by the actual conditions is, without a doubt, quite a frequent event [18]. To prevent those situations maximum values are being applied, which results in the non-optimal configuration.

Dynamic control

In the most advanced approach, the control system of the lighting installation is based on the lighting profile concept. It is defined as an aggregate of the environment state (area type together with non-constant parameters involved in lighting class selection described in section 2.1.2.1) and the relevant luminaires' luminous flux ratio. Therefore, as real-time calculations of lighting profiles would introduce another degree of complexity to the computations, in this strategy, a collection of dimming levels for grouped or individual luminaires are being prepared, which are suited for specific conditions of the environment. These conditions are regularly updated by sensors (induction loops, rain or ambient light sensors) and the dimming level of lighting fixtures affected by the changes are shifted accordingly [19] (2.7). Not only does it ensure that a selected lighting class satisfies the requirements imposed by the standards in a given area, but it also optimizes lighting level for particular circumstances.

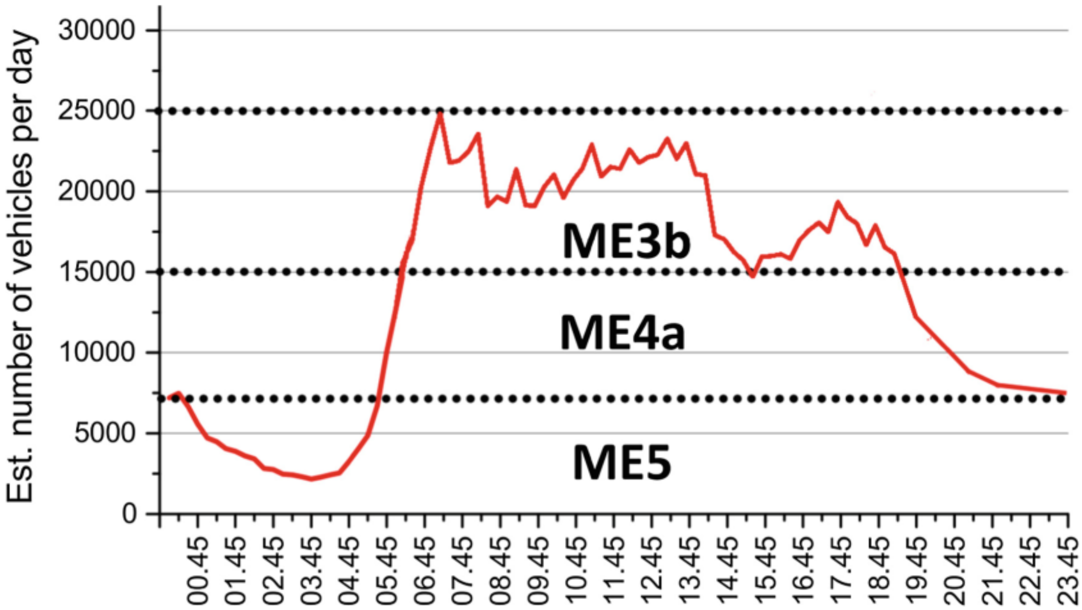

Figure 2.7: The 24 h traffic load measured in 15 min time windows (source: [24])

Although, in comparison to the statistics-based scheme, the application of this methodology yields further savings [23], it has to be pointed out that the design creation for certain areas is an extremely challenging task. A major issue of the dynamic control system is related to the necessity of defining numerous alternatives of each lighting scene corresponding to the actual environment state. If the designer decides to include other important factors like ambient light levels or weather conditions in the computations, the number of options raises, thus the process gets even more complicated [8]. As the design prepared by means of industry-standard software is impossible to be achieved in an acceptable time, it has to be backed by an AI engine capable of reducing the architect's workload. Taking advantage of advanced, photometric design tools mentioned earlier, together with the application of sophisticated heuristics may dramatically limit the number of possibilities to be analyzed. Additionally, further system performance improvements can be obtained by means of parallel computation [17].

Retrofit scopes summary

Even though optimal street lighting system design is an incredibly complex process, the developments in this area will bear fruit in the future. Optimal selection of the luminaires, together with its installation and control parameters lead to a number of extremely beneficial outcomes. First of all, from the investment perspective, reducing energy expenditure affects financial outlays like capital expenditure (CAPEX) and operational expenditures (OPEX). In this context, the former, which is fundamentally the implementation cost of a new system, is decreased by simply installing lower-wattage luminaires, which are significantly cheaper. The latter costs, however, are reduced by better maintenance and optimal illumination management. While generally malfunctioning luminaires are discovered by road users' complaints, streetlight control networks notify the infrastructure owner about any safety risk regarding certain fixture, thus it can be localized and repaired in a timely manner. Moreover, as the installation illuminates only the required area, the minimum energetic outcome is consumed. Not only does it satisfy the economic aspect of the installation's performance, but it also substantially contributes to reducing the environmental impact of fossil fuel combustion (CO_2, sulphur oxides, etc.). Furthermore, an adequate illumination level guarantees safe and convenient movement in public spaces for both drivers and pedestrians and mitigates the light pollution phenomenon, which has an adverse influence on, among others, human health and numerous ecosystems.

2.1.4 Data Collection

The success of street lighting modernization projects relies on access to precise geospatial data and its utilization. The collection process of the aforementioned data is, however, burdened with several issues like high cost, long duration and most importantly unsatisfactory accuracy and completeness of the data [25]. The selection of a suitable collection method is also dependent on their performance against the listed matters. Two types of relevant data to be collected are distinguished:

- infrastructure and road layout,
- road traffic.

2.1.4.1 Infrastructure and road layout inventory

Although there is a number of available collection techniques concerning the infrastructure and road layout data, they can be broken down into two major categories: land and air/space-based techniques. Land-based methods consists of field inventory or integrated GPS/GIS mapping systems, which are currently the most popular choices. Besides them, a photo/video log and mobile LiDAR (Light Detection and Ranging) methods are recognized, with the latter being a relatively new type of a mobile mapping system, and the most promising. As it comes to methods which use air or space as a medium for collecting the data, aerial and satellite imagery is being distinguished together with an airborne version of the LiDAR. All of them exhibit distinct advantages and disadvantages, therefore it is crucial to select one best suited for the specific project. Particular features of five, most relevant collection methods has been presented by Jalayer [26], who carried out a comparative study using the multi-criteria analysis process (MCA). MCA is especially useful when a determination of the superiority of one criterion over another is not possible by means of any quantitative valuation. Every considered criterion had a rating of 1 to 5 assigned, with the former being the worst and the latter the best, to relatively compare the performance of a certain method over the others. Additionally, a weighting factor was chosen for each characteristic, which helped to identify the proportionate importance of each criterion (it is determined by consulting the stakeholders). Finally, by summing the scores of each criterion and multiplying it by its weighing factor, a total weighted score is acquired.

Table 2.1: Evaluation Matrix for Inventory Data Collection Methods (source: [26])

	Criteria	GPS Data Logger	Robotic Total Station	GPS Enable Photo/Video Log	Satellite/ Aerial Imagery	Mobile LiDAR	Weighting Factor
Field Data Collection	Equipment Cost	3	2	4	5	1	0.25
	Labor Cost	2	1	4	5	3	0.25
	Data Collection Time	2	1	4	5	3	0.25
	Safety	2	1	4	5	3	1.00
	Data Completeness	3	4	2	1	5	2.00
	Data Quality	3	4	2	1	5	2.00
	Disruption to Traffic	2	1	4	5	3	1.00
Field Data Reduction	Software Cost	5	4	3	2	1	0.25
	Labor Cost	5	3	4	2	1	0.25
	Data Reduction Time	5	3	4	2	1	0.50
	Data Storage Size	5	4	2	3	1	0.25
Total Weighted Score		24	23	23	21	29	

As the results indicate, when data completeness and quality are the first concern, the utilization of mobile LiDAR technique would yield the finest results. Nonetheless, due to the enormous amount of information collected, it requires substantial effort to reduce the data volume in order to distil the desired inventory data. It is also worth mentioning, that making use of mobile LiDAR technology is associated with the highest expenses.

On the other hand, a traditional field survey does not need data reduction, as only the required data is collected. While it is also considerably cheaper and easier to perform, it is definitely more time-consuming. Additionally, it endangers the working crews to the risks related to roadway traffic. Moreover, as the GPS Logger is the most widely used technique in the outdoor lighting sector, it creates a significant, industry-specific problem. In this method, the collector locates an inventory element in the field (pole) and obtains its georeference data. Then, the attributes like road width are ascribed to each record. As it would be problematic to derive precise roadway width at each determined point, an average (for a given road segment) is usually assigned instead and, as a result, makes the implementation of the custom approach to the lighting design impossible.

2.1.4.2 Road traffic data collection

Real-time road traffic data is generally obtained by sensors. Their popularity is mostly owed to the maturity of the technology and precision of the operation, which greatly improves the roadway conditions. Two main sensor types are being distinguished, namely: intrusive and non-intrusive [27]. The former, together with data collectors, are installed on the road surface or just beneath it. Their placement is burdened with additional costs associated with installation and maintenance, but while it is inevitable to interfere with the traffic during these activities, this kind of sensors is characterized by a tremendous accuracy rate. They can be grouped into three categories: piezoelectric sensors, pneumatic tube sensors and inductive loop detectors, with the last being the most popular [28].

The utilization of the latter, however, minimizes the negative effects on roadway users linked with traditional methods. Non-intrusive sensors, which use a variety of techniques such as infrared, ultrasonic, radar or video image detection [27], are also able to effectively detect and record desired data without occasional traffic disruptions. However, not only are the initial implementation expenses higher in comparison to the conventional techniques, they are also vulnerable to the weather conditions (for instance rain, fog or snow).

Recently, a new, mobile technique of gathering traffic information has emerged. Floating car data (FCD) relies on probe vehicles, which serve as traffic sensors located by means of GPS or mobile phones. Vehicles provide regular information about their geographic position, as well as their driving speed. Then, basing on the collected data, valuable insights like traffic status are being derived and sent back to the roadway users. FCD is not usually used independently to predict the traffic intensity, however, Leduc [29] described it as a great complementary technique for the existing methods and emphasized the benefits resulting from combining traditional and mobile road measurements.

Last, but not least, data can be obtained from mapping services such as OpenStreetMap [30], which is a free, open-source world map. OSM has been collaboratively created by people all over the world (over 2 million registered users), mostly by means of portable GPS devices and aerial imagery. It is an excellent spatial information source; nonetheless, the quality of the data, as well as its coverage fluctuates, due to the voluntary manner of the creation of datasets. The data is generally more precise and available in highly developed countries or commonly frequented areas.

2.1.4.3 Data collection summary

While sensors gather an enormous amount of useful data, the most relevant and important characteristic in the current context is the traffic intensity, which is commonly forecasted on the basis of the historical datasets. This, however, entails certain issues like analysis based on scarce data (resulting from a deficient amount of available sensors). While municipalities are, to a greater or lesser extent, equipped with traffic flow measurement devices in conflict areas such as junctions, other places are usually left without any surveillance at all. Nevertheless, this case still does not justify the installation of additional sensors in these areas, due to the financial aspect of the process. Other than that, the traffic volume is highly dependable on irregular roadway circumstances and therefore, the predicted value might not be coherent with the actual situation on the road. To overcome these problems, new traffic flow prediction algorithms, which base on various AI and statistical tools are being researched [31]. Moreover, due to the developments in the road traffic data collection area, the desired pieces of information are available from the variety of sources and their unification might provide even more descriptive conclusions as a result. This could be implemented by combining the long-established sensor technologies with the novel, mobile data collection techniques and freely available road network data from OpenStreetMap [29].

2.1.5 How can data integration contribute to optimization of lighting design process?

To achieve the desired optimization results, the process of lighting design requires specific data as described in section 2.1.2. GIS data considering lighting infrastructure and road layout can be collected by various methods described in the previous section in different forms including CAD files/drawings, Excel sheets, JSON and XML files or various GIS formats (ESRI Shapefiles, GeoJSON, GeoPackages, etc.). Similarly, data collected from either traditional or mobile sensing devices, complemented by freely available data from world maps like OSM, can support road network analysis and as a result – the selection of lighting classes for the examined segments. Nevertheless, the collected GIS data which is intended for the design procedure requires preprocessing, which involves the extraction of applicable calculation fields in accordance with CEN 13201-3 standard [32]. This is possible only after the identification of relevant areas together with their types (roads, sidewalks etc.) and luminaires associated with them. This, however, is an extremely demanding task for a computing system, specifically for projects covering entire cities [33]. This process is currently done manually, as the industry-standard software is not capable of batch processing. It is not feasible (even with the help of the software) to manually examine all potential lamp settings in a timely manner [33]. Therefore, to handle this problem, the design process has to begin from scratch, which involves merging infrastructure inventory data and data regarding the illuminated areas. The unification process of spatially-described data collected from different sources is called geospatial data integration. An integrated environment, capable of performing data modelling and visualization, is especially important as the decision-making process is usually established on the joint analysis of various datasets [25]. Regarding road lighting optimization, data integration allows the reasoning to be based not only on lamp attributes (spacing, height, etc.) but also on spatial relationships between the luminaires and the areas being illuminated. This feature not only increases the precision of the computations but is also imperative to overcome the aforementioned problems related to the automation of the process. Moreover, by imposing constraints resulting from spatial analysis of the data, the pool of possible states might be significantly decreased. An optimization attempt performed in [19] resulted in reducing the search space (considering both photometry calculations and lighting profiles administered by the control system) 55,296 times, which makes the approach viable.

Fortunately, what is impossible by means of industry-standard software is achievable using Artificial Intelligence and Machine Learning techniques available, among others, within the environment of the Python programming language. General ease of use, together with a variety of tools capable of manipulating and visualizing the data, have made Python one of the most promising alternatives for tasks related to data analysis. There are many open-source Python libraries capable of performing these tasks like pandas [34], scikit-learn [35] or matplotlib [36]. The former is equipped with unique data structures (Series, DataFrames) and methods, which are designated to ease and fasten working with tabular data. Pandas is generally capable of carrying out complete of range actions associated with data scientist workflow in a high-performance and flexible manner, including:

- loading data from a variety of sources,
- cleaning and handling missing data,
- processing data (aggregating, transforming, merging)
- data analysis and modelling,
- preparing the results for visualization

On top of pandas and Shapely [37] (which is intended to work with geometric types) libraries, another open-source project has been developed, namely, GeoPandas [38]. It considerably simplifies handling geospatial data by enhancing the data types utilized by pandas with the ability to operate on geometric objects (GeoDataFrames). Moreover, GeoPandas ensures, that these operations do not need a database for querying and managing geographical information (e.g. PostGIS) as they are enabled within the programming language.

As it comes to geospatial data visualization, it is possible using the already mentioned matplotlib library, however, it is often done by means of dedicated GIS software like QGIS. Formerly know as Quantum GIS, it is an open-source, cross-platform, desktop application, which not only allows data visualization but also supports edition and analysis of the data. It is compatible with a number of raster, vector and database formats, therefore provides a convenient way of geospatial information management [14].

Nevertheless, the collection of geospatial data together with its processing, visualization and analysis are the essence of Geographical Information Systems (GIS). However, in order to improve the economic and environmental aspects of transportation systems, increase safety and comfort of moving in the public spaces, while at the same time conforming with the current standards, an integration of different data sources and technologies (automated tools capable of data processing, data analytics etc.) is required. This, as a result, will significantly boost the decision-making process in the optimization of lighting design and control processes, among others [39].

2.2 Considered approaches to the problem

The proposed approach is based on automatic processing using available (but separate) GIS datasets and tools such as PhoCa [10] to determine the optimal result. This is a fully automatic approach, i.e. it assumes tuning procedures and selecting thresholds based on the visualised data, rather than manual data correction on a point-by-point basis. Due to nuances in the data, this process may need to be guided by the operator, but assumes no manual adjustment of individual data values, however, the premise is to achieve the best possible approximation automatically. There are plenty of scientific papers which propose solutions to the optimisation of parameters of lighting installations, but after the lighting situation (understood as a road segment with assigned luminaires) is manually modelled by the designer. Since it is more of a practical issue, there are no publications which describe automatic approaches to the investigated matter. Therefore, software obstacles, which prevent automation of the process will be briefly characterized in this section.

Typically, each street is assessed by lighting engineers with regard to all parameters which affect lighting, such as pole distribution, pole spacing, road width, pole setbacks, etc. Nevertheless, each lighting situation needs to be configured separately by hand during this process. As tools such as DIALux do not analyse the input setting in any way, they expect the lighting situation to be already determined and described by an engineer (2.8). Programs dealing with photometric calculations do not recognise the geometric object along which the street lights are arranged, what they see, however, is a lighting situation with the poles configured in a specific distance between them. Therefore, as the manual examination of all lamp settings is not feasible in a timely manner, engineers tend to reduce the number of modelled situations, thus reducing precision and risking over and under-lighting.

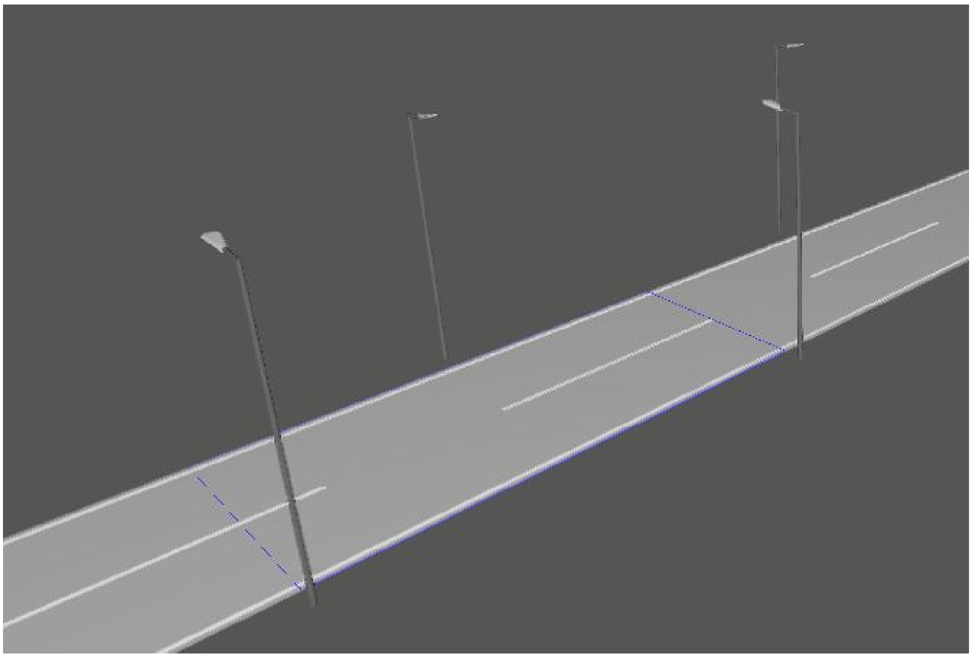

Figure 2.8: A lighting situation modelled in DIALux software

On the other hand, there are also GIS-like systems, which are being used for continuous operational lamp management (e.g. Streetlight Inventory [40]). In these systems, lamps actually appear alongside the roads; however, only the layer of street lights is modelled. It is possible to view and check the parameters of particular lamps in these systems (Fig. 2.9), however, as there is no representation of the roads as such, associating the points with the areas they illuminate is usually impossible. Street lights may even have a street name assigned to them (to ease the management); however, most of the time there is either no reference to the street, or the reference is too general, and thus, it cannot be concluded that the lamps are actually positioned alongside a geometric object. As the map visualizations underneath are based on one of the web mapping services available, the system is not aware of the streets which are there, and therefore it is inaccurate to claim that they light up a certain area.

These systems are excellent as GIS management tools, however, the problem is that they only model

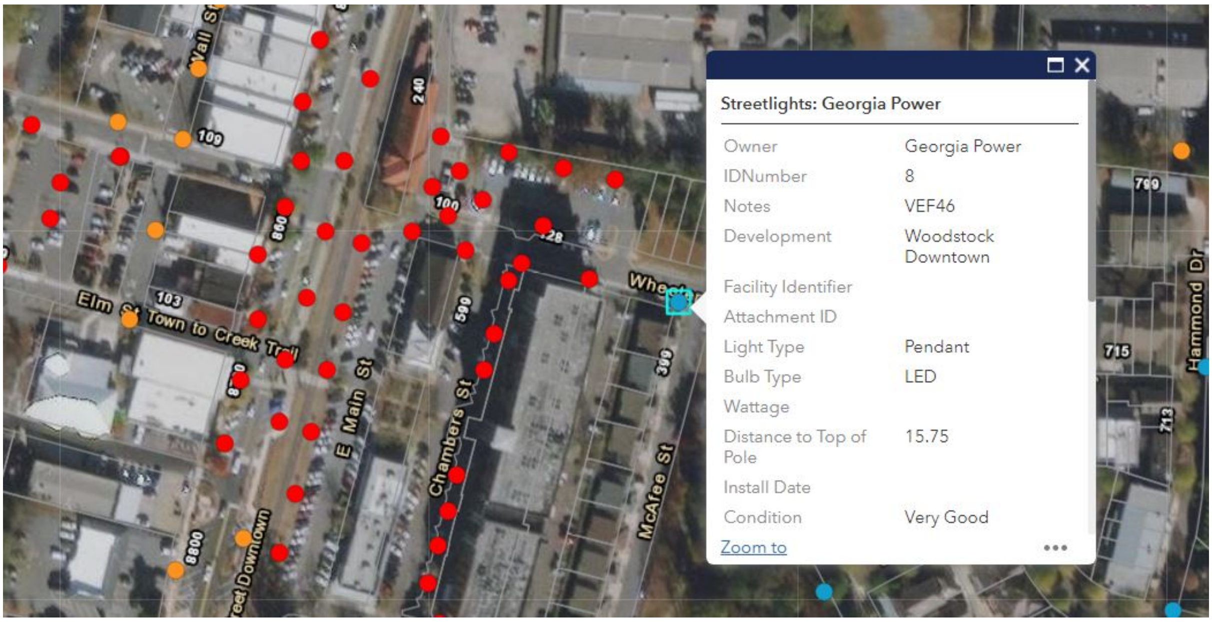

Figure 2.9: Representation of lighting infrastructure (street lights only) in Streetlight Inventory software [40]

the lamp data. While they do it with exceptional precision, they do not understand the situations that they model, i.e. they have neither the possibility of manually modelling the lighting situations as there is no representation of illuminated area nor do they have any possibility of automating the process.

For the above reasons, the processing of the road segments is currently made by a human. The manual approach is understandable in case of deficient and rough data; however, considering the ubiquitous availability of structured and semantically described data, the attempt to automate the process is definitely worth pursuing, especially as the designer's time is generally much more valuable than that of the CPU. This is particularly helpful during the strategic planning phase, when the decision-makers need an accurate estimation of the possible power savings and investment costs, e.g. to prioritise the modernisation in different parts of the city.

Chapter 3

Road lighting efficiency improvement through data processing and automated computation

3.1 Problem statement

3.1.1 Motivation

Power and energy-related projects are often difficult to plan and execute due to insufficient and not integrated data. Smart Cities tend to collect vast amounts of various data, but often the result is not one integrated dataset, but a set of separate databases or files with no relation between them. Replacement of old road and street lighting infrastructures with new, energy-efficient devices is good practical example. As old installations often cannot be relied on with regard to a proper determination of lighting requirements (which depend on road types, their users, traffic intensity, navigation difficulty, and other factors), simply retrofitting old fixtures with similarly-powered LEDs will reduce energy consumption, but won't improve the quality or safety. Moreover, data that is intended for the lighting design procedure requires preprocessing, which involves the extraction of applicable calculation fields in accordance with the CEN 13201-3 standard [32]. This is possible only after the identification of relevant areas together with their types (roads, sidewalks, etc.) and luminaires associated with them. This, however, is an extremely demanding task for a computing system, especially for projects covering entire cities. Commercially available software is not capable of batch processing of the geospatial data, therefore this process is currently done manually. Each lighting situation for a particular area is described by hand, which leads to sub-optimal solution due to introduced simplifications. It is not feasible to examine all potential lamp settings in a timely manner. Therefore, in order to automate the process, the procedure has to begin from determination of spatial relationships between the road and street light infrastructure layers.

3.1.2 Goals

The primary goal of the thesis is to investigate the possibility of utilizing existing GIS datasets to develop automatic algorithms to handle the tedious task of data integration and optimization in road and street lighting modernization projects. This will be done by developing a tool to process geospatial data in order to automatically determine spatial relationships between datasets (e.g. road segment is lit up by a certain fixture). The solution will be based on a set of procedures build upon the examination of the analyzed collections of data, starting from manual review of questionable results and ending with fully automatic processing. The tool will encompass the following functionality targets:

- Cleaning, conversion, and unification of the data,
- Automatic assignment of segment linestrings to the lit area shapes,
- Automatic assignment of lamps to the lit areas,
- Evaluation of the results of the automatic process and road segments quality analysis,
- Preparation of the data for photometric optimization.

The determination of spatial relations of analyzed datasets is non-trivial most of the time and it produces a number of challenges to overcome. Therefore, in order to reach the desired goal a number of problems will be addressed in the thesis, such as: identification of relevant data within the analyzed dataset, differentiation of lamps which light up two areas and ambiguity of the assignment results.

3.2 Solution overview

3.2.1 Selection of tools and libraries

During the planning phase of the thesis, a decision related to the environment of the experimental procedure development had to be made. The selection process was based on ease of use, efficiency and, most importantly, on the availability of the pre-existing packages within a certain programming language with the following characteristics:

- ability to read GIS vector data from different formats such as ESRI Shapefile, GeoPackage or GeoJSON,
- ability to deal with different projections of the data (transform the data from one projection to another),
- ability to clean and manipulate the data,
- ability to perform spatial operations on the geo-referenced data.

Due to the compliance with the mentioned requirements, the Python programming language has been chosen as the primary technology to be used in the approach. As there is a variety of open-source libraries designed to perform geospatial operations, Python emerges as a front-runner in GIS-related applications.

Three main Python packages will be utilized during the development of the proposed solution:

- pandas, which will be used to cleanse and manipulate the data,
- Shapely, which allows to perform numerous geospatial operations on a variety of geometries such as points, linestrings, and polygons,
- GeoPandas, which is built on top of the two preceding packages (extends their functionalities). GeoPandas supports many different input data formats and is equipped with a number of features for spatial analysis and processing, which significantly eases the work with geospatial data. GeoPandas is priceless as it comes to performing bulk operations on geometric attributes or during spatial joins.

3.2.2 Description of data processing steps

3.2.2.1 Cleaning, conversion and unification of the data

The data is provided as datasets and may come in different forms, including Excel sheets, JSON and XML files, AutoCAD drawings, and various GIS formats (ESRI Shapefiles, GeoJSON, GeoPackages, etc.), and therefore, in case of incompatible data collections, the standardization process will be applied to ensure that data is consistent. Datasets will be read by the GeoPandas module and both their spatial data types (Points, LineStrings, Polygons) and coordinate reference systems will be analyzed. Imported data will have a certain spatial reference already assigned, however it is usually much more convenient to convert latitude and longitude values into projected coordinates, which enables efficient calculations, including distance. Precise selection of the projected coordinate system together with its unification within each dataset will be performed to ensure the accuracy of the results, as the inconsistencies in this aspect might become a reason for algorithm misbehavior. Then, after the identification of attributes and their semantics, unnecessary columns will be dropped from the datasets. Moreover, the data collections will be filtered out by the values in specific columns, which will not be useful in further calculations. Finally, a number of cleaning operations will be performed on all datasets, such as:

- Unifying remaining attributes' names and their values (e.g. transforming all string values to lowercase),
- Converting strings to numeric values for future calculations (e.g. light pole height),
- Converting units from imperial/US units to the metric system,
- Replacing various string representations of N/A values with consistent numpy equivalent (e.g. np.nan),

• Saving clean datasets in one common format

3.2.2.2 Automatic assignment of segment linestrings to the lit area shapes

Once the data is consistent and in a calculation-ready form, the process will be initialized by performing a spatial join operation on road polygons and segment linestrings. Spatial join combines the data in a similar fashion as the standard join, nonetheless, the spatial affinity of two features is the relation basis [13]. This relationship between geometric objects is defined by means of Shapely's binary predicates, which return a boolean value, as they are implemented as methods [37]. The following predicates can be used to determine how the spatial join operation will be utilized, by specifying the `op` parameter [38]:

- `within` – this predicate returns `True` if the first object (considering both its interior and boundary) intersects only with the interior of the other object,
- `contains` – the inverse of within predicate,
- `intersects` – the default spatial join setting, which returns `True` if the objects have common points [37]

The last predicate will be used in the spatial join, as all road polygons which intersect with a specific road segment linestring need to be acquired. Therefore, the attributes of the latter will be joined into the former, and as a result, each road polygon would get the attributes of a linestring that intersect with the polygon. This operation would enable the future calculations – the assignment of lamps to the lit areas – to be performed not only basing on the physical distance between the examined layers but also on other significant segment attributes like the street name or its type.

Figure 3.1: The visual representation of a group of road polygons (grey) intersecting with a specific road segment linestring (blue)

Then, a DataFrame that consist of road segments and a set of intersecting road polygons will be created.

The information about the segment's geometry will be merged into the established object and a GeoDataFrame with an appropriate coordinate reference system will be eventually constructed. At this point in the process, a buffer will be performed on the segment linestring, which is one of the most commonly used operations in spatial analytics. By applying a buffer to any of the geometries, a circumferential polygon is created around the original object within a certain distance, as presented in Fig. 3.2 [13].

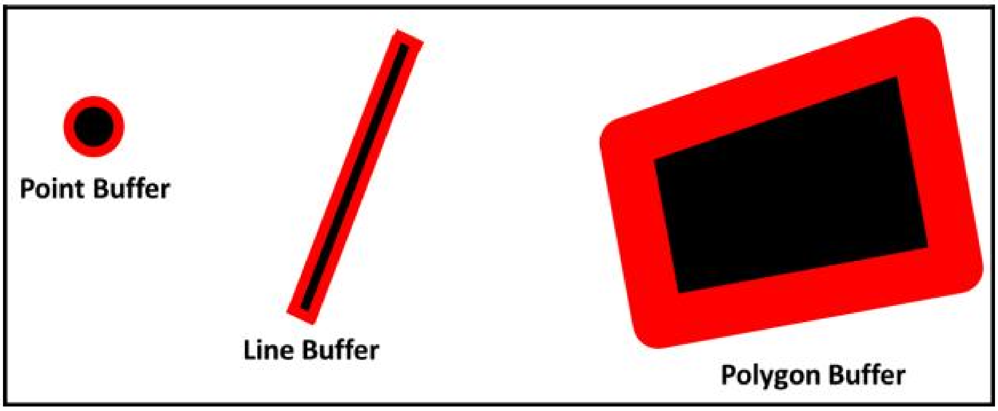

Figure 3.2: The original objects (black) and the polygons resulting from the buffer operation (red) result (source: [13])

Then, in order to derive road segments polygons, road geometries grouped by the linestring they intersect with, will undergo a union operation, which combines multiple shapes into a single object. It has to be noted, however, that from the acquired union polygon, only the parts which intersect with a certain segment's buffer will be taken into consideration for further examination (Fig. 3.3).

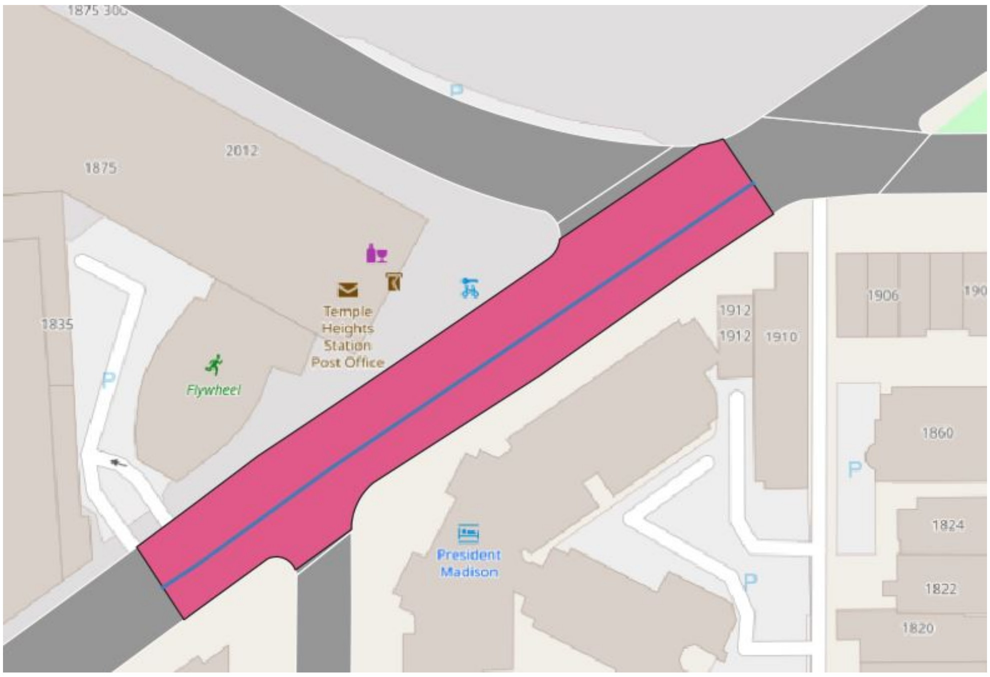

Figure 3.3: The union of polygons (violet) intersecting with a specific segment's linestring buffer (blue)

3.2.2.3 Automatic assignment of lamps to the lit areas

Again, the process will be initialized by performing a spatial join operation, however, this time the computation will consider previously created road segment polygons and street light points. Before the operation, points that refer to the position of light poles will be buffered by a specified distance. By doing so, it will be possible to obtain several candidate roads assigned to each of the lighting points (3.4).

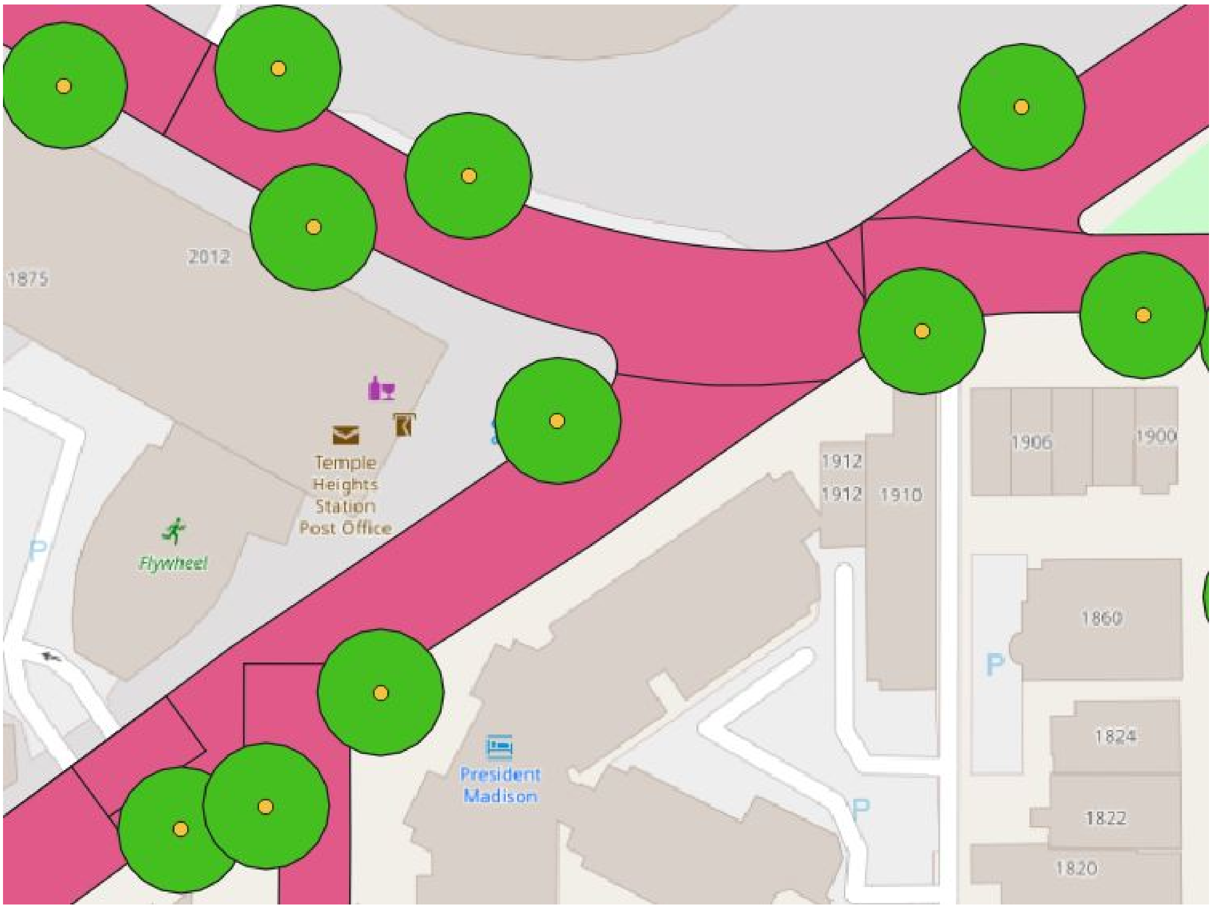

Figure 3.4: Buffered street light points (green) together with the candidating road segments (violet)

The spatially-joined table will be then grouped by the identification number of the light pole and, for each road segment in a group, a number of factors will be calculated, including the physical distance from the pole to road segment or Levenshtein distance between street names of both datasets.[1] Based on these values, a normalized score within the range of (0;1) will be obtained for each candidate road segment. The coefficients obtained from the following methods will determine how accurate each assignment is:

- distance score – which calculates a partial score based on physical distance from street light to the road segment.
- name score – which compares Levenshtein distances between street name attributes. Essentially, it splits the street names ascribed to both datasets into a list of strings and calculates the minimum distance in terms of similarity between any of the examined elements in the collection.

[1]Levenshtein distance is defined as a minimum number of character operations (e.g substitute, delete or insert) to transform an expression into another [41].

- type score – which evaluates the type of road the light is being ascribed to. It is based on the assumption that the lamps will most likely illuminate the streets of dominant categories (primary, residential etc.), while the others will receive a slightly lower rating (e.g. service road).

- number of lamps in a segment – which applies the partial score of 1 if the number of luminaires assigned to a given road segment is more than one, otherwise, the partial score of 0 is applied. The logic behind it is presented on the explanatory figure below 3.5.

Figure 3.5: A situation with a main road, which is being lit, and block a without any street lights. Assuming similar scores from other factors, the green point representing lamp might be incorrectly ascribed to this area, thus the penalty is being applied.

Weights for each of the partial scores will be determined empirically, the overall score will be computed and the road segment with the highest overall score will be ascribed to the street light. The outcome of the automatic process will be evaluated by manually reviewing the cases. The examination of the results will then allow to introduce additional heuristics or adjust a number of parameters to improve the accuracy of the algorithm, for example:

- The distance parameter of the buffer operation – if the distance is too low, the buffer of the street light might not intersect with any of the road segments,

- The lowest acceptable score based on which the assignment is made,

- Weights of specific scores.

3.2.2.4 Road segment quality analysis

After the adjustment of the results, road segments will be analyzed by calculating and examining several road properties, which will serve as quality indicators of the modeled lighting situation. First of all, the width of the road from the perspective of each street light assigned to the segment will be acquired.

For that purpose, a straight line will be constructed through the point which represents the position of the street light and the nearest point between the mentioned point and the road segment. The latter geometry will be obtained with the help of Shapely's `nearest_points` function, which calculates the closest points of two geometries. Again with the help of Shapely, the linestring will be extrapolated in both directions by means of a scale function, which is available from the affinity module. The intersection length between the linestring and road segment polygon will therefore characterize the width property. Example results of this process are presented in Figure 3.6.

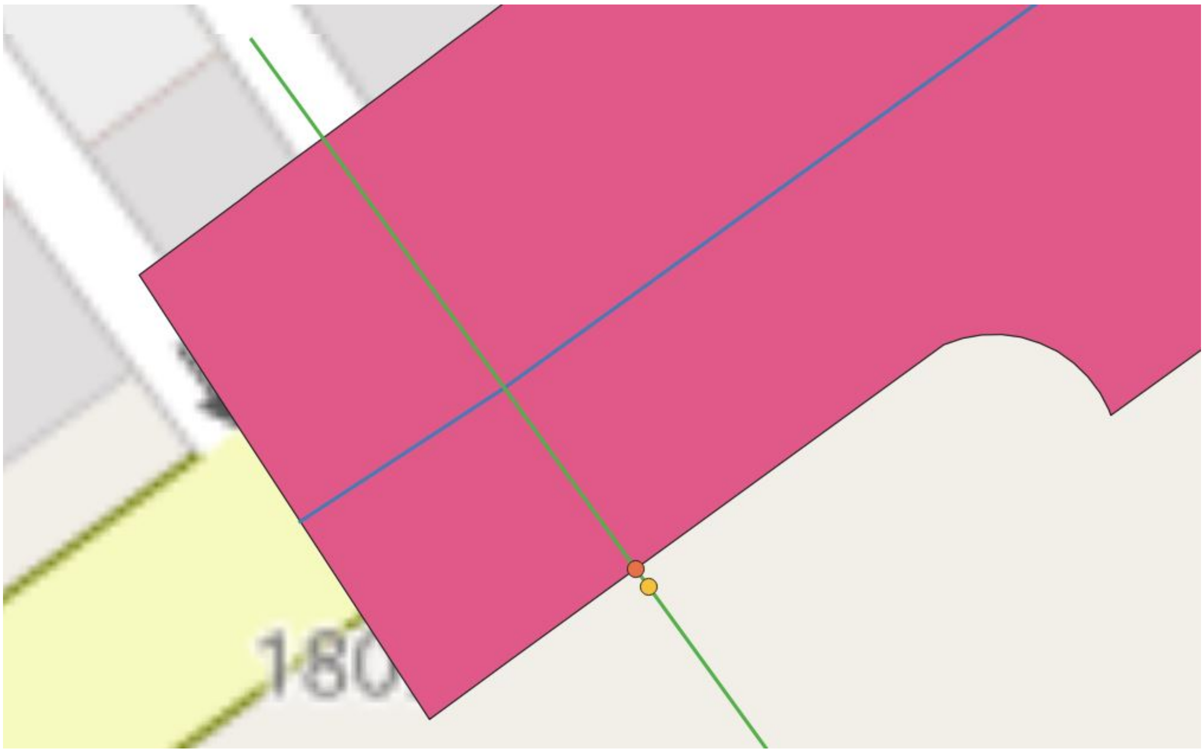

Figure 3.6: Width calculation process. Yellow point represents a light pole, orange point depicts the nearest point between both geometries and the green straight line is constructed through these points. The length of intersection between the linestring and the violet polygon defines the width.

Then, information related to the street segment linestring geometry will be merged to the main GeoDataFrame, which consist of street lights and associated road segment polygons, and basing on the spatial relationship between them, the spacing between the lamps in each road segment will be calculated. This indicator will be obtained in the following steps:

- First, points representing the street lights will be projected on the linestring which depicts the segment using Shapely's `project` method. As a result of this operation, the distance between the lamp and the beginning of the linestring will be obtained [37].
- Then, each group of lamps will be sorted by the acquired distance and the position of the lamp in a sequence will be defined.
- Finally, the spacing between each neighboring lamp in a sequence will be calculated. This operation will be performed with pandas' `diff` function, which computes the difference between two

elements of the DataFrame [34] (in this particular case the difference between the value and the next element in the row)

Both indicators will undergo a statistical analysis, and after reviewing questionable outcomes, the results will be tuned.

3.2.2.5 Preparation of the data for photometric optimisation

The final results will be converted into the input format of PhoCa software, which will be used for the optimization of examined lighting installation. The platform requires each row of the dataset to be a distinct street light, with the following attributes:

- unique identifier of fixture-segment pair,
- street light identifier,
- street name,
- road segment identifier,
- number of fixtures,
- road width,
- number of lanes*,
- lighting class*,
- arrangement of the street lights (one-sided or double-sided)*,
- spacing of street lights in a segment,
- distance from the street light to the nearest edge of road segment,
- height of the light pole,
- arm length of the light pole,
- fixture type (road, park or decorative)*,
- fixture wattage,
- latitude,
- longitude

Attributes marked with an asterisk refer to the data, which is not available from within the available datasets, thus it will be derived analytically. Moreover, as the software does not accept projected coordinates, they will be converted back to the spherical coordinate system (WGS 84).

Chapter 4

Case study

4.1 Dataset description

The study will cover the District of Columbia, with the data publicly available from Open Data DC[1] and GEOFABRIK[2] websites, with the datasets from the latter being based on the OpenStreetMap data. Both of them provide numerous spatial datasets for the analyzed area, available for download in several formats, including ESRI Shapefiles. Calculations will be performed on the following datasets:

- Street Lights (ESRI Shapefile) – containing street lights together with their locations and attributes (source: Open Data DC),
- Street Segments (ESRI Shapefile) – illustrated as linestrings representing the axis of particular roads in the District (source: GEOFABRIK),
- Roads (ESRI Shapefile) – which consists of precise roadway shapes in the form of polygons (source: Open Data DC)

4.1.1 Street lights

The streetlights dataset (70,956 records) is the fundamental dataset used for the analysis of the current installation. It models each lighting pole as a point with features, which characterize its geometry and additional parameters. Semantic description of all the attributes available in the dataset (thematically grouped) is given below. Furthermore, statistical analysis regarding the most significant features has been also presented below.

- Information regarding the entity and date of data capture – `add_to_gis`, `added_by`,
- Description of the asset type with the corresponding data related to its operational status. In the

[1] https://opendata.dc.gov
[2] https://geofabrik.de

current dataset asset types are described numerically; however, by examining older data collections from Open Data DC, names corresponding to each value has been derived – `asset_type` `(e.g. street light or alley light)`, `asset_status`, `why_inactive`,

- Attributes related to the location of the light pole like the street name, ward, or the city quadrant, within which the records are located and the association with a street segment from another dataset from Open Data DC (described in the following section) – `street_name`, `cross_street`, `house_no`, `street_segment`, `ward`, `quadrant`,
- Data concerning the US equivalent of European road lighting classes – `road_classification`
- A set of information regarding lighting poles – `facility_id`, `pole_height (in feet)`, `pole_type`, `pole_color`, `last_pained`, `pole_composition(e.g. wood or steel)`, `owner_description`, `power_feed (e.g. underground or overhead)`,
- Data describing the style, length, and number of arms of a specific light pole – `arm_style (e.g. pendant or truss)`, `number_arms`, `arm_length1`, `arm_length2`,
- Fixture parameters, which consists of its type, style, number of lights together with their wattage and other informative data regarding the other equipment installed, condition of the fixture, its manufacturer, and the modification history – `object_id`, `light_type(e.g. street light or alley light)`, `fixture_style (e.g. cobra head or post-top)`, `number_light`, `wattage1`, `wattage2`, `other_equipment`, `cct_description (color temperature)`, `shield_description`, `condition`, `light_manufacturer`, `light_history`, `led_in_operation`,
- Corrections related to any attribute with a date of last modification – `is_modified`, `what_modified`, `last_modified`,
- The most important attribute, which characterizes each record in the dataset geographically – `geometry`

Figure 4.1 demonstrates street light fixture types, which indicates that High Pressure Sodium lamps are dominant in the dataset, while the following figure (4.2) displays the asset types, which are present in the street lights dataset. The asset type is an attribute, which is extremely useful as a basis for data selection process.

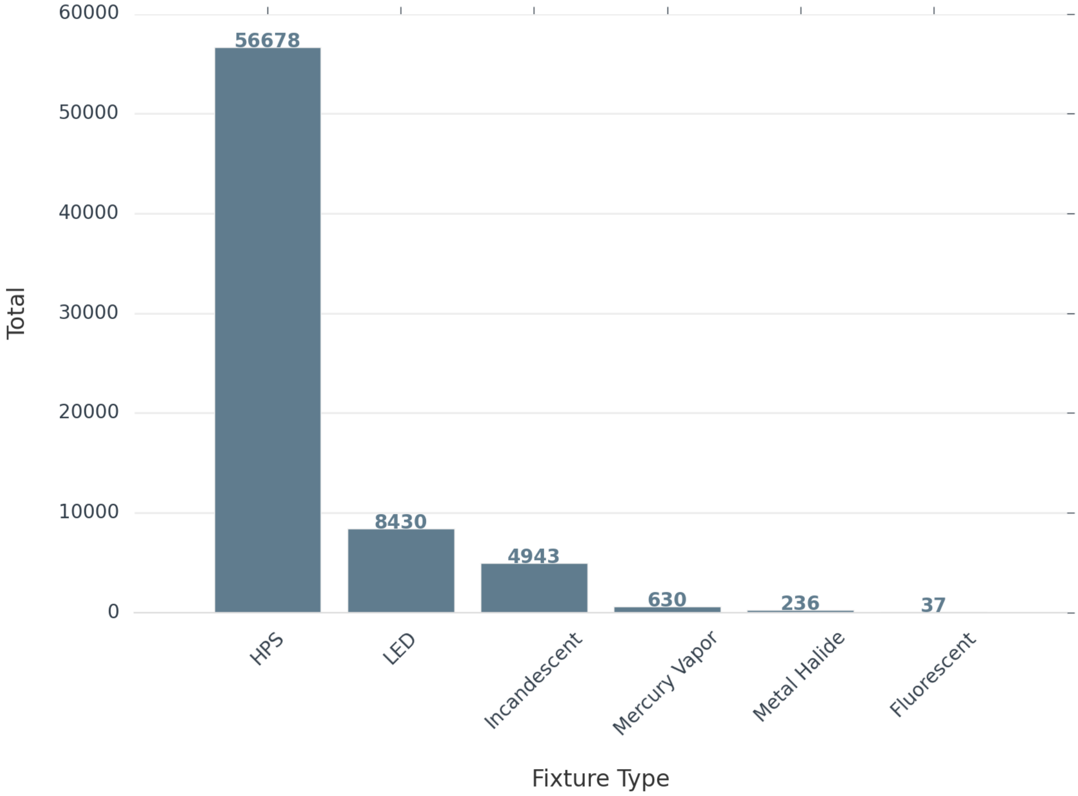

Figure 4.1: Fixture types

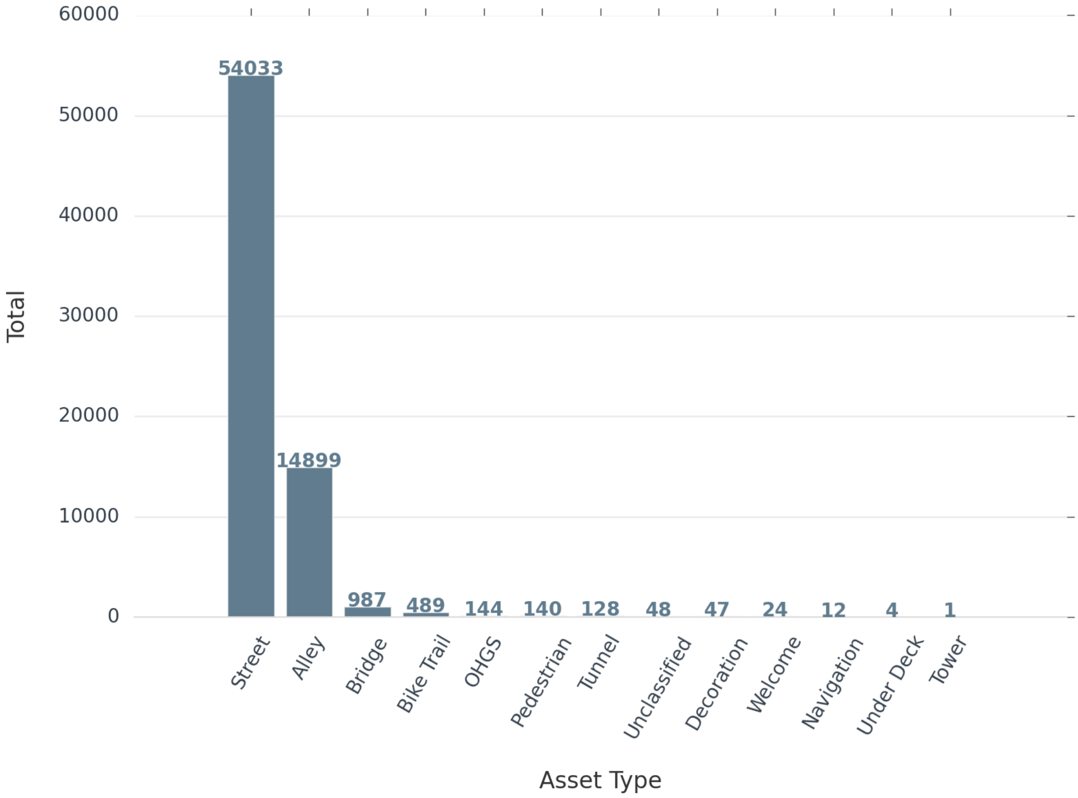

Figure 4.2: Asset types

Figure 4.3 is a classification of poles concerning the number of fixtures installed (lighting poles with 1 fixture were ignored in the following figure, as they make up the vast majority of the dataset i.e. 67,705 records):

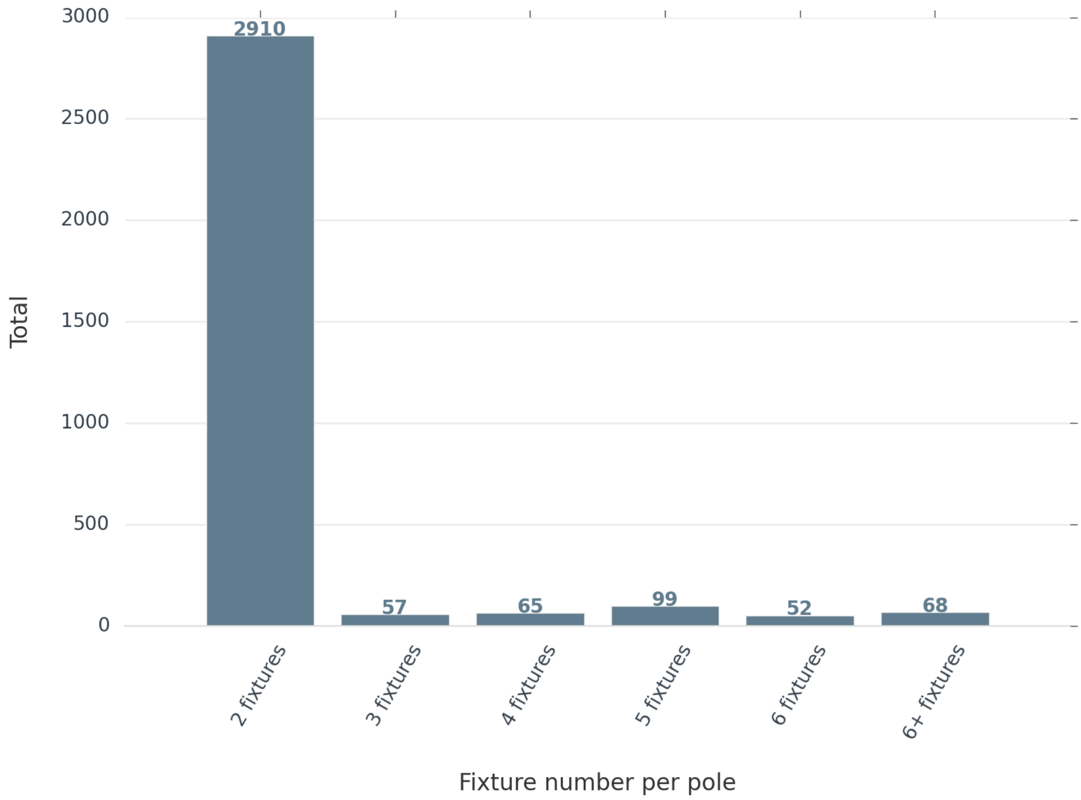

Figure 4.3: Number of fixtures installed per pole

As the figure 4.3 indicates, there are 3,251 lighting poles with more than 1 fixture. The dataset, however, does not provide the information about areas they illuminate, therefore is it important to determine a number of carriageways associated with each pole. A differentiation between poles with multiple fixtures is given below.

- 2-fixture pole, with arms positioned perpendicularly to the carriageways, where each luminaire illuminates a distinct road (Fig. 4.4). In alternative version of this setup, one of the fixtures illuminates the sidewalk,
- 2-fixture pole, with arms positioned in parallel to the road (Fig. 4.5),
- 2-fixture pole, with arms generally arranged at the angles within the range of 45-90° between them. Fixtures are illuminating either one or two roads,
- Multiple-fixture pole with decorative/park designation (Fig. 4.6),
- Other combinations of aforementioned setups.

Figure 4.4: Lighting pole with 2 fixtures which illuminate two parallel roads

Figure 4.5: Pole with 2 fixtures, poletop-mounted

Figure 4.6: Decorative lighting pole with 5 fixtures

Considering the multitude of street light types, it is important to categorize them by a number of areas they illuminate. This can be done by applying heuristics based on other features such as the arm lengths, fixture style etc. Therefore, to ensure the accuracy of lighting design, the following analysis has been performed:

- Street lights with more than one light and number of arms different than 1 have been acquired (2,903 records). Most of them have 2 lights (2,835 records), however, records with 3 (4 records) and 5 (64 records) lights are also present. Moreover, as it comes to the number of arms, they are either 2 (735 records) or N/A (2,168 records).
- Lights with an arm number of N/A has been dropped. This assumption is based on the fact that all these records, have a "Posttop" value assigned in the `fixture_style` attribute and, as it can be seen in Figure 4.2, the light in poletop-mounted street lights is focused on a single area only (735 records remaining).
- The observation of the records with double arms suggests that the shorter arm lights up pavement, and therefore lights with asymmetrical arm lengths has been filtered out (661 records remaining).
- Lights has been also analyzed in terms of `arm_style` attribute. Fixtures with the arm style of either "Decorated Straight" or "WP Decorated Scroll" have been identified to be positioned in parallel with the road and has been dropped consequently (507 records remaining).
- The aforementioned conclusion was also applicable for lights with `asset_type` of "Bridge Light", thus these 98 records have also been rejected.

Eventually, 409 street lights, which illuminate 2 areas, have been determined. All lights that have been dropped during analysis, will be considered as a single lights in further calculations.

4.1.2 Street segments

The street segments dataset (32,841 records) represents all streets as linestrings. Initially, the calculations were meant to be based on another Open DC dataset – Street segments – however, the streets in the mentioned data collection are modelled as traffic arteries i.e. a street with three distinct carriageways is represented as a single line situated in the middle of the road. Due to the fact, and despite already having a direct association between a segment and a light pole (`street_segment` attribute), the decision was made to use the OpenStreetMap dataset instead, to improve the accuracy of the results. "Street segments" contains, except the ever-present identification number and the geometry coordinates of the shape, several valuable attributes including the type of the road segment with the corresponding code (`fclass`, `code`) (Table 4.1), street name (`name`), highway location reference (`ref`), maximum allowed speed (`maxspeed`) and the street layer (useful for identification of e.g. underground roads). Moreover, boolean indicated attributes such as `bridge` or `tunnel` specifies if the segment crosses a certain road structure and the `oneway` feature implies the traffic direction.

Table 4.1: Road segment types

	Total
service	12865
footway	6842
residential	4954
primary	2248
tertiary	1464
secondary	1460
motorway	445
path	434
motorway_link	401
trunk	331
cycleway	291
other	1106

4.1.3 Roads

The roads dataset (52,339 records) contains polygons, which represent the roads in the District of Columbia. The precise shapes have be obtained by combining LiDaR data with orthophoto imagery and other map sources. The records contain an information about their area type e.g. road, paved drive, intersection, median island, parking lot, etc. The distribution of road types in presented in Table 4.2. Other attributes in the dataset include identification numbers, the capture date, as well as the geographical features of the shape such as the area or length. This dataset, is not related to neither the street lights nor the street segments datasets.

Table 4.2: Road types

	Total
Paved Drive	15696
Road	13521
Parking Lot	8911
Intersection	7467
Alley	3361
Median Island	2702
Hidden Road	410
Other	271

4.2 Details of processing

4.2.1 Cleaning, conversion and unification of the data

After reading the data with the GeoPandas module, the datasets have undergone an analysis of the coordinate reference system assigned to the geometry attribute. It has been found that all the shapes from the original datasets have been stored in the WGS 84 geographic coordinate system. Therefore, to ease the computations and support better analysis, the spatial data has been transformed to a projected coordinate system. Local projection, which is used in the area of Maryland state and its surroundings, has been selected for each dataset (NAD83). Then, a normalization process was performed on the street name attributes of both street lights and street segments datasets, as in the further stage of calculations these values will be directly compared. For that reason, the values have been lower-cased, missing values have been replaced with an empty string (as the similarity analysis function cannot take NaN values) and both the street types (e.g. 'street', 'place', 'alley', 'avenue', 'terrace'), and the positional indicators (e.g. Northwest, SW), which were included in each record, have been deleted. As an example, the value of "New Jersey Avenue Northwest" has been converted to "new jersey" to possibly reduce the risk of a wrong comparison. Moreover, several attributes (`pole_height`, `number_arms`, `arm_length1`, `arm_length2`, `wattage1`, `wattage2`), has been incorrectly stored in the dataset as a string type (sometimes together with its units e.g. '30 ft'). Therefore, unit abbreviations have been removed, values have been turned into numeric counterparts and the conversion to the metric system has been performed. Also, a variety of string representations of NaN values (e.g. "N/A", "NaN" or empty string) has been replaced with a coherent equivalent, available in the numpy module (`np.nan`). Finally, filtering operations, described below, have been carried out individually on each dataset, and afterward, attributes which have not been perceived as particularly useful have been dropped.

- Street Lights (filtered using the `asset_type` attribute) – As the analysis scope will exclusively cover street lights along the main roads, only records with asset types of either 'Street Light' or 'Bridge Light' will be taken into account in further calculations. The size of the GeoDataFrame object has been decreased by 22.5% (70,956 to 55,020 records) and the final set of attributes consists of the following: `id`, `asset_type`, `street_name`, `road_classification`, `pole_height`, `number_arms`, `arm_length1`, `arm_length2`, `arm_style`, `fixture_style`, `light_type`, `number_light`, `wattage1`, `wattage2`, `geometry`.

- Street Segments (filtered using the `fclass`, `tunnel` and `layer` attributes) – Driven by the same reasoning, segments with the following 'fclass' values have been excluded: 'footway', 'path', 'cycleway', 'steps', 'pedestrian', 'track', 'living_street', 'track_grade1', 'track_grade3', 'bridleway'. Moreover, records that suggest underpass in the 'tunnel' attribute and values below 0 in the 'layer' feature have been omitted, as underground lighting will not participate in the computations. The size of the GeoDataFrame object has been decreased by 25% (32841 to 24662 records) and the final set of attributes consists of the following: `id`, `fclass`, `street_name`, `maxspeed`, `layer`, `bridge`, `geometry`.

- Roads (filtered using the `description` attribute) – Similarly to the aforementioned datasets, only records characterized by 'Road', 'Intersection' or 'Hidden Road' values in the description column have been preserved, with the latter retained to keep the consistency of several carriageways. The size of the GeoDataFrame object has been decreased by 60% (52,339 to 21,398 records) and the final set of attributes consists of the following: `id`, `description`, and `geometry`.

Cleaned datasets have been saved in GeoPackage format, which proved to be more flexible than Shapefiles e.g. for the visualization purposes.

4.2.2 Automatic assignment of segment linestrings to the lit area shapes

A spatial join performed in the first iteration of the process resulted in a GeoDataFrame consisting of 50,621 records. 5,679 segments and 140 roads have not intersected at all, which corresponds to 23% and almost 1% of the initial size of the datasets respectively. Then, by grouping the GeoDataFrame object by the `id` of a road segment and aggregating all identifiers of intersecting road polygons, the following DataFrame (Table 4.3), with the row length of 18,983, has been created:

Table 4.3: Structure of the DataFrame containing road segments and ids of intersecting roads

osm_id	road_id	count
105492131	{20913}	1
105649640	{5584, 5553, 5569, 5638}	4
105673914	{6129, 5034, 5023}	3
105742485	{12433, 12434, 12435, 12438, 12440, 12410}	6
105742763	{12396, 12398, 12407}	3

After merging the segment's geometry information to the above table, a buffer of 10 meters has been applied to the object's spatial attribute. A distance of 10m has been selected to avoid shape protrusion which may affect the assignment process. Polygons in each group underwent the union operation and the shapes have been trimmed by previously retrieved linestring buffer. The distribution of the number of polygons intersecting with each street segment has been presented in Figure 4.7. The procedure was successful for all road segments, and as a consequence, a total of 18,998 road segments have been created. Most of the constructed objects have yielded a satisfactory result (Fig. 3.3), however, after a scrupulous review, a major issue has been identified. During the process, a significant number of multipolygons (2,930 out of 18,998) has been formed, which is undesirable for further calculations. Problematic geometries have been identified as originating from the following reasons:

- segment intersects with roads only on its ends,
- segment crosses the road centrally through the previously filtered median island,
- occurrence of barely noticeable holes between road polygons,
- inaccurate road shapes.

In the following subsections, particular issues will be described and solutions will be proposed.

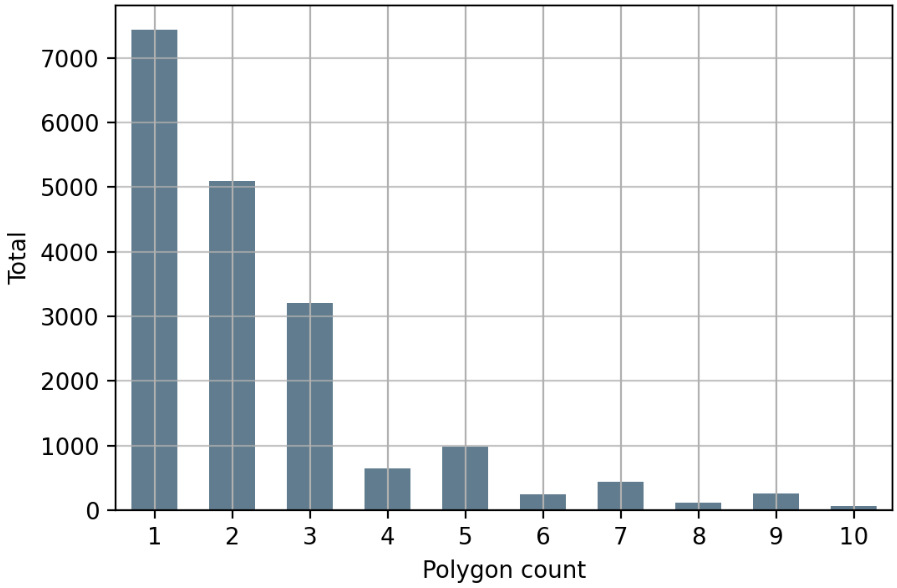

Figure 4.7: The distribution of the number of polygons intersecting with each street segment (1–10 polygon groups only, as the larger groups constitute of only 2.5% of cases)

4.2.2.1 Issue 1 – Segment intersects with roads only on its ends

The vast majority of problematic unions have been created in a way presented in Figure 4.8. Segment linestring crosses the road, which is not involved in calculations; however, it also intersects with the main carriageways on its ends. A proposed algorithm extension, in this case, would be to calculate the total intersection percentage between linestring and road segment polygon and drop geometries, which will obtain a value below the acceptable minimum. The minimum score will be acquired empirically, by manually reviewing the cases.

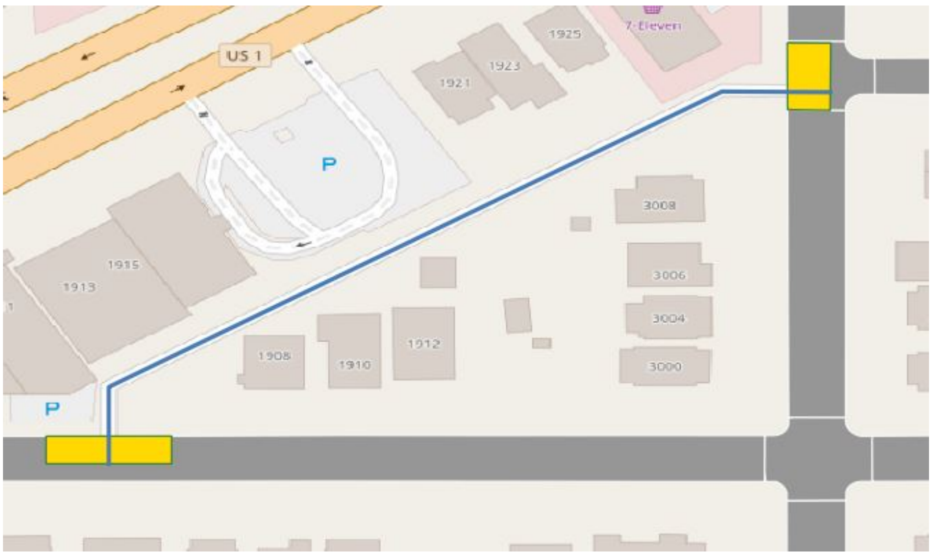

Figure 4.8: Issue 1 – example

4.2.2.2 Issue 2 – Segment crosses the road centrally through previously filtered median island

Another issue has been depicted in Figure 4.9. It can be seen that the segment linestring crosses the road centrally through the previously filtered median islands, hence creating a MultiPolygon object. The situation is only relevant if the polygons around the median are split, rather than being a unified shape surrounding it. To overcome the problem, median strips might be included in the second iteration of the procedure and the evaluation of the results will be repeated.

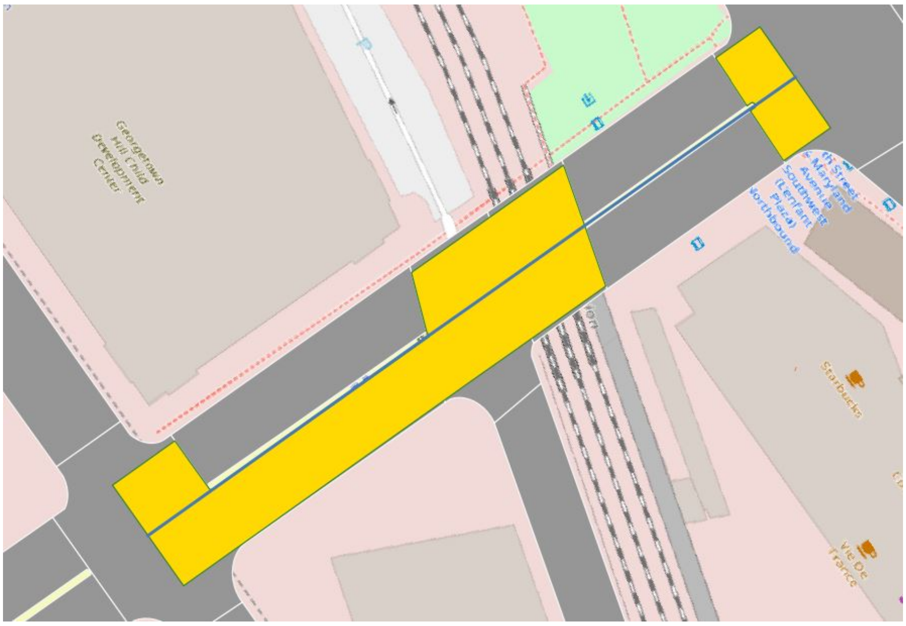

Figure 4.9: Issue 2 – example

4.2.2.3 Issue 3 – Occurrence of barely noticeable holes between road polygons

Figure 4.10 displays yet another problem related to the occurrence of small gaps between the shapes of the roads. To address this issue, a buffer will be applied to all geometries in a road segment, the shapes will be unified, and again a buffer operation with an opposite distance value will be performed, to retain the original structure of the road.

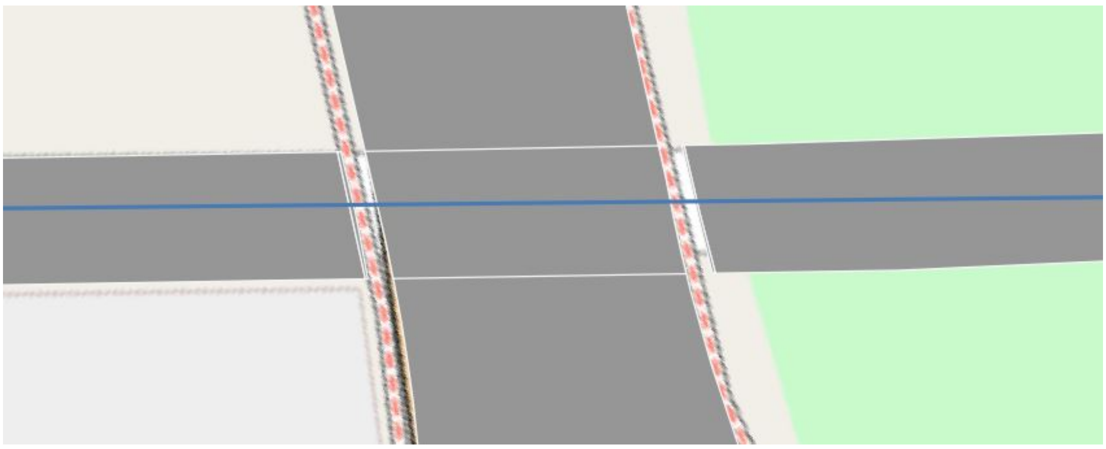

Figure 4.10: Issue 3 – example

4.2.2.4 Issue 4 – Inaccurate road shapes

The creation of wrong geometries, in this case, is mostly due to the invalid structure of road polygons (Fig. 4.11). These shapes, instead of being modelled as a single carriageway, consist of multiple roads, and therefore, if a linestring buffer crosses these parts of the road, a MultiPolygon object is created. A heuristic, which would remove all shapes which do not intersect with a segment linestring will be applied. As a result, a single polygons will be obtained.

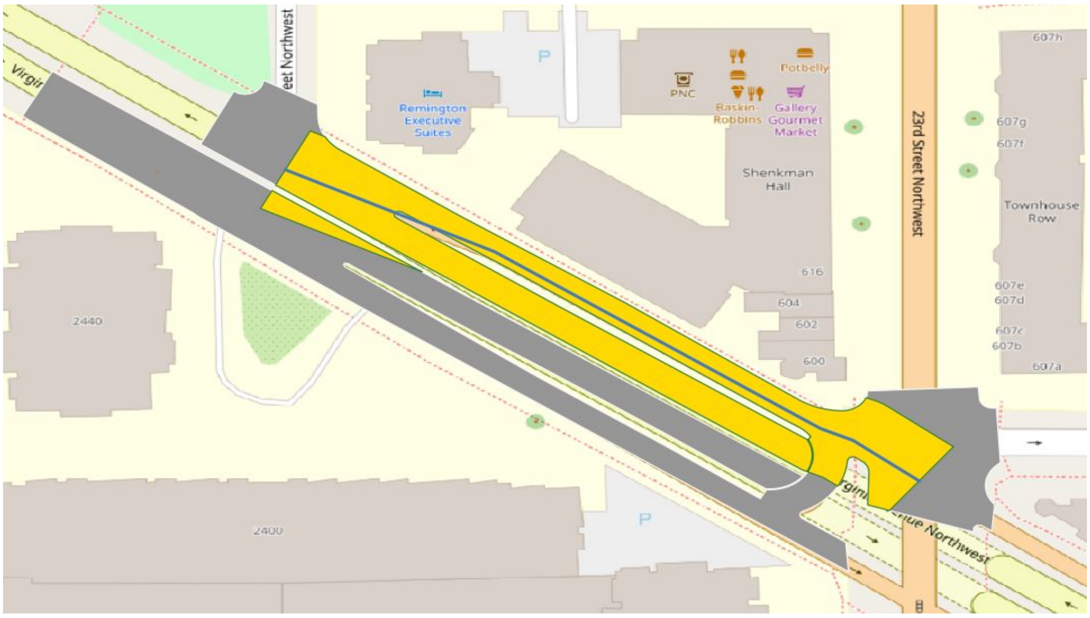

Figure 4.11: Issue 4 – example

4.2.3 Automatic assignment of segment linestrings to the lit area shapes - algorithm extension

In the second approach to the process of assigning segment linestrings to the lit areas, solutions to the aforementioned issues have been applied. First of all, the median islands have been included in the dataset and linestring length has been computed for all the records. Parts of MultiPolygon objects, which do not intersect with segment linestring, have been excluded, resulting in a decrease of unwanted geometries number from the initial 2,935 to 2,406. Then, the intersection rate between linestrings and road segment polygons has been obtained. In order to achieve the most accurate results, the interiors of MultiPolygons has been dropped (cases when a linestring crosses median island). By gradually reviewing the geometries starting from the lowest percentage, it has been concluded that all shapes below the value of 0.4 are invalid, and therefore will not take part in further calculations. Thus, the total number of MultiPolygons declined from 2,406 to 77. The distribution of intersection percentage between segment linestring and multipolygons has been presented in Figure 4.12. Afterward, an algorithm extension described in 4.2.2.3 has been applied. Also, by manual review, a minimal percentage of polygons for which the process will be applicable has been acquired and amounted to 96.7%. The shapes have been merged and the overall number of multipolygons has been decreased by 21.

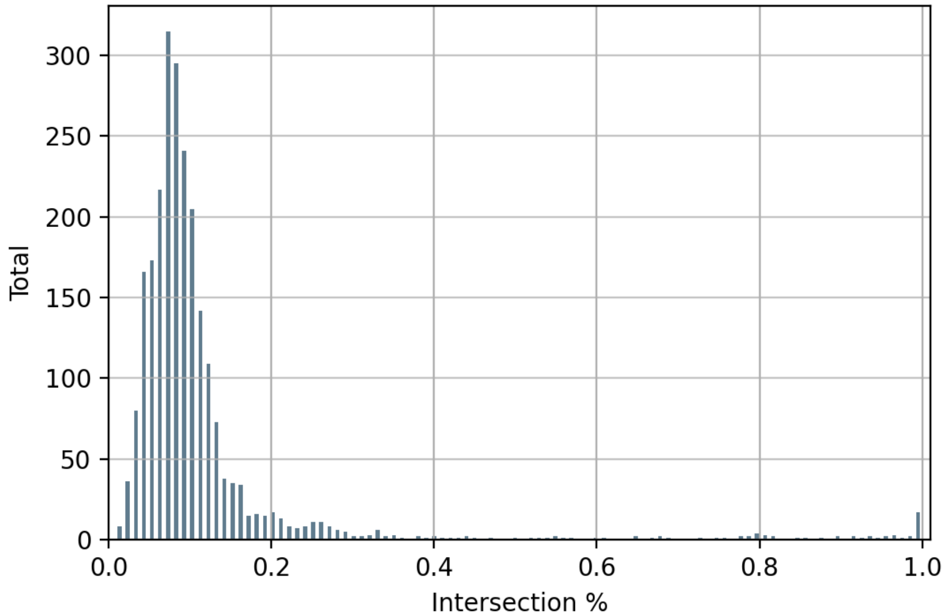

Figure 4.12: Distribution of intersection percentage between segment linestrings and multipolygon geometries

Finally, a solution to the second issue has been examined; however, the results were unsatisfactory. It turned out that by including median islands in the dataset, the resulting from unary union shapes have still been inaccurate (Fig. 4.13). Therefore, for these and remaining cases (54 shapes), only the parts of the geometries, which have the longest intersection with the segment linestring have been kept in the form of polygons.

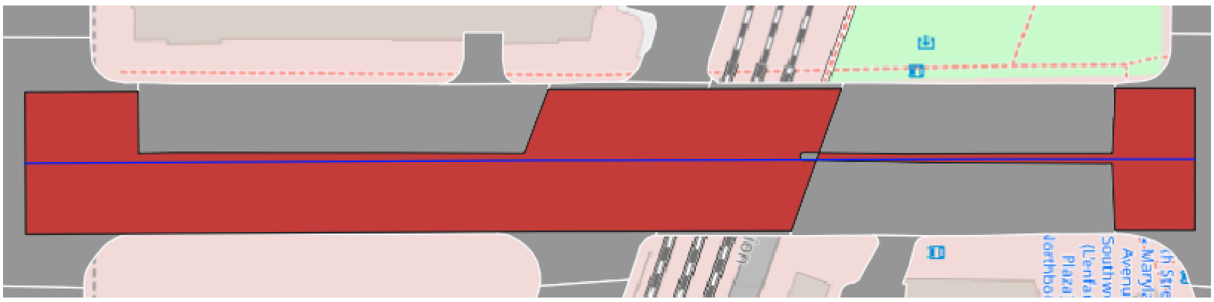

Figure 4.13: Inaccurate shape resulting from unary union operation

4.2.4 Automatic assignment of lamps to the lit areas

Similarly to 4.2.2, the procedure was started by executing a spatial join (again based on the intersection relationship) between previously created road segment polygons and buffered points, which refer to street lights (a buffer of 5m has been initially applied to the dataset containing street light data). The operation resulted in a GeoDataFrame consisting of 73,184 records with 648 non-intersecting lamps. By reviewing records that were not included in the geopandas data structure, it was found that the applied distance was to low, as several lamps, which in fact illuminate a certain area, has been disregarded.

Therefore, spatial join has been conducted again with a higher distance of 7 meters, producing a new GeoDataFrame (79,743 records) which will be the basis for the further calculations. The assignment will be performed in accordance with the distinction made in 4.1.1 (number of areas illuminated by street lights).

4.2.4.1 Assignment of lamps, which illuminate two areas

As described in 4.1.1 a total of 409 lights with this characteristic have been identified recognized, which translates into 1,015 records in the GeoDataFrame resulting from the spatial join. 8 lamps, which do not intersect with any nearby area has been identified. Moreover, further 30 records, which intersect with one area only have been found. After manual examination of the aforementioned street lights, it was concluded that both records which were not included in the spatial join and records which did not intersect will the minimum of 2 areas, will be considered as double lights which light up a single segment. The following assignment accuracy measures will be applied for the remaining 371 fixtures and the candidate road segments:

- distance score – The setback of 0 to 3 meters is practically an equally good assignment and it is a matter of adjusting the mounting angle of a fixture by several degrees. At some point, as the distance is significantly greater, this possibility vanishes. Due to this fact, the normalization formula will be an exponential function, instead of standard linear dependency and has been described in (4.1).

$$distance\_score = 1 - (1/1 + e^{5-distance}) \tag{4.1}$$

  The exponent in 4.1 has been also empirically tuned in accordance with the maximum distance between the lighting pole and road segment, which was formerly defined as 7 meters. Therefore, candidates with the setback of 0-3 meters will receive scores between 1 and 0.88 respectively, a setback of 5m will correspond to the score of 0.5, and road segments with a distance of around 7 meters will be assigned a score around 0.

- type score – the lamps, which illuminate the streets of dominant categories will receive a score of 1, otherwise a score of 0.8 will be applied (service road, unclassified road),

- intersection percentage – as the road segment polygons with low intersection percentage has not been omitted (to prove the accuracy of the algorithm and not to accidentally skip correct candidates), this measure will be additionally used to determine to most suitable assignment candidates.

Weights for each score have been set as equally important and the overall score has been computed. Then, the assignment process has been performed, taking into consideration only the candidate roads, with the overall score higher than 0.5. The results of the assignment have been satisfactory (Fig.4.14), and therefore the weights of particular scores have not been adjusted. However, several incorrect ascriptions have been spotted, thus the minimum score needed to be increased. In order to acquire an accurate minimum score, all assignments within the score of 0.5–0.7 have been examined. As a result, the minimum assignment score of 0.6 has been obtained. Again, street light with an insufficient score (3 records), will be considered as a single-area double lights in the further calculations.

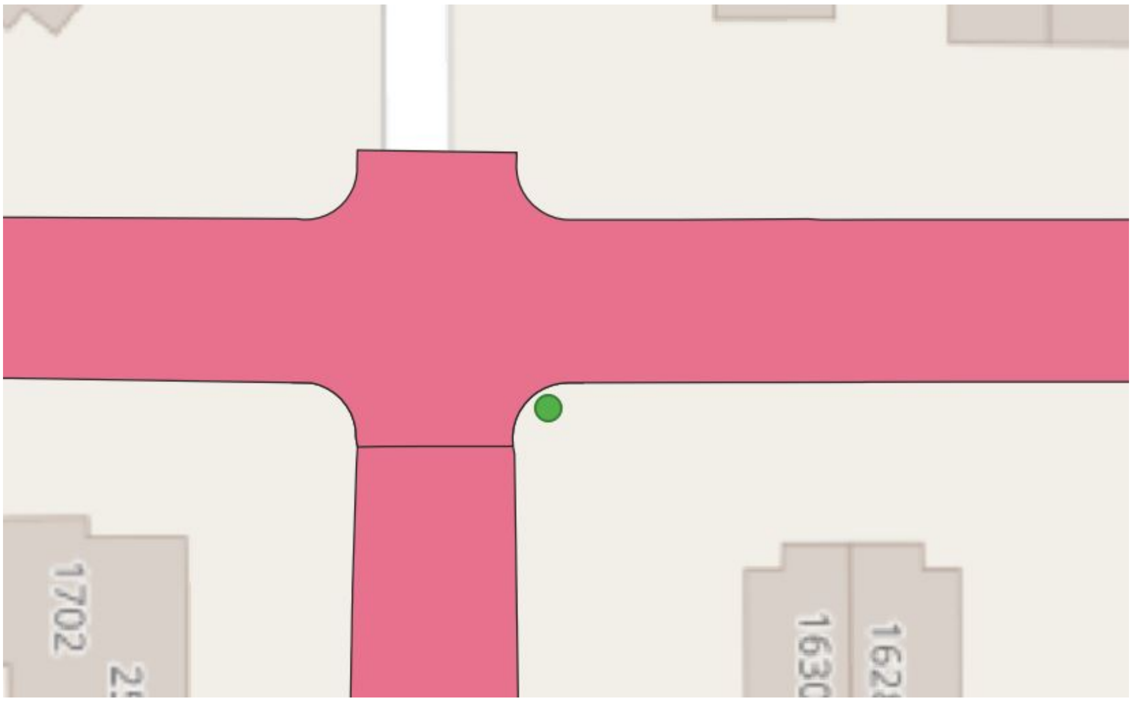

Figure 4.14: Street light assigned to both neighbouring areas

4.2.4.2 Assignment of lamps which illuminate a single area

Initially, the already assigned street lights have been dropped from the GeoDataFrame (78,734 records remaining). 509 records have not been ascribed, however, after manual examination it was decided not to increase the maximum distance, as it would not increase the accuracy of the algorithm. The matter is much more related to:

- the absence of data in certain areas (either road polygons or segment linestrings),
- incorrect asset type in the initial dataset (e.g. alley lights marked as street lights),
- partially unresolved problem described in 4.2.2.2 (creation of holes in the shapes due to the removal of multipolygon parts with low intersection percentage)

The same assignment scores will be applied for the candidate road segments in this case, together with a pair of additional measures, namely, "number of lamps in a segment" and the "name score". Comparison result between street name attributes of the latter score will also be normalized exponentially, however, the score will be much more strict. Candidates with the Levenshtein distance of 0 and 1 will receive scores of 1 and 0.92 respectively, a distance of 2 will correspond to the score of 0.5, and road segments with the higher distance will be assigned a score around 0. The formula has been described in (4.2).

$$name\_score = 1 - (1/1 + e^{5-(2.5*min\_lv\_distance)}) \tag{4.2}$$

Weights for each score have been also initially set as equally important and the overall score has been computed. The assignment process has been performed, taking into consideration only the candidate roads, with the overall score higher than 0.5. The results of the assignment have been satisfactory in general; however, several ascriptions have not been made, which was found to be related to wrong weight distribution. First of all, due to the exaggerated weight of the measure which favors multiple lights in a road segment ("number of lamps in a segment"), several cases such as the situation presented in Figure 4.15 has been observed. Moreover, the overestimated weight of "intersection percentage" has made the assignment of correct shapes impossible (Fig. 4.16). Therefore, the weights of the aforementioned scores have been halved.

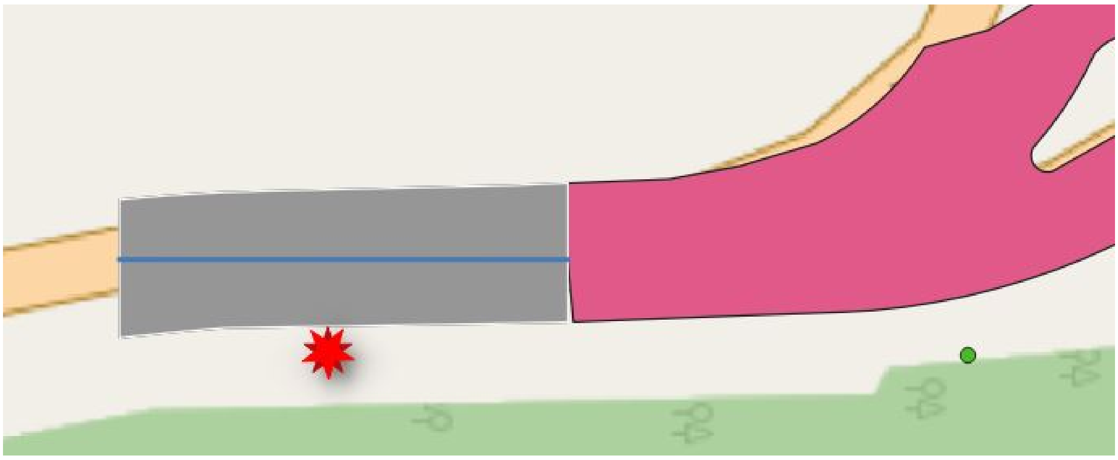

Figure 4.15: Example of no assignment due to upscaled weight of "number of lamps in a segment" score

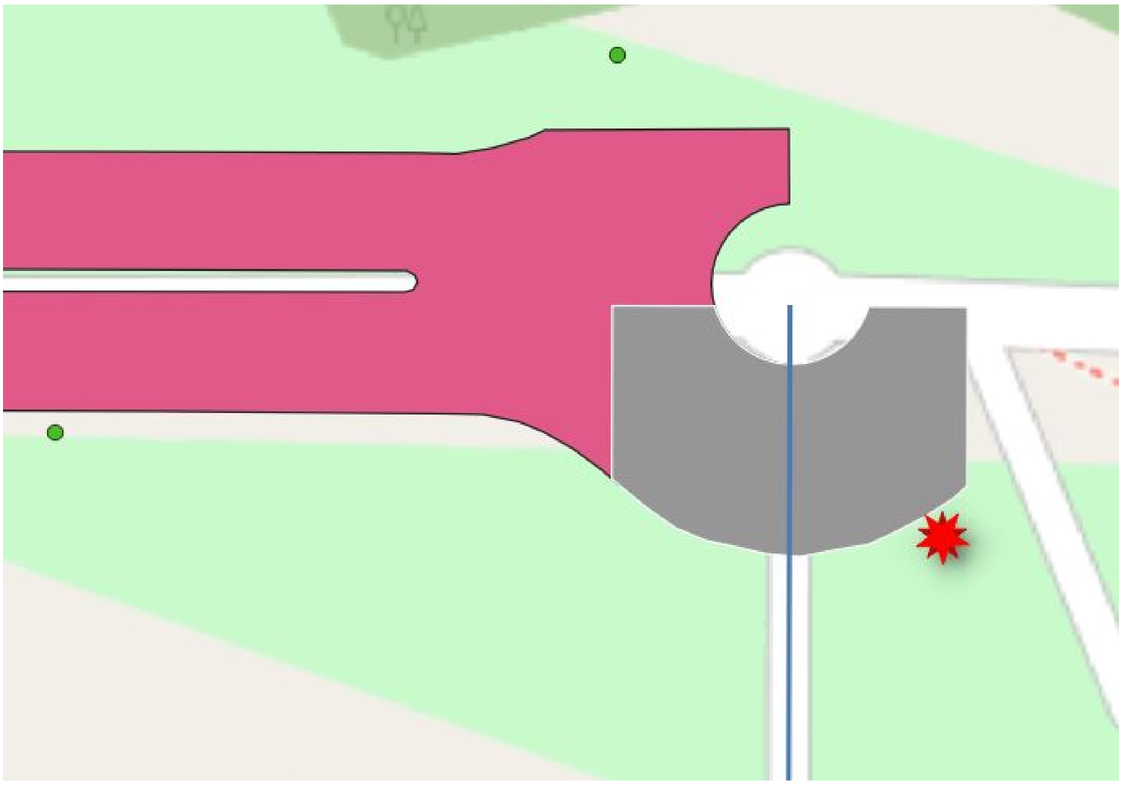

Figure 4.16: Example of no assignment due to upscaled weight of "intersection_%" score

After the weight adjustment, the process has been repeated. While the importance tuning handled the issue of wrong assignments, the minimum overall score needed to be increased as some undesired ascriptions have been made. Therefore, all assignments within the score of 0.5-0.6 have been examined and as a result, the minimum assignment score of 0.547 has been obtained. Consequently, 26 records below this score have not been assigned (otherwise, they would be assigned to the wrong area). On the other hand, the remaining 54,117 has been assigned the most suitable candidate road segment. Furthermore, street lights, which light up two areas, have been duplicated and transformed into single-area lamps (each record has been split into 2 separate rows, with a unique id and corresponding attribute values such as arm length or fixture wattage) and have been merged to the data structure (a total of 54,853 records). Based on the road segment identification numbers assigned to street lights, a dataset of all ascribed road segments (8,259 records) has been also acquired for further calculations.

4.2.5 Road segment quality analysis

4.2.5.1 Width of the road segment

The width of the road from the perspective of each street light assigned to the segment has been calculated in accordance with the procedure described in 3.2.2.4. However, after examining the results, several cases with width over 20 meters have been identified, which is especially odd, as the linestring buffer (4.2.2) had been previously specified to be 10 meters (theoretically disqualifying width values above 20m). Therefore, records with problematic width values have been visualized and the following reasons of incorrect measurement have been determined:

- Street light positioned diagonally or laterally in relation to the road segment (Fig. 4.17),
- Inaccurate attribute values (Fig. 4.18),
- Inaccurate differentiation of street lights, which illuminate two road segments (Fig. 4.19)

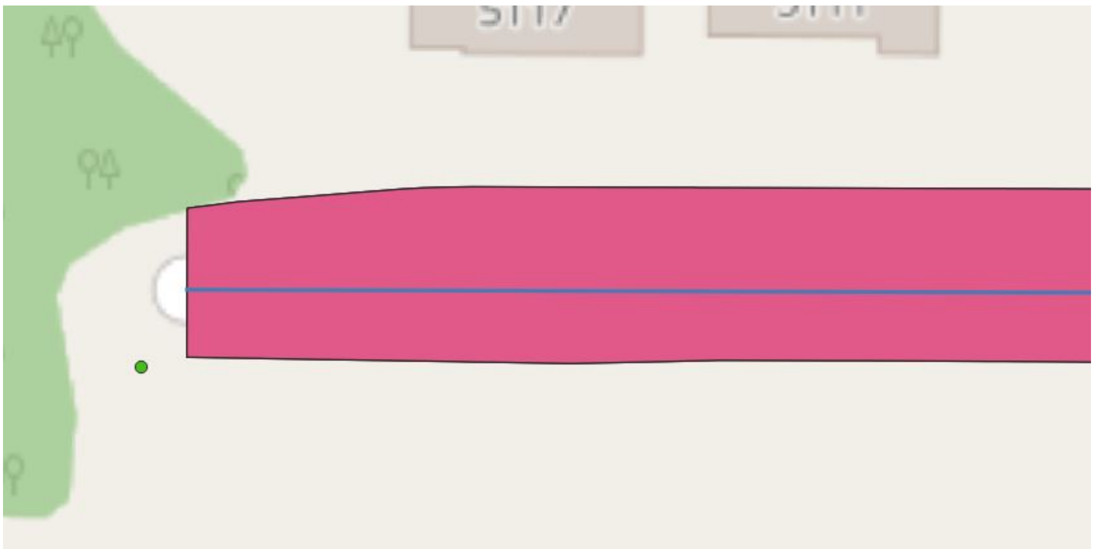

Figure 4.17: Example of street light positioned laterally in relation to the road segment

The most common circumstance with wrong width measurement is associated with the position of street light relative to the road. As the algorithm is built upon calculating the length of a line constructed through street light point and the nearest point between both geometries, width is not measured correctly (not perpendicularly) in case of lighting poles located diagonally or laterally in relation to the segment. Moreover, as presented in 4.18, the street light has been wrongly assigned to the yellow segment, due to the incompatibility in the street name attribute (a continuous road with different street names), thus resulting in an inaccurate width value.

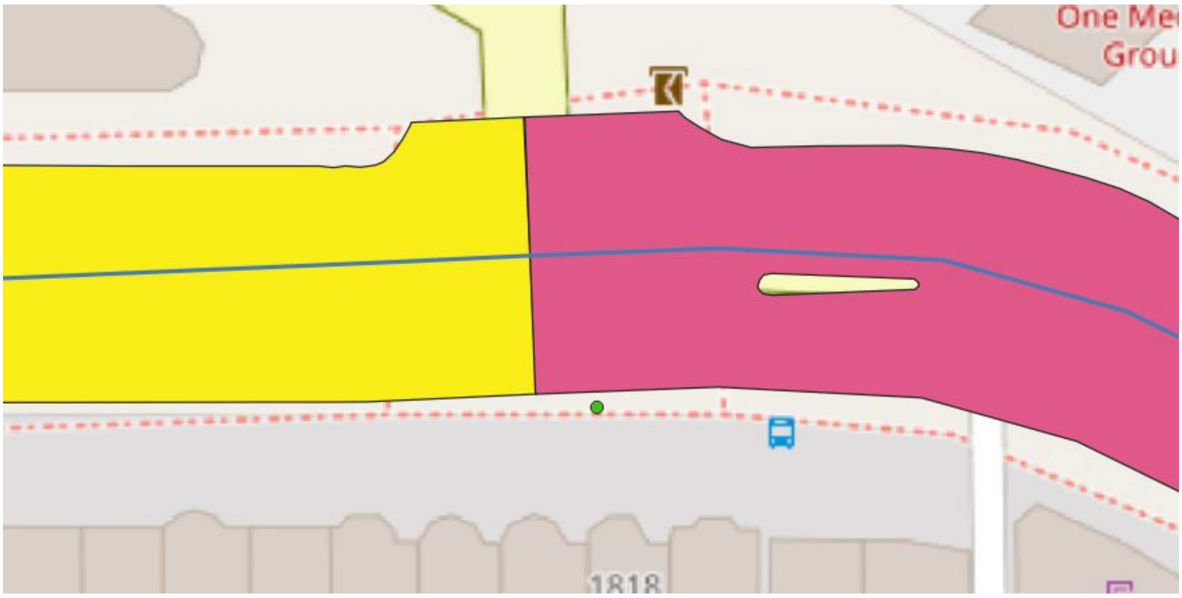

Figure 4.18: Example of wrong assignment due to inaccurate street name attribute values

Figure 4.19: Example of wrong differentiation of lamps which light up two segments

The last case considers the inaccurate differentiation of double-fixture street lights. In this particular example (Fig. 4.19), heuristics described in 4.1.1 (e.g. equal arm length) indicated that the lighting pole will light up two areas. Visualization of the examined records confirmed that in fact, both street lights are focused on one segment only.

Precise width values are essential, especially as the road width median of each segment is necessary during the preparation of the data for photometric optimization. Therefore, in order to acquire representative width values, the following procedure has been implemented:

- If there is only one street light in a segment and has a width value of over 20 meters assigned, the dataset width median (10.3 m) will be applied to the value instead,

- If there are two street lights in a segment and one of them has a width value of over 20 meters assigned, the value will be leveled with the width of the second street light,

- If there are more than two street light in a segment and more (or equal) than half of the street lights ascribed have a width value of over 20 meters assigned, the dataset width median will be applied to all values instead.

The final distribution of road width has been presented in 4.20

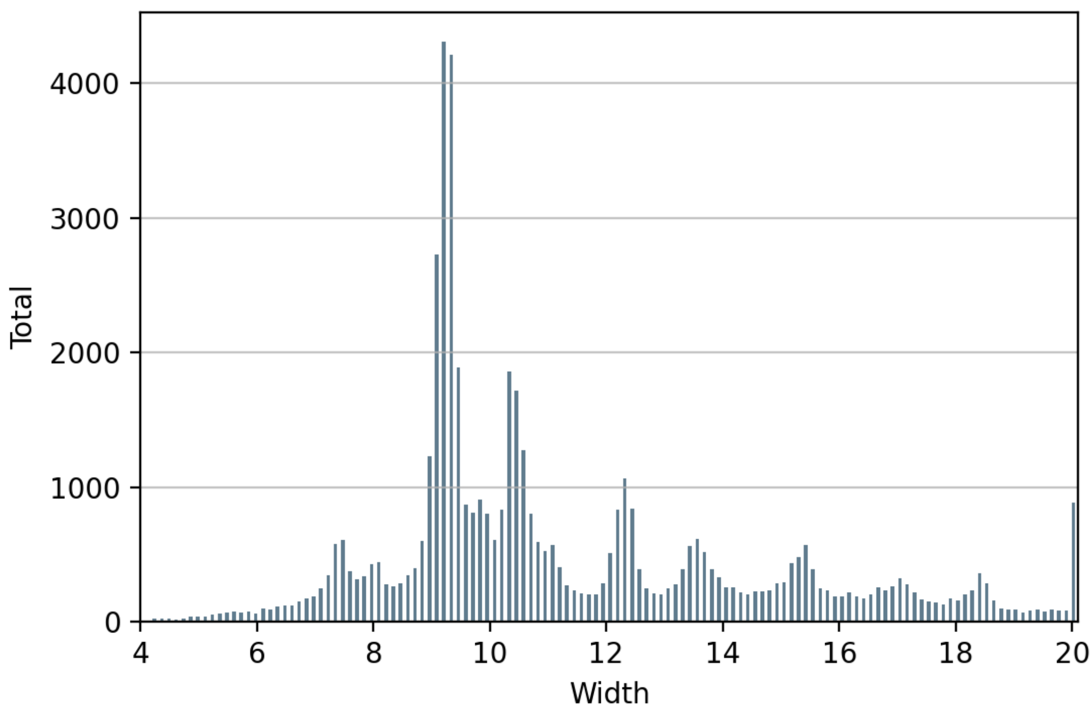

Figure 4.20: Road width distribution

4.2.5.2 Spacing of lamps within a road segment

First of all, the spacing of street lights within each segment has been calculated in accordance with the procedure described in 3.2.2.4. Then, a median and a variance of spacing have been computed for each road segment, with the latter being used as a segment quality indicator. Generally - the lower the variance, the better. This is due to the fact that the median of spacing in a particular road segment is the calculation basis during the photometric optimization process, rather than its discrete spacing values and thus the outcomes are much more accurate. Spacing variance results have been analyzed and the following value groups have been distinguished:

- N/A value – 31.7% – Value associated with the pole count of less than 3 within a road segment. Segments with single street light do not have the spacing value at all and segments with two lights have only one, thus the variance is not computed in both cases. The lack of the indicator makes quality evaluation of these road segments impossible.
- Value lower than 100 – 48.3% – The analysis demonstrates, that low spacing discrepancy has been identified among the majority of examined road segments, which indicates their excellent quality (Fig. 4.21).

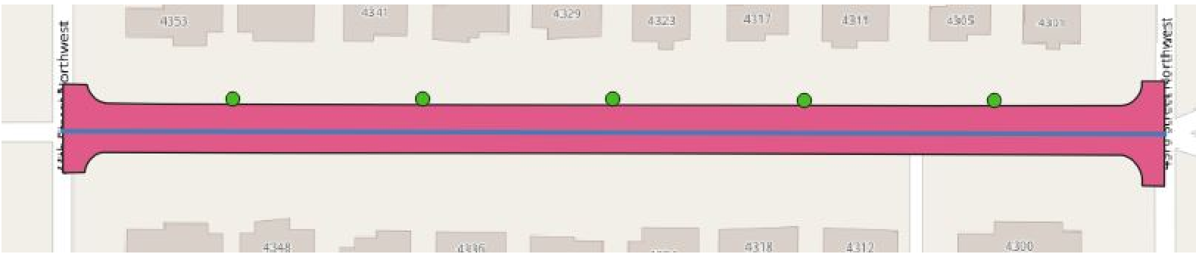

Figure 4.21: Example of a road segment with spacing variance below 100

- Value between 100 and 200 – 10.5% – Acceptable road segments with slight differences in spacing values (Fig. 4.22)

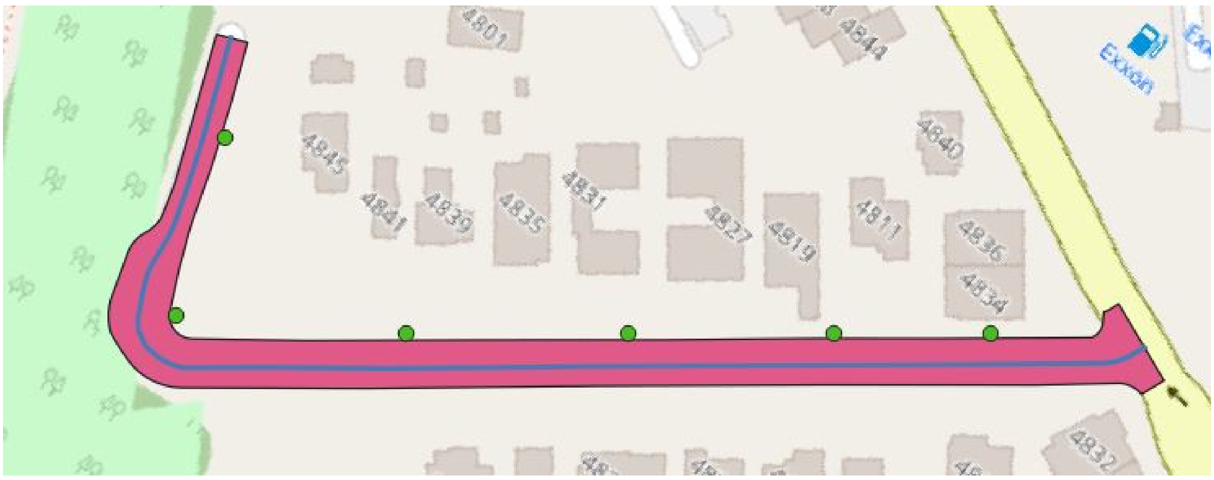

Figure 4.22: Example of a road segment with spacing variance between 100 and 200

- Value over 200 – 9.5% – Almost one-tenth of segments obtained a value of over 200, which suggests spacing issues within a segment. Therefore, records with inappropriate spacing variance have been visualized and the following reasons of excessive values have been determined:
  - Absence of a street light in a sequence (Fig. 4.23)

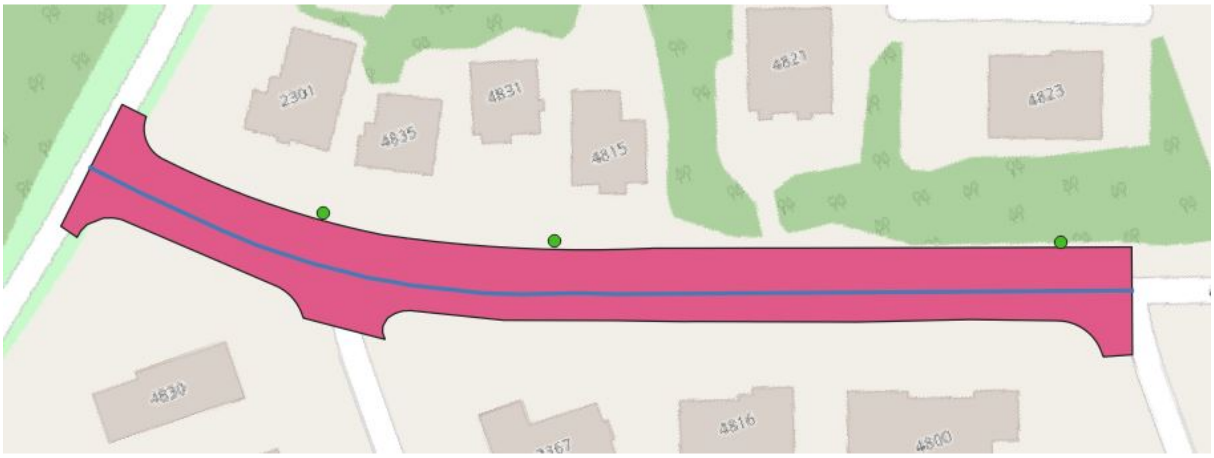

Figure 4.23: Example of a road segment with spacing uniformity disrupted by an absence of a street light in a sequence

  - Intersections disrupting spacing uniformity (Fig. 4.24)

Figure 4.24: Example of a road segment with spacing uniformity disrupted by intersections

  - Bilateral arrangement - While the first two cases are self-explanatory, this particular situation requires clarification. The distribution of spacing within segments with the perfectly opposite arrangement is divided into two groups with regular values and figures around 0, which may result in ridiculously low spacing median values (Fig. 4.25).

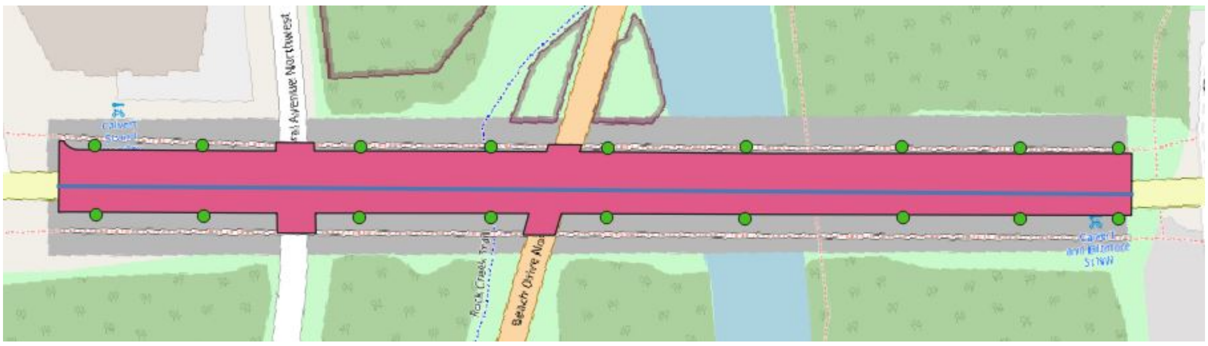

Figure 4.25: Example of bilateral arrangement

As the automatic solutions to the first two problematic segment groups will be a part of the future work, these road segments, will feature in further calculations in their intact form. However, an arrangement of street lights in a segment will be calculated (unilateral/bilateral) to resolve the last issue. By providing the information about the bilateral arrangement to the software, it will automatically split the segment into two independent lighting situations, thus improving the accuracy of the optimization. The arrangement and adjusted spacing median will be obtained by implementing the following procedure to each road segment:

- Calculate the average segment spacing and split the set of the spacing values into two subsets (over and below average),
- Empirically acquire the maximum value of spacing, under which the segment might be classified as bilateral (3 meters),
- Calculate the ratio of spacing values lower than 3 meters in the collection that holds values below average,
- If the ratio is higher than 0.5, the road segment will be considered as bilateral. The ratio level of over 0.5 has been chosen as several cases have been observed, where the arrangement of the road segment could be considered bilateral, despite locating an occasional value higher than 3 meters in the collection,
- The spacing median of all records with detected bilaterality will be adjusted to the sum of the medians of values in groups over and below average (e.g. 20.5m + 0.5m, instead of 0.5m).

By applying the aforementioned methodology, 21% of road segments with problematic spacing variance have been corrected. Finally, the median results have also shown that 1,417 records (one lamp in a segment) have received an "N/A" value, thus the dataset spacing median (29.45 meters) has been assigned to these records instead.

4.2.6 Preparation of the data for photometric optimisation

Having calculated the arrangement of street lights within a segment in the previous subsection, the remaining required data, which is still not available is the number of lanes, road classification, and the type of fixture. The first mentioned information will be acquired by dividing the road median of each segment by 3.5, which is a common industry practice. As it comes to the road classification, it is one of the attributes of the initial street lights dataset, however, the volume of the pedestrian traffic(high, medium, low marked as "H", "M" and "L" respectively) is not defined, which is required in the U.S. standard. Therefore, as other sources are not accessible, the design will be performed basing on the classification obtained from custom conversion rules presented in 4.4. The logic behind it is straightforward, i.e. the rules assume that the higher the theoretical pedestrian traffic volume of a particular road, the stricter the lighting requirements (e.g. "Local" road should have higher pedestrian traffic volume than the "Collector" road, therefore the former has been assigned a value of "H", while the latter a value of "M"). It has to be noted that "Arterial" roads have been classified as "MAJOR_A" and "Other Freeway and Expressway" value has been converted to "FREEWAY_A" as there was no possibility to differentiate between the types through other attributes.

Table 4.4: Conversion of road classification available in the dataset to the official classification values accepted by PhoCa software

Dataset Classification	Converted Classification
Local	LOCAL_H
Collector	COLLECTOR_M
Minor Arterial	MAJOR_L
Principal Arterial	MAJOR_L
Interstate	FREEWAY_A
Other Freeway and Expressway	FREEWAY_A

Last, but not least, the fixture type had to be decided between road, park, or decorative. It was assumed (and confirmed by visualization), that all records with "Posttop", "Mushroom" and "Chinatown Luminaire" values in the fixture style attribute will be assigned a fixture type of "park" (and "road" for the remaining street lights). The final results have been converted into the input format of PhoCa software and contained information about all assigned street lights (54,853 records).

4.3 Results

The quality of the data preparation process in lighting retrofit projects can be evaluated with the number of extracted lighting situations and the extend to which they reflect the reality. Generally, the more configurations can be generated by a certain approach, the more precise the design will be and its representation of reality will be more accurate. Considering the available data, the following approaches to the retrofit of lighting installation in the District of Columbia are viable:

- luminaire replacement based on the conversion chart described in 2.1.3.2,
- luminaire replacement and photometric optimization based only on the initial data,
- luminaire replacement and photometric optimization based on the processed data, which underwent the procedures developed in the thesis.

The input format of the most photometric calculation software assumes that each row represents a fixture together with the assigned road segment parameters. However, as it would be inefficient to consider each luminaire separately, they group these situations by the parameters significant for the optimization process i.e. road width, spacing, pole height, arm length etc. Therefore, the aforementioned approaches will be assessed based on the number of groups created from the available parameters.

The first approach is burdened with immense error, as there is no guarantee regarding the quality of the existing installation - it may display severe under- or over-lighting. The number of different configurations is equal to the count of unique values representing energy expenditure in the `wattage` attribute. Based on these values, a LED fixture with an appropriately lower wattage is installed.

The second approach assumes utilization of photometric calculations rather than using a conversion chart. It is based on a notably greater number of parameters, as the initial dataset (besides the wattage data) contains the information about the height of the pole, arm length and fixture type. However, there is no data related to the distance from the pole to the edge of the road, road width or pole spacing. These values have to be estimated using assumptions based on the type of the road i.e. higher pole spacing on the freeways than on the residential roads, etc.

Finally, the third approach, based on all the aforementioned parameters together with the ones acquired via the developed tool. Unlike the previous procedure, it assumes either minimum or no estimations at all (considering width and spacing). All approaches have been grouped by the available parameters and the results have been presented in Fig. 4.26.

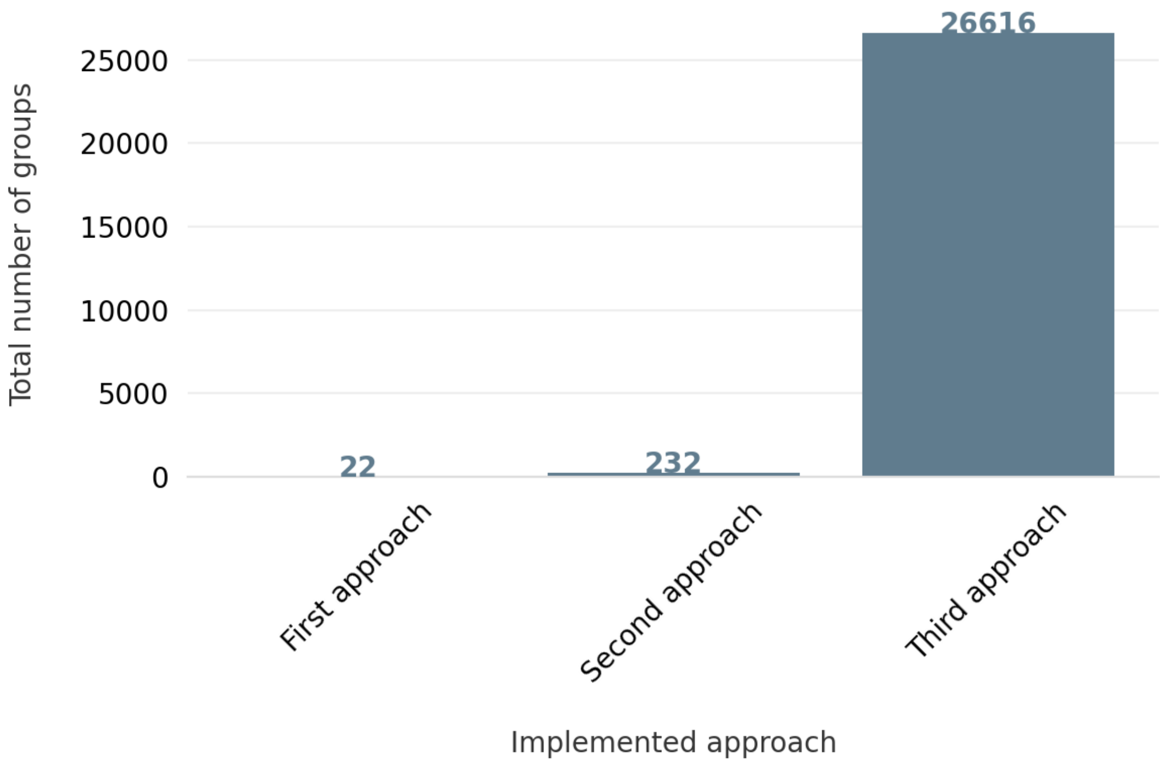

Figure 4.26: A total number of configurations depending on the implemented approach

Despite the fact that the number of distinguished groups is incomparably greater in the approach presented by this thesis, it is also based on values which are closer to the real situations. Even though the procedure developed in the thesis still assumes some estimates, they are much more accurate in comparison to the other approaches. The proposed procedure allows to calculate lighting situations more precisely and, more importantly, without the involvement of human work. The tool is not capable of performing a detailed design (at this point); however, it allows to precisely estimate the power required for lighting installation in a fully automatic manner. Not only is it reusable for other areas (assuming similar sets of data), it is also much more accurate than the other approaches.

Chapter 5

Conclusions

Road lighting design accuracy depends on correct extraction and interpretation of lighting situations (a set of parameters which take part in the photometric optimisation process) from the available data. Due to the enormous computation scale and the inability to batch process in most of industry-standard applications, the operation is currently done manually (which leads to many simplifications). Therefore, an attempt to automatically derive the lighting situations from the data concerning the District of Columbia has been performed. The presented methodology was based on the determination of spatial relationships between three sets of data, namely: street lights (which models each lighting pole as a point with additional features), street segments (which represent streets as linestrings), and roads (which refer to the shapes of the carriageways in a form of polygons). Initially, the precise road shapes have been associated with intersecting street segments and have been split in accordance with the outer points of the linestrings. The process was followed by the assignment of light points to the previously created road segments, thus developing a set of lighting situations. The calculations were carried out iteratively and the procedures were extended until satisfactory outcomes were achieved (99% of street lights assigned to the road segments). While the accuracy of the assignment cannot be evaluated precisely (as it would require manual review of more than 50,000 records), it has to be emphasized, that lighting design would be impossible to perform in a timely manner by means of the standard procedures for a project of this scale. Therefore, a thoughtful selection of ascription criteria together with careful adjustment of their corresponding weights, made it possible to obtain the most optimal results possible. These methodology assumptions minimized the ambiguity of the assignments, which was one of the successfully overcome challenges described in 3.1. The remaining problems, i.e. the identification of relevant data within the analyzed dataset and the differentiation of lamps which light up multiple areas have been also solved. With regard to the former, the relevant computation data (main carriageways in Washington, D.C.) have been obtained through analysis of attributes and visualisation of records. The latter, however, was managed by applying heuristics based on numerous attributes of street lights such as arm lengths and fixture style. Consequently, 409 records have been recognized to illuminate two distinct road segments. Ideas for improving the algorithms have been presented in the following section (5.1).

Finally, as the results described in section 4.3 suggest, the lighting design performed with the help of the developed tool yields superior outcomes in comparison with the other considered approaches. Not only are they significantly more precise, they also reflect the reality in a much more accurate way. The number of possible configurations created by means of the proposed solution was respectively 1,210 and 114 times greater than by utilization of other procedures described in 4.3. The tool allows to automatically (no human interference) estimate the possible power savings in a certain area (assuming similar data), which is especially desirable during the planning phase of the lighting retrofit projects.

5.1 Future Work

The following extensions to the algorithm will be considered in the future work:

- Optimal (power usage/cost criterion) gap filling in lamp sequences (Fig. 4.23),
- Automatic segment splitting based on varying spacing/width values (Fig. 4.24),
- Applying an additional assignment criterion (4.2.4), which would measure the level of spacing uniformity disruption along a specific segment (even greater reduction of the results ambiguity),
- Clustering of lamps, which share common characteristics (i.e. reasonably low pole height, low wattage and others). In case of the absence of information related to the asset type of lighting pole (if it lights up an alley, street or park etc.), the heuristic would assume that if most of the cluster elements have not been assigned to a segment, then those that got assigned during the process should be detached (e.g. lamp in the park assigned to unlit road outside the park).

Bibliography

[1] M. Quicheron, G. Zissis, A. d. Almeida, P. Bertoldi, European Commission, Joint Research Centre, and Institute for Energy and Transport. *Accelerating the deployment of solid state lighting (SSL) in Europe*. Publications Office, Luxembourg, 2013. OCLC: 1044369989.

[2] A. De Almeida, B. Santos, B. Paolo, and M. Quicheron. Solid state lighting review – Potential and challenges in Europe. *Renewable and Sustainable Energy Reviews*, 34:30–48, June 2014. ISSN 13640321. doi: 10.1016/j.rser.2014.02.029. URL `https://linkinghub.elsevier.com/retrieve/pii/S1364032114001506`.

[3] Solid-State Lighting Research and Development. Technical report, U.S. Department of Energy, Apr. 2014.

[4] Energy Savings Potential of Solid State Lighting in General Illumination Applications. Technical report, U.S. Department of Energy, Jan. 2012. URL `http://www.osti.gov/servlets/purl/1218257/`. Issue: 1218257.

[5] Light Pollution, . URL `https://www.darksky.org/light-pollution`.

[6] N. Group. Global Smart Street Lighting & Smart Cities: Market Forecast (2019 – 2028). Technical report, Apr. 2019.

[7] Making Outdoor Lighting More Efficient, Safe, and Affordable With Open-Standard Control Networking Technology. Technical report, Echelon, 2015. URL `http://goo.gl/mfpGYT`.

[8] A. Sdziwy and L. Kotulski. Towards Highly Energy-Efficient Roadway Lighting. *Energies*, 9(263):16, Apr. 2016. ISSN 1996-1073. doi: 10.3390/en9040263. URL `http://www.mdpi.com/1996-1073/9/4/263`.

[9] A. Sędziwy. A New Approach To Street Lighting Design. page 25, 2015.

[10] L. Kotulski, J. D. Landtsheer, S. Penninck, A. Sędziwy, and I. Wojnicki. Supporting energy efficiency optimization in lighting design process. page 12, 2013.

[11] *EN 13201-2:2015 Road lighting - Part 2: Performance requirements*. .

[12] *CEN/TR 13201-1:2014 Road lighting - Part 1: Guidelines on selection of lighting classes*. Polski Komitet Normalizacyjny, Warszawa, 2016. ISBN 978-83-275-5702-5. OCLC: 956557545.

[13] J. Lawhead. *Learning geospatial analysis with Python: understand GIS fundamentals and perform remote sensing data analysis using Python 3.7*. 2019. ISBN 978-1-78995-927-7. OCLC: 1154476837.

[14] Documentation for QGIS 3.10, . URL `https://docs.qgis.org/3.10/en/docs/`.

[15] E. Westra. *Python geospatial development: learn to build sophisticated mapping applications from scratch using Python tools for geospatial development*. Packt Pub., Birmingham, 2013. ISBN 978-1-78216-153-0 978-1-68015-362-0. URL `http://site.ebrary.com/id/10714624`. OCLC: 849917253.

[16] A. Sędziwy and M. Kozień-Woźniak. Computational Support for Optimizing Street Lighting Design. In W. Zamojski, J. Mazurkiewicz, J. Sugier, T. Walkowiak, and J. Kacprzyk, editors, *Complex Systems and Dependability*, volume 170, pages 241–255. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-30661-7 978-3-642-30662-4. doi: 10.1007/978-3-642-30662-4_16. URL `http://link.springer.com/10.1007/978-3-642-30662-4_16`. Reporter: Complex Systems and Dependability Series Title: Advances in Intelligent and Soft Computing.

[17] I. Wojnicki and L. Kotulski. Street Lighting Control, Energy Consumption Optimization. In L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Zurada, editors, *Artificial Intelligence and Soft Computing*, volume 10246, pages 357–364. Springer International Publishing, Cham, 2017. ISBN 978-3-319-59059-2 978-3-319-59060-8. doi: 10.1007/978-3-319-59060-8_32. URL `http://link.springer.com/10.1007/978-3-319-59060-8_32`. Series Title: Lecture Notes in Computer Science.

[18] A. Sedziwy, A. Basiura, and I. Wojnicki. Roadway Lighting Retrofit: Environmental and Economic Impact of Greenhouse Gases Footprint Reduction. page 11, 2018.

[19] A. Sdziwy. Control Driven Lighting Design for Large-Scale Installations. page 10, 2018.

[20] CityTouch, . URL `https://www.lighting.philips.com/main/systems/lighting-systems/citytouch`.

[21] LightGrid, . URL `https://products.gecurrent.com/control-systems/lightgrid-outdoor-wireless-control-system`.

[22] Owlet IoT, . URL `https://www.schreder.com/en/products/owlet-iot-smart-city-control-systems`.

[23] I. Wojnicki, S. Ernst, and L. Kotulski. Economic Impact of Intelligent Dynamic Control in Urban Outdoor Lighting. *Energies*, 9(5):314, Apr. 2016. ISSN 1996-1073. doi: 10.3390/en9050314. URL `http://www.mdpi.com/1996-1073/9/5/314`. Number: 5 Reporter: Energies.

[24] A. Sedziwy, L. Kotulski, and A. Basiura. Enhancing Energy Eciency of Adaptive Lighting Control. page 10.

[25] B. S. Balasubramani and I. F. Cruz. Spatial Data Integration. In S. Sakr and A. Zomaya, editors, *Encyclopedia of Big Data Technologies*, pages 1–8. Springer International Publishing, Cham, 2018. ISBN 978-3-319-63962-8. doi: 10.1007/978-3-319-63962-8_218-1. URL http://link.springer.com/10.1007/978-3-319-63962-8_218-1. Reporter: Encyclopedia of Big Data Technologies.

[26] M. Jalayer, H. Zhou, J. Gong, S. Hu, and M. Grinter. A Comprehensive Assessment of Highway Inventory Data Collection Methods. *Journal of the Transportation Research Forum*, June 2014. ISSN 1046-1469. doi: 10.5399/osu/jtrf.53.2.4219. URL http://journals-dev.library.oregonstate.edu/index.php/trforum/article/view/4219.

[27] J. Guerrero-Ibáñez, S. Zeadally, and J. Contreras-Castillo. Sensor Technologies for Intelligent Transportation Systems. *Sensors*, 18(4):1212, Apr. 2018. ISSN 1424-8220. doi: 10.3390/s18041212. URL http://www.mdpi.com/1424-8220/18/4/1212.

[28] C. R. Bennett, H. de Solminihac, and A. Chamorro. Data Collection Technologies for Road Management. (30):8, 2006.

[29] G. Leduc. Road Traffic Data: Collection Methods and Applications. page 56.

[30] Openstreetmap, . URL https://www.openstreetmap.org/.

[31] M. Bielecka, A. Bielecki, S. Ernst, and I. Wojnicki. Prediction of Traffic Intensity for Dynamic Street Lighting. pages 1149–1155, Sept. 2017. doi: 10.15439/2017F389. URL https://fedcsis.org/proceedings/2017/drp/389.html. Meeting Name: 2017 Federated Conference on Computer Science and Information Systems.

[32] *EN 13201-3:2015 Road lighting - Part 3: Calculation of performance*. Polski Komitet Normalizacyjny, Warszawa, 2016. ISBN 978-83-275-5869-5. OCLC: 1150419442.

[33] A. Sędziwy and L. Kotulski. Multi-agent System Supporting Automated GIS-based Photometric Computations. *Procedia Computer Science*, 80:824–833, 2016. ISSN 18770509. doi: 10.1016/j.procs.2016.05.372. URL https://linkinghub.elsevier.com/retrieve/pii/S187705091630847X. Reporter: Procedia Computer Science.

[34] pandas, . URL https://pandas.pydata.org/.

[35] scikit-learn, . URL https://scikit-learn.org/stable/.

[36] Matplotlib, . URL https://matplotlib.org/.

[37] Shapely, . URL https://shapely.readthedocs.io/en/latest/manual.html.

[38] GeoPandas, . URL https://geopandas.org/index.html.

[39] K. K. Lwin, R. C. Estoque, and Y. Murayama. Data Collection, Processing, and Applications for Geospatial Analysis. *Data Collection*, page 20. Reporter: Data Collection.

[40] Streetlight inventory, . URL https://solutions.arcgis.com/local-government/help/streetlight-inventory/.

[41] Levenshtein distance, . URL https://en.wikipedia.org/wiki/Levenshtein_distance.