POLITECNICO DI TORINO

Master's Degree Course in Energy
and Nuclear Engineering

Master's Degree Thesis

Assessment of rooftop solar photovoltaic potential through spatial analysis and implementation of machine learning techniques

Supervisor

Prof. Andrea Lanzini

Candidate

Francesco Petrosino

Co-supervisors

Ing. Daniele Salvatore Schiera

Ing. Riccardo Novo

Academic year 2021/2022

Abstract

The necessity to reduce greenhouse gas emissions to meet the global warming regulations has increased the demand for renewable energy sources; unlike power generation from fossil fuels, renewables have relatively low geographic density and they're unpredictable. Solar energy deployment is gaining greater attention and it's a technically and economically feasible solution as a sustainable alternative that might alleviate aspects of the current climate crisis, especially in cities, where it is crucial to promote the use of solar technologies. The roof surfaces within urban areas are constantly attracting interest as they supply huge potentials for the mitigation strategy to minimise the environmental impact by achieving the sustainable development goals.

The local generation of renewable electricity through roof-mounted photovoltaic (PV) systems on buildings in urban areas can play a significant role within the transition to a low-carbon energy system as the resulting large-scale deployment is quite straightforward once the methodology of the solar energy potential assessment has been developed. Unfortunately, some roof surfaces are unsuitable for installing photovoltaic systems, in fact one of the major challenges today is to evaluate the suitability of PV systems' installations on buildings' roofs, as they are generally lowered by superstructure.

To date, the lack of high-resolution data and also the large uncertainties related to existing processing methods impede the accurate estimation for measuring the rooftop solar energy potential over a heterogeneous urban environment containing flat and pitched roof surfaces at different slopes and directions; the idea is to rate the roof surfaces regarding their solar potential and suitability for the installation of photovoltaic systems.

In this thesis work, we will try to address these issues and therefore the gap between existing methodology and technological development through remote-sensing data and the implementation of Machine Learning (ML) algorithms to simplify the intricate topography of cities. This technique is based on aerial images that are analysed using image recognition, Geographic Information Systems to estimate the rooftop photovoltaic potential of buildings in an urban environment, the city of Aosta, Italy.

ML techniques, combined with satellite images, allow to overcome the constraints of lack of information in providing this mapping at large scale. The scalability of the trained model allows to predict the existing solar panels deployment at the Italian national scale which might be able to extract predictive models in urban areas and may be a valuable input for policymakers and for investing in distributed energy infrastructures.

The methodology, however, is generalizable to any region where similar data is available and could therefore be useful for researchers, energy service companies and municipalities to assess the rooftop PV capacity of the region.

Table of Contents

List of Figures

List of Tables

Abbreviations

SDG	Sustainable Development Goals
WEO	World Energy Outlook
GHG	Green House Gases
RES	Renewable Energy System
IoT	Internet of Things
QGIS	Quantum Geographic Information System
EU	European Union
GSE	Energy Services Manager
PAN	National Action Plan
DEM	Digital Elevation Model
DSM	Digital Surface Model
DTM	Digital Terrain Model
EP	Economic Potential
GG	Geographical Potential
GUI	Graphical user interface
NGEU	Next Generation European Union
PP	Physical Potential
PV	Photovoltaic
RES	Renewable Energy Sources
SEBE	Solar Energy on Building Envelopes

SEP	Spatial Energy Planning
TP	Technical Potential
UMEP	Urban Multiscale Environmental Predictor
WFS	Web Feature Services
WMS	Web Map Services
GSE	Energy Services Manager
EC	Energy Consumption
STC	Standard Test Condition
PR	Performance Ratio
PAN	National Action Plan
SVM	Support Vector Machines

Chapter 1

1. Introduction

1.1 Background and Problem Statement

The climate of Earth has been constantly changing throughout history. Nowadays the global average temperature is more than 0.85°C higher than it was in the last period of the 19th century. Besides that, from 1983 to 2012, it was the warmest year period of the last 1400 years in the Northern Hemisphere. This temperature increase has several and crucial consequences for the planet earth, such as the declining of mountain glaciers and hotspots of biodiversity. Most of the causes that promote climate changes in the global average surface temperature was generated by anthropogenic forcings, such as the increase of Greenhouse Gas (GHG) concentrations. The source of these anthropogenic GHG emissions is from fossil fuel combustion and industrial processes which contributed 78% of all emissions starting from 1970. Sectors of the economy responsible for most of this include electricity and heat production (25%); agriculture, forestry, and land use (24%); industry (21%); transportation including automobiles (14%); other energy production (9.6%); and buildings (6.4%). [1], [2]

Energy is one of the central topics to the achievement of the 2030 Agenda for Sustainable Development Goals (SDG). A shift towards sustainable energy production and solutions is crucial to the achievement of the Paris Agreement objectives adopted under the United Nations Framework Convention on Climate Change. [3] The energy sector has been experiencing its most complex transformation since its creation: electricity is increasingly the "fuel" of choice due to its share in global final consumption which is approaching 25% and it's destined to rise further. Political support and reduced technology costs are leading to rapid growth, but they require the whole system to change and operate differently in order to ensure a reliable, affordable and sustainable supply; these are the three most important keywords of WEO. [4]

This problem statement is multi-level and there are several approaches that should be mixed together in order to find the best compromise, as a complete solution for this problem will be so hard to find and achieve. Key factors for a sustainable and secure energy transition to cleaner renewable sources are essentially 3:

- Declining cost: as the technology is more and more widely used, it is growing, it is gaining important market positions and investments are repaid. Cost reduction is important as mature technologies, including hydroelectric and geothermal energy, have been competitive, but the sharp drop in costs has surprised even the most optimistic observers: solar and wind energy can now beat the technologies of conventional generation in many of the world's major markets.
- Public opinion: it is a powerful force for change, too. Consumers increasingly prefer to use products and services with a greater sustainable footprint, and civil movements put pressure on governments, which should introduce a market incentive programme to promote the transition from fossil gas to RES, and companies to reduce air pollution and carbon emissions.
- Technological innovation: higher efficiencies of solar photovoltaic modules and taller wind turbines, have played an important role also to create renewable hydrogen generated from electrolysis. Innovations are opening new frontiers, for example digital technologies, such as smart grids, IoT, big data, and artificial intelligence could help to raise efficiency and accelerate emerging smart generation and distribution systems.

We can compare the results creating possible scenarios and performing a sensibility analysis of the models applied that meets the criteria mentioned above, but this requires a documentation effort, as well as a simple way to provide data and detailed information. Each scientific paper and study have its strengths and weaknesses, but often overlooks aspects that cannot be quantified, such as social and environmental risks or opportunities, or reflects only a limited spectrum of possible developments. [5]

1.2 Thesis Structure

This master's thesis starts with an introduction chapter covering the background and problem statement of the work, just read. Chapter 2 gives an overview of the current territorial and energetic framework of Italy and provides a brief point of view on the case study we are going to analyse right after. The following chapter, number 3 consists of literature review based on current methodologies, and their detailed investigation, from the various potentials analysed, which really affect the research, to the most interesting output of the study, it will be very descriptive, but also very important, as the key aspects will be fully explained. Chapter 4 is the core of the thesis and describes the original and scalable methodology that was carried out for the assessment of the solar technical potential from photovoltaic sources. Firstly, the data collection, tools and software are introduced, the most important of this section and for the whole master thesis is QGIS, the geographic information system which is implemented in the methodology and deeper explained in the following subchapters. The last sections of this chapter are focused on the introduction of new techniques based on machine learning algorithms. Chapter 5 is dedicated to the application of the methodology and the spatial analysis on Aosta Valley, taken as a case study, and the numerical results are presented and discussed. In this chapter the methodology has been implemented on two different raster resolutions and comparisons between different methods are mentioned. Finally, Chapter 6 contains the conclusions of the whole thesis. The main results obtained are summarised trying to give a wider view on the problem analysed, while giving ideas and suggestions for future improvement of the methodology.

Chapter 2

2. Current Status and Research Objectives

2.1 European Commission Green Energy

Decarbonizing the EU's energy system is critical to reaching 2030 climate objectives and achieving the long-term objectives as carbon neutrality by 2050, since the production and use of energy account for more than 75% of the greenhouse gas emissions.[6] 3 key principles for the clean energy transition to reduce pollutant emissions and enhance the quality of life of the citizens can be listed as following:

1. Ensuring a secure and affordable EU energy supply
2. Developing a full integrated, interconnected, and digitalized EU energy market
3. Prioritising energy efficiency, improving the energy performance of the buildings and developing a power sector based largely on renewable sources.

The main objectives of the Commission are:

1. Building interconnected energy systems and better integrated grids to support renewable energy sources.
2. Promoting innovative technologies and modern infrastructure.
3. Boosting energy efficiency and eco design of products.
4. Decarbonising the gas sector and promoting smart integration across sectors
5. Empowering consumers and helping EU countries to tackle energy poverty
6. Promote EU energy standards and technologies at global level

The Commission adopted a set of proposals to make the EU's climate, energy, and transport cleaner also through taxation policies for reducing net greenhouse gas emissions by at least 55% by 2030 compared to 1990 levels. The main objective is to make Europe the first climate neutral continent in the world. [7]

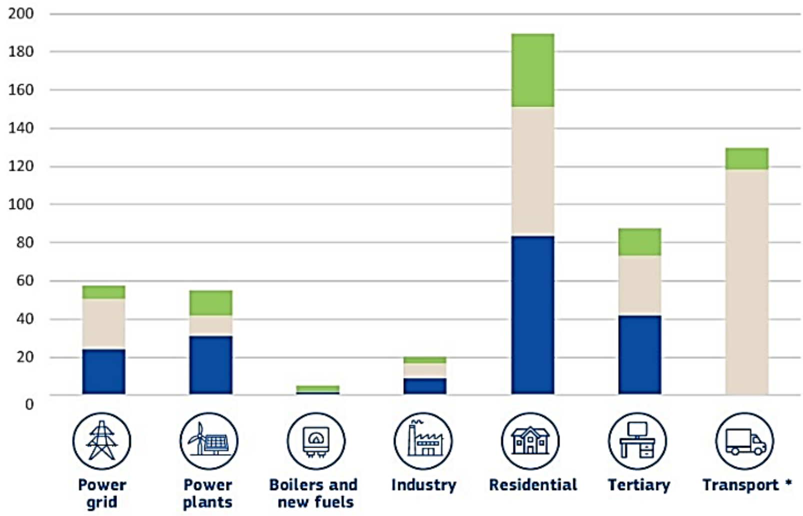

Figure 1: Yearly mean investment in two different periods in billions of euros, relatively 2011-2020, 2021-2030 and additional greenhouse reduction between 2021 and 2030

Blue color represents the historic yearly investments in the energy system from 2011 to 2020. Grey colour is used to demonstrate the additional current 2030 policies in 2021-2030 in comparison with the previous period from 2011 to 2020. Green colour demonstrates the additional greenhouse reduction between 2021 and 2030 to achieve -55%. [8]

Year and annual trajectory	2016	2017	2025	2030
Renewable production	110.5	113.1	142.9	186.8
Hydropower	42.4	46.2	49.0	49.4
Wind	17.7	17.7	31.0	41.5
Geothermal	6.3	6.2	6.9	7.1
Bioenergy	19.4	19.3	16.0	15.7
Solar	22.1	24.4	40.1	73.1
Gross inland consumption of electricity	325.0	331.8	334	339.5
RES-E share (%)	34.0%	34.1%	42.6%	55.0%

Table 1: Growth targets 2030 for the renewables share in the electricity sector (TWh)

To sum up, growth in renewable energy directly contributes to minimising the adverse effects of volatile global fossil fuel prices and exchange rate risks related to energy geographic and logistic monopoly. The newly agreed upon ambitious 2030 renewable energy and energy efficiency targets will aid in reducing the EU's reliance on imports of fossil fuels as well as its vulnerability to global shocks and uncertainty regarding the price of them. Achieving the sustainable transition will require innovation, which will be sparked by investments in energy efficiency and renewable energy. These investments will also put the EU on the path to compliance with the Paris Agreement. [2]

2.2 Italian Territorial Context

In general, in Italy, the growth of RES has been stronger than expected and final consumption has declined. As a result, in terms of % we are ahead of schedule with respect to the EU's Agenda in all the three sectors:

- *Power sector:* 107.6 TWh in 2014 vs 79.0 expected (99 in 2020)

  – The target % for the power sector is 26.4%. In 2014 Italy had already reached 33.4%, while the expected intermediate target was 21.7%.

– Renewable power targets will be revised from 26% to 32-35% (120-130 TWh)

• *Thermal sector*: 9.9 Mtoe (2014) vs 5.5 expected (10.5 in 2020)

• *Transportation sector*: 1.5 Mtoe (2014) vs 2.14 expected (3.45 in 2020)

Many of the data presented below are the result of the collection of statistical data on the production of energy from renewable sources on the Italian territory, the source is the report provided by the GSE (Energy Services Manager), divided between the Electrical and Thermal sector and Transportation. [9] The purposes reside in both ordinary statistical production and the monitoring of energy consumption targets from RES established by Directive 2009/28/EC and the National Action Plan for renewable energy, PAN.

The graphs below show the RES consumed in the electric sector under a gross power and percentage point of view, also considering the PAN trajectory of the previous strategy. Italy is performing well and even better than the dark blue line meaning that we are overlapping our objectives and there is still possibility to further improve the situation as the sharing of electricity from RES is increasing. [9]

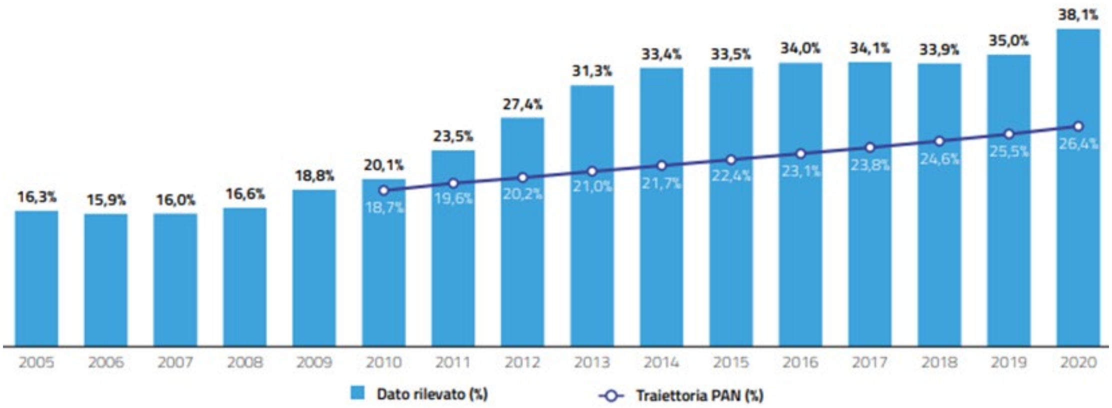

Figure 2: Share of final energy consumption in the electricity sector covered by RES in Italy (%)

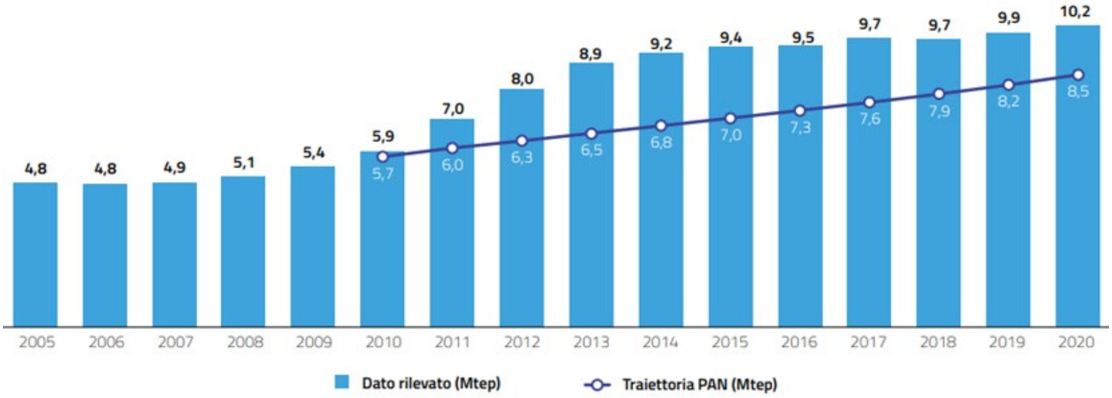

Figure 3: Share of final energy consumption in the electricity sector covered by RES in Italy (Mtoe)

The electricity production plants powered by renewable sources installed in Italy were, at the end of 2020, just under 949,000; these are mainly photovoltaic systems (98.6% of the total), which have increased by almost 56,000 units compared to 2019 (+ 6.0%). Gross efficient power of installed plants is equal to 56,586 MW, with an increase of approximately 1,091 MW compared to 2019 (+ 2.0%); this trend is mainly generated by the growth dynamics recorded in the solar (+785 MW) and wind (+192 MW) sectors. [10]

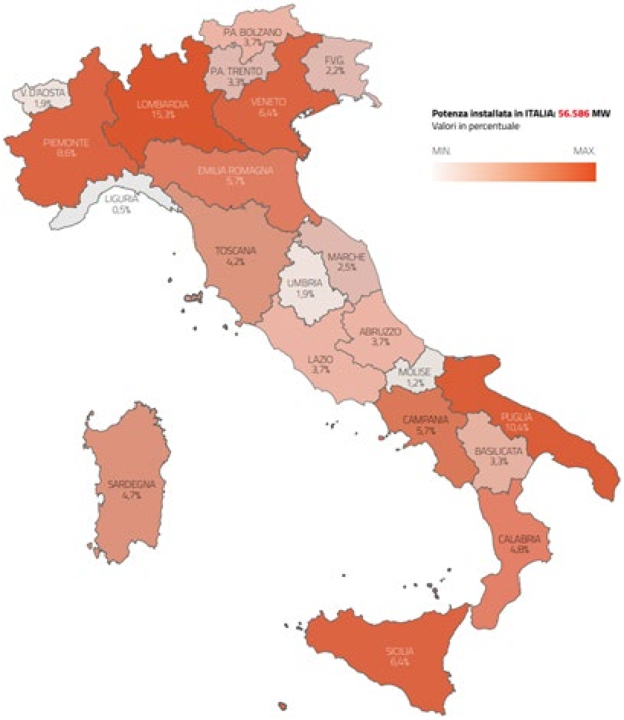

Figure 4: Number and power of photovoltaic plants in the regions

Regione	2020			2021		
	Numero impianti	Potenza installata (MW)	Produzione Lorda (GWh)	Numero impianti	Potenza installata (MW)	Produzione Lorda (GWh)
Lombardia	145.531	2.527	2.441	160.757	2.711	2.545
Veneto	133.687	2.079	2.179	147.687	2.204	2.258
Emilia Romagna	97.561	2.170	2.402	105.938	2.270	2.394
Piemonte	65.004	1.714	1.827	70.400	1.792	1.884
Lazio	62.715	1.416	1.778	67.889	1.496	1.736
Sicilia	59.824	1.487	1.911	64.464	1.542	1.902
Puglia	54.271	2.900	3.839	58.914	2.948	3.881
Toscana	48.620	866	946	52.723	908	955
Sardegna	39.690	974	1.155	41.831	1.001	1.166
Campania	37.208	877	981	40.293	924	952
Friuli Venezia Giulia	37.168	561	600	39.698	591	609
Marche	30.953	1.118	1.351	33.262	1.150	1.314
Calabria	27.386	552	681	29.476	573	661
Abruzzo	22.512	755	945	24.200	774	910
Umbria	20.809	499	562	22.144	513	551
Provincia Autonoma di Trento	17.946	197	203	19.271	207	201
Liguria	10.126	119	117	10.846	127	122
Basilicata	8.894	378	491	9.456	388	477
Provincia Autonoma di Bolzano	8.871	257	272	9.349	268	271
Molise	4.470	178	231	4.726	181	221
Valle D'Aosta	2.592	25	28	2.759	26	28
ITALIA	**935.838**	**21.650**	**24.942**	**1.016.083**	**22.594**	**25.039**

Table 2: total number of solar plants installed, the installed power and the gross production of electricity, sorted by region.

The 935,838 photovoltaic systems installed in Italy as of 31 December 2020 correspond to a power equal to 21,650 MW. In the solar year 2020 alone, just over 58,000 systems were installed, for an installed capacity total of 749 MW2; 27% have a power lower than or equal to 3 kW, 68% between 3 kW and 20 kW, the remaining 4% greater than 20 kW. Compared to 2019, the plants that entered into operation in the calendar year decreased by 4.5%; the variation of the installed power, on the other hand, is negligible.

In 2020, Lombardy is the region with the highest concentration of installed power of RES plants for electricity production (15.3% of the total power nationwide); among the northern regions, followed by Piedmont (8.6%) and Veneto (6.4%).

Aosta Valley is instead one of the regions with lowest concentration of installed power and energy generation, this is why we want to focus our dissertation on this region as it has one of the highest energy potentials of development in order to reach the Italian sustainable development goal. [11]

An effective parameter to measure the production performance of a plant is analysing the equivalent hours of use, obtained from the ratio of the gross production generated in a given year and the gross efficient installed power: in the north of Italy this coefficient is about 1300 hours, gross of the conversion yield of the whole system.

A similar indicator is the capacity factor, which is obtained by dividing the production generated in a year by the production that the plant could have generated if it had operated continuously at full power, which can also be calculated as the ratio between the equivalent hours of use and hours of the year (8,760). Anyway, we should keep in mind that photovoltaic plants are more conditioned by exogenous factors, mainly of a climatic nature and that's why these factors aren't increasing by the years.

Considering the numbers of photovoltaic systems installed in Italy, approximately 92% of the plants have a power of less than 20 kW, while 35% of the installed power is concentrated in plants of size between 200 kW and 1 MW. Overall, the power of photovoltaic systems represents 38% of that relating to the entire national renewable plant park.

From the graph below we can also understand the different typology of installation considering the place where they are installed, distinguishing the ones put on the ground from the ones mounted on other structures such as rooftops mainly.

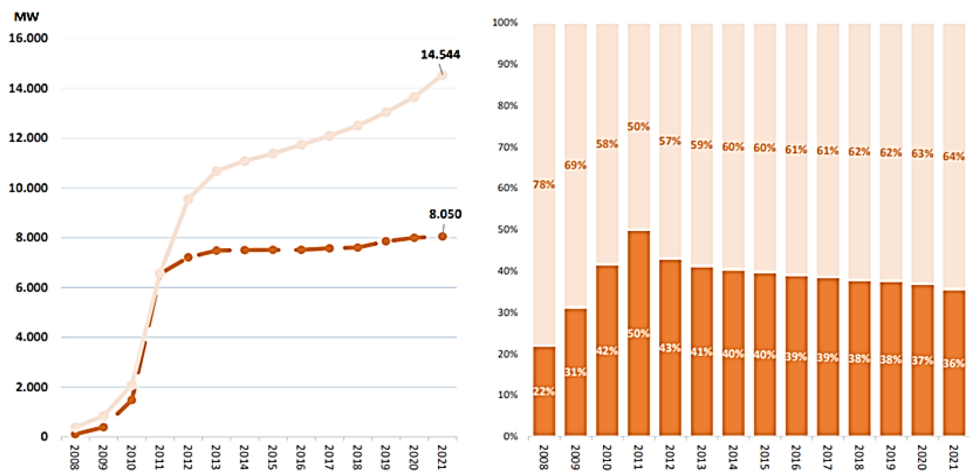

Figure 5: Power of PV installed on the ground and non-ground

2.3 Aosta Valley Case Study

Reading the above-mentioned analysis, it can be clear why we decided to focus our attention on the city of Aosta, in the middle of the Aosta Valley, located in a valley completely surrounded by mountains.

From a climactic point of view Aosta is this determines a more continental climate than tending to the alpine one. Summers are short, but still very hot. During the winter peaks of 9-10 °C are reached, in February temperatures drop rather rigidly, often remain below zero, but in broad daylight, especially at the end of the month, environments easily reach 12-13 °C. The monthly record for February dates back to 2017, when it reached 19°C, however the maximum temperatures, especially in the presence of strong sun with African anticyclone, rose rapidly upwards, stagnating in the basin and reaching 25-26 °C. In June and July, the peak of heat is reached, as well as the peak of precipitation and heat storms. They are the only months in which mild values are recorded in the morning, around 15°C, and very high maximum values, with peaks usually around 32-35 °C. Aosta's absolute record of heat belongs to June 27, 2019, when the capital reached 40 °C. At the end of August, temperatures drop sharply and the last thunderstorms are observed. In September, normal values and daily maximum values in the morning are 30 ° C, while In October, the climate cools down considerably, with the first values below zero in the middle of the month, and the maximum diurnal values struggle to exceed 20 °C. In November and December, the minimum values are often below zero, even abundantly, and snowfalls are very frequent and intense.

This meteorological overview was meant to explain that even if the region is located in the northeast part of Italy, it is somehow very sunny and there are high limits to overstep in order to achieve better results in terms of generation and sharing of solar renewable energy.

These tables below explain well the bad situation on energy production and number of PV plants installed in the Aosta city, that’s why it is at the end on the Italian list by region; also, the variation between 2020 and 2021 does not bode well in terms of technological progress. [11]

	Production (GWh)		Incidence on the national total (%)		*Variation % of the production*
	2020	2021	2020	2021	*2021/2020*
Aosta Valley	27,8	27,9	0,1	0,1	*0,3*
Italy	24.941,5	25.039,0	100,0	100,0	*0,4*

Table 4: Gross production of photovoltaic systems installed in Aosta Valley compared to Italy

	2020				2021				*% 21 / 20*	
	n°	%	MW	%	n°	%	MW	%	n°	MW
Aosta Valley	2.592	0,3	25,4	0,1	2.759	0,3	26,4	0,1	*6,4*	*4,1*
Italy	935.838	100,0	21.650,0	100,0	1.016.083	100,0	22.594,3	100,0	*8,6*	*4,4*

Table 3: Number and power of PV systems installed in Aosta Valley compared to Italy

With the help of the Energy Centre, a research centre, located in Turin and belonging to the Polytechnic (based on the themes of energy, sustainability and innovation with the aim of launching a series of actions and projects that will provide support and strategic advice to local authorities, national and transnational bodies, on energy policies and technologies to be adopted) we managed to go in details about Aosta energy situation and to understand how to improve their technical potential related to the solar energy production; working hours are taken as 992 h/year.

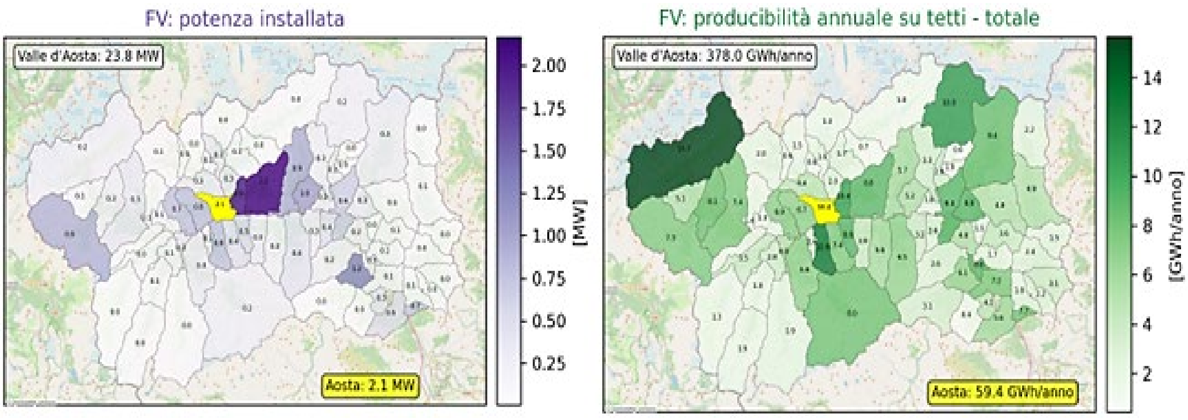

Figure 6: Installed power and annual producibility on the Aosta roofs

Chapter 3

3. Literature Review

This following chapter briefly investigates the literature theory of this thesis going through the various methodologies available to determine the urban rooftop PV potential and finding the best available methods to improve the existing ones for future research. This chapter holds 3 sections as follows: section 3.1 introduces various affecting factors and most shared commons methodologies to investigate our important studied factors such as physical, geographic, technical and economic potential. Section 3.2 investigates different methodologies for observing the rooftop PV potential, section 3.3 demonstrates several information of methodologies to obtain a better outlook for future work.

3.1 Affecting Factors and Data Sources

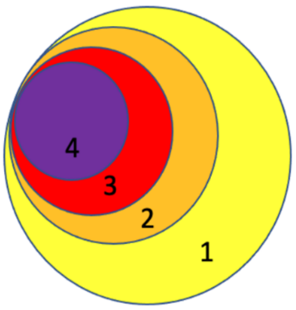

Figure 7: Hierarchical distribution of physical, geometric, technical, and economic potentials

A hierarchical, justifiable methodology is considered as the most reasonable means of evaluating the potential of renewable energies. [12]

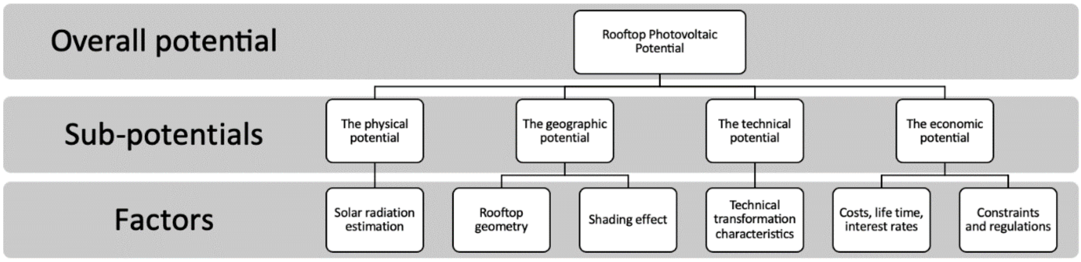

Figure 8: Rooftop photovoltaic potential essential factors [14]

So, as we can see the scheletric hierarchical categorization of our affecting factors is quite common and well shared among the various different scientific papers that have been analyzed. Even if there are some differences in plotting them as circles within each other or in a more parallel way, the concept to develop the methodology remains the same for each of them.

3.1.1 Physical potential

Physical potential can be defined as the solar energy irradiated to the surface of the earth on a year basis. The radiative flux incident at a specific location depends on the time of a day year due to the rotation of Earth and on the geographical position on earth which are called longitude and latitude. The radiative flux is reduced upon traversing the atmosphere towards the surface due to reflection, scattering and absorption of radiation in the atmosphere. The fraction of incoming radiation reflected back into space is called the albedo of the earth atmosphere system. [12]

It is symptomatic of the maximum energy limit in the resource contemplated. In the varying applications, this potential can be evaluated in different ways. [13] In Izquerdo's paper it is assessed as the horizontal irradiation, which is calculated using log-standing procedures as follows: computing the monthly extra-terrestrial radiation entrenched the geometry of the sun- earth system, computation of monthly clearness index for the locations with hourly meteorological data, creating monthly irradiation maps, investigating the effect of hourly shadows on monthly values taken into account with geometrical calculations with digital terrain model. [14]

3.1.2 Geographical potential

Geographical potential is assessed as the influence of the human-made environment and the location limitations. The geographical potential of the renewable resources is achieved by eliminating the reserved zones like roads, beaches, lakes, and rivers, just as preserved areas such as national parks. [15] The suitable portion of theoretical potential that can be used because the land or location is suitable and readily accessible. The geographical potential will only comprise roof surfaces suitable for solar installations because only photovoltaic plants on roofs are taken into account. [16]

3.1.3 Technical potential

Technical potential is the usable amount of energy under technical considerations within a specified location and time. It is also convenient to describe the technical potential as an irradiation that is technically usable taking also into account the efficiency of photovoltaic modules. [16] In this case technical potential is the amount of energy generated by photovoltaic panels installed on building roofs in the available area in Aosta Valley, Piedmont, Italy, whilst considering shading effects.

3.1.4 Economical potential

Economic potential is the proportion of the technical potential that can be utilised economically. It considers costs and socioeconomic factors such as fuel and electricity prices, other opportunity costs, and land prices. After determining the system size and business model, electricity generation from the rooftop solar PV system should be calculated. Later on, the profitability of the rooftop solar PV and eventually the economic potential of the solar PV system can be estimated. [17] Economic and implementation potentials will not be included in this thesis.

3.2 Current Methodologies

Analysis of the correlation between the generation of electricity and the width of the study area of the literature review. The idea of this chapter is to individually analyse every single paper in order to deeply describe our literally review, it will be quite lengthy and verbose but it's the only way to fully understand which methodologies have been studied since now and which ones of them could potentially be more investigated and developed to extract the real objective of this thesis project. Below, a very interesting graph is presented: it puts in relation the width of the studied area extension, generally called geo-scale (which will be explained in detail later) and the annual production of the electricity in that particular area. The aim is to show that this correlation follows a similar pattern through all the papers but there is not really a proportionality affecting the two variables, in fact there are a lot of factors affecting the solar potential calculation, such as the solar irradiance on the ground (physical potential) and the rooftop area suitability geographical potential). Our scope is to investigate why this happens and how to optimise the methodology in order to obtain a result as accurate as possible, and of course scalable.

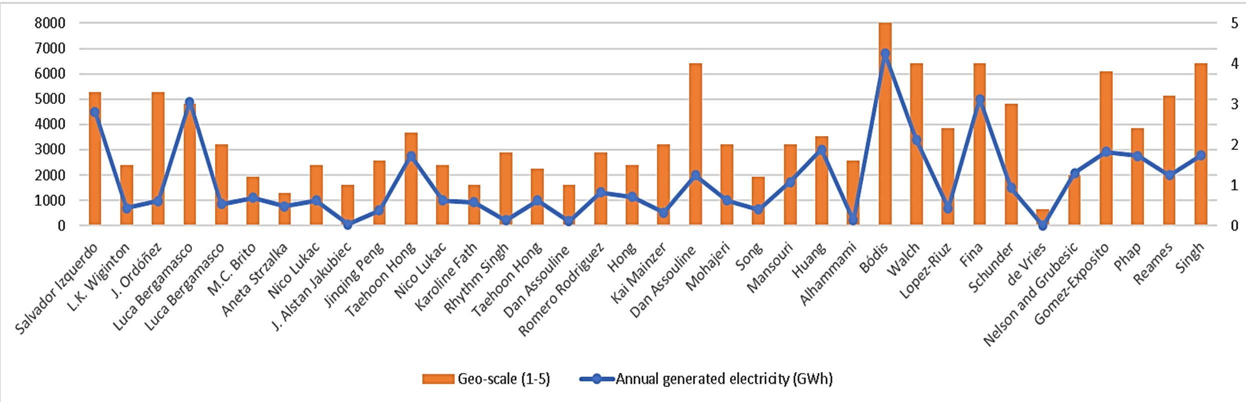

Figure 9: Analysis of the correlation between the generation of electricity and the width of the study area of the literature review

The 36 scientific papers are ordered by year of publication:

1. In April 2008, urban areas of Spain were investigated based on a bottom-up approach in order to estimate the technical potential of rooftop photovoltaic systems. Research was performed based on statistically representative GIS vector maps with land uses' and building densities' data. Statistical construction data and Google Earth digital urban maps that are

exported and scaled with AutoCAD are used to obtain the energy capacity of photovoltaics on rooftops. Correction factor was used to eliminate the errors that might have been caused by building type, location, shading effect, tilt angle of roof and orientation. [14]

2. In January 2010, residential buildings in Andalusia (Spain) were analysed by a top-down approach that was based on statistical sampling. Urban satellite maps that were obtained from Google Earth and statistical data were used as inputs to be scaled on AutoCAD. [18]

3. In January 2010, rooftop solar photovoltaic potential of residential buildings in Ontario (Canada) was investigated by a bottom-up approach consisting of five essential steps to obtain roof areas "Feature Analysis" extraction tool in ArcGIS was used. Later, in order to increase the accuracy, shading and orientation were considered. Finally, by utilising market prices, an economic assessment was performed. [19]

4. In February 2011, Piedmont region was observed as a case study to obtain the photovoltaic solar potential by using hierarchical methodology. Numerical technical regional maps were used for geographical cadastral analysis. Maps were analysed by ArcGIS and processed in MATLAB. Shadowing effect was considered. [20]

5. In September 2011, a city level top-down approach was considered for the study based on ortho image analysis. Aerial imagery, 3D models and satellite maps were used for the analysis and it was performed on Matlab; it's the deeper evaluation of the previous scientific paper, made by the same authors. [21]

6. In November 2011, Photovoltaic potential assessment in Lisbon was obtained by a top-down method using LiDAR data to create DTM and DSM, population distribution, and solar radiation modelling by ArcGIS solar analyst extension tool. [22]

7. In January 2012, a bottom-up method in combination with extraction algorithms on PV system simulations was assigned to determine rooftop

photovoltaic potential for family houses in Stuttgart. Geoinformation systems, 3D models, and LiDAR data were used as inputs for the method. [23]

8. In September 2012, a bottom-up method that was a combination LiDAR data, pyranometer measurements of solar irradiances, a Heuristic vegetation shadowing and multi-resolution shadowing model was developed to obtain solar potential and sustainability of photovoltaic systems installation in Maribor. [24]

9.In May 2013, a bottom-up method was applied in order to obtain the PV potential for hotel and commercial buildings' rooftops in Hong Kong. LiDAR and statistical data were used as inputs. [25]

10. In May 2013, to predict photovoltaic potential for a bottom-up method was created by using the combination of 3D models, GIS, and LiDAR with Dayism irradiation simulation engine, rooftop temperature and meteorological climate data in Cambridge, USA. The result can be combined by online mapping and financial modules to interest the potential building owners for installing photovoltaic panels on buildings. [26]

11. In March 2015, a city level bottom-up method was generated to estimate the generated electricity of photovoltaic systems in South Korea. The method also contained a sensitivity analysis of impact factors and resulted differently according to varying regional factors, slope, and azimuth of the installed panel. Satellite maps and statistical data were used as inputs. [27]

12. In March 2015, a bottom-up method that was to estimate nonlinear photovoltaics potential was proposed in Maribor. This method included the effects of topography, vegetation and shadowing that are crucial for the accuracy of the case study and they were obtained by comparing the constant and nonlinear efficiency characteristics of solar photovoltaic inverters and photovoltaic module types. LiDAR and 3D models were used as inputs. [28]

13. In June 2016, Photovoltaic potential of Mumbai (India) was calculated by using a top-down methodology of GIS image analysis, high-granularity land public data and sunshine simulation in PVSyst. The same methodology was also applied to the mathematical models and micro-level

simulations in the PVSyst. Aerial imagery, Satellite maps and statistical data were used as inputs. [29]

14. In November 2016, a top-down approach was applied on urban areas also considering the facades. LiDAR, 3D models and statistical data were used to perform the analysis. [30]

15. In April 2017, a bottom-up method was developed after the observation of the high impact of shadowing factors to estimate the physical and technical potential in South Korea. The method was created by using Hillshade analysis, and the main objective of this method was to analyse building shadowing. Satellite maps and statistical data were used to complete the analysis. [17]

16. In May 2017, a bottom-up method was applied for urban areas within communes that was based on support vector machines (SVM). Aerial imagery, LiDAR, and 3D models were used as inputs. [31]

17. In September 2017, CityGML geography description and 3D models for simulations, analyses, and visualisation on SimStadt platform was used as a top-down methodology for regional and urban scale of photovoltaic potential in Ludwigsburg. 3D models and statistical data were used as inputs. [14]

18. In June 2017, a big district level bottom-up approach was applied based on using Hillshade tool on ArcGIS. 3D models and statistical data were used as inputs. [32]

19. In July 2017, a city level top-down approach was held based on a methodology consisting of image recognition and machine learning. Aerial imagery, satellite and statistical data were considered as inputs. [16]

20. In September 2017, a country level bottom-up approach by random forest model was considered for the research. Methodology can be summarised as: collect the data related to variables, train a random forest model based on the data, and use the model to predict the variables in unknown locations. High resolution aerial imagery and statistical data were used as inputs. [33]

21. In June 2018, a city level bottom-up approach was considered in the research including feature selection, scaling of the dataset, data labelling,

support vector machine (SVM) testing and training. LiDAR, 3D model and statistical data were conducted as inputs for the analysis.[34]

22. In November 2018, using remote-sensing data, the intricate topography of a city level geographic area was modelled as a digital elevation model. Later, a bottom-up approach was applied, and the file of roof segments was created, and then run through a dissolve function which merges contiguous polygons with a specific common characteristic to produce continuous suitable areas. Next, the rooftop polygon file was converted to a raster file and reclassified. [35]

23 In November 2018, simulation of the monthly and annual solar radiation on rooftops at an hourly time step to estimate the solar PV potential was performed for a small-town level. This is a bottom-up approach based on rooftop feature retrieval from remote sensing images. 2D rooftop outlines and 3D rooftop parameters retrieved from high-resolution remote sensing image data. [36]

24. In April 2019, a bottom-up approach on a city level was evaluated in order to calculate the PV potential. The total available area of building roofs has to be obtained, then the annual radiation available per unit area, which is based on the area found in the first step and the weather conditions and finally U-Net deep learning method was applied by using 3D models and satellite images as inputs. [37]

25. In June 2019, a bottom-up approach was applied on a city level geographic area. Number of buildings, their types and rooftop area were calculated by using Esri ArcGIS software to estimate rooftop solar system energy potential and economic performance of Khalifa and Zayed (Abu Dhabi). [38]

26. In July 2019, a continent level top-down approach was considered for the research. The methodology is the combination of satellite-based and statistical data sources with machine learning to provide a reliable assessment of the technical potential for rooftop PV electricity production with a spatial resolution of 100m across the European Union. [39]

27. In December 2019, a country level bottom-up approach that was the combination of Machine Learning, GIS processing and physical models for the

treatment of large spatial-temporal datasets was considered for the research. Uncertainties were quantified from the statistical distribution of the variables involved in the potential estimation in the form of standard deviations. LiDAR, 3D models and statistical data were used for the analysis. [15]

28. In March 2020, a city level top-down approach that transforms night light intensity captured by the satellite radiometry into electricity consumption was considered for the research. The average electricity consumption per household in different clusters was obtained at the end of the process. [40]

29. In June 2020, country level top-down approach was performed in order to estimate a cost optimal large scale economical potential. Statistical data was used for the analysis. [41]

30. In July 2020, Remote sensing, LiDAR, footprint, and Google's Project Sunroof data was used to detect the residential rooftop solar potential in Erie Country, USA. LiDAR data was used to obtain Digital Terrain Model, Digital Surface Model, and normalized Digital Surface Model. The work has ended up with a result of the low-income population having relatively low access to rooftop solar. [42]

31. In September 2020, the Quick-scan method was used to estimate the solar potential in Eindhoven. Methodology consists of reconstructing virtual 3D roof segments and developing a fitting algorithm for photovoltaic modules on rooftops. A comparison of rooftop solar energy potential estimation by UAS and LiDAR data was carried out. ArcGIS solar analysis toolbox was used to determine the rooftop solar radiation. [43]

32. In October 2020, technical potential of rooftop photovoltaics to the future electricity mix in Spain was obtained by ArcGIS and the national geographical database, considered five sustainable scenarios each comprising different shares of centralized renewables, rooftop PV and storage. [44]

33. In October 2020, one quarter of the city was considered as a geographical area for this research that was using a methodology classified as bottom up. Digital elevation model through extremely high-resolution UAS data was obtained. Edge detection and object delineation was performed in order to obtain viable and optimal roof extraction and energy calculation eventually. [45]

34. In November 2020, four different cities in the USA were considered as a geographical area for the research. Bottom-up approach that was merging national datasets to estimate rooftop solar potential was applied. The distribution of rooftop solar systems, and census tract-level socioeconomic and demographic characteristics. 3D models, satellite images and statistical data were considered as inputs. [46]

35. In November 2020, a city level top-down approach that contained artificial intelligence was performed. Remote sensing technology, GIS, high-resolution satellite image analysis technology was used in analysis. [47]

36. In December 2020, 13 cities were considered for the study that was using a top-down approach. The analysis was performed based on a solar-city plan for Mumbai city modelled using mathematical models. [48]

From this deep analysis should be noticed that machine learning techniques are used more and more frequently to support geospatial environmental data modelling and to estimate geometries for solar prediction on building rooftops. It should also be mentioned that the average complexity brought by the machine learning in these studies did not significantly increase the total computational time of the methodology, while improving the general accuracy of the estimations.

Author	Year	Technique	Data	Tools and submethods
Dan Assouline	2016	SVM	LiDAR	Scikit-Learn
Mainzer	2017	CNN	Aerial satellite data	Image recognition
Dan Assouline	2018	RF-Ensemble	DEM	Scikit-Learn
Mohajeri	2018	SVM	LiDAR	Kernel
Bodis	2019	SVM	Satellite earth observation data(ESM)	Scikit-Learn
Huang	2019	CNN	Satellite image	Pytorch library
Walch	2020	RF-Ensemble	Satellite data	Scikit-Learn
Reames	2020	CNN	Satellite image	DeepSolar
Phap	2020	CNN	Satellite image	Pytorch library

Table 5: Different ML methods found in the literal review

The most used techniques are:

SVM-Support Vector Machines: are models of supervised learning algorithms for regression and classification. Given a set of examples for training, an algorithm builds a model that assigns the new examples to one of two different classes, obtaining a non-probabilistic binary linear classifier. Representation of the examples as points in space, mapped in such a way that the examples belonging to the two different categories are clearly separated by as large a space as possible.

CNN-Convolutional Neural Network: they are a fundamental tool in the field of deep learning. In particular they are suitable for image recognition. You can use it to train a network, and use it later to get a categorical or numeric label. It is also possible to extract features from a previously trained network, and use them to train a linear classifier.

RF-Random Forests: is one of the most used supervised algorithms due to its accuracy, simplicity and flexibility. The fact that it can be used for classification and regression tasks, coupled with its non-linear nature, makes it highly adaptable to a range of data and situations.

3.3 Comparison of Different Methodologies

Classification				Model		Methodology				Input data					
#	Author	Geographic area covered	Geo-scale	Top down/Bottom u	White-Grey-Black Box	Method	Tools	ML	GIS	Aerial	LiDAR	3D model	Satellite	Statistical	Year
1	Singh	13 cities, 79M people	4	Top down	White-Box	Based on a	Micro-level	no				+		+	2020/12
2	Phap	Capital city (30 districts)	2,4	Top down	Black-Box	Artificial	This study	yes	+				+	+	2020/11
3	Reames	4 metropolies	3,2	Bottom up	Black-Box	Merging	GIS software,	yes	+			+	+	+	2020/11
4	Nelson and Grubesic	North part of the city	1,25	Bottom up	Grey-Box	Digital	The LiDAR	no	+	+	+	+			2020/10
5	Gomez-Exposito	Andalucía, Pais Vasco and Madrid	3,8	Top down	Grey-Box	Interpolation	Matlab -	no	+					+	2020/10
6	de Vries	Several buildings	0,4	Bottom up	White-Box	Quick-scan	Fitting	no	+	+	+				2020/09
7	Schunder	City	3	Bottom up	Grey-Box	Normalized	Google's	no			+			+	2020/07
8	Fina	Country	4	Top down	White-Box	Cost-optimal	Building	no	+					+	2020/06
9	Lopez-Riuz	Capital city	2,4	Top down	Grey-Box	Transforms	Spatially	no					+	+	2020/03
10	Walch	Country	4	Bottom up	White-Box	Combinantion	Data mining -	yes	+		+	+		+	2019/12
11	Bódis	European Union	5	Top down	Black-Box	Combination	Eurostat, EU	yes	+				+	+	2019/07
12	Alhammami	City	1,6	Bottom up	White-Box	New policy foı	GIS-based	no	+					+	2019/06
13	Huang	City	2,2	Bottom up	Black-Box	The total	U-Net - deep	yes	+			+	+		2019/04
14	Song	Small town	1,2	Bottom up	White-Box	Simulation of	Segmentatio	no	+			+	+		2018/11
15	Mansouri	City	2	Bottom up	Grey-Box	Using remote-	Multi-criteria	no	+	+	+			+	2018/11
16	Mohajeri	City	2	Bottom up	Grey-Box	Feature	Support	yes	+		+	+		+	2018/06
17	Dan Assouline	Country scale	4	Bottom up	Grey-Box	Machine	GIS solar	yes	+		+			+	2017/09
18	Kai Mainzer	City level	2	Top down	Grey-Box	Image	GIS	yes	+	+			+	+	2017/07
19	Hong	Big district	1,5	Bottom up	White-Box	Deep building	Hillshade tool	no	+			+		+	2017/05
20	Romero Rodriguez	Federal State	1,8	Top down	White-Box	Urban level	Java-based	no	+			+		+	2017/04
21	Dan Assouline	Urban areas within communes	1	Bottom up	Grey-Box	Supervised	Support	yes	+	+	+	+			2016/11
22	Taehoon Hong	Urban areas including building	1,4	Bottom up	White-Box	Hillshade	ArcGIS	no	+				+	+	2016/06
23	Karoline Fath	Urban areas including building	1	Top down	Grey-Box	Simulating	ArcGIS Solar	no	+		+	+		+	2015/03
24	Rhythm Singh	Buildings within urban areas	1,8	Top down	Grey-Box	GIS image	QGIS	no	+	+			+	+	2015/03
25	Nico Lukac	Large urban area	1,5	Bottom up	White-Box	nonlinear		no			+	+			2014/01
26	Taehoon Hong	City level(due to the location	2,3	Bottom up	Grey-Box	GIS-based	RETScreen(f	no	+				+	+	2013/12
27	J. Alstan Jakubiec	City level	1	Bottom up	Grey-Box	3D model-	Daysim	no	+		+	+	+		2013/05
28	Jinqing Peng	Hotel and commercial	1,6	Bottom up	White-Box	estimate	FORTRAN	no			+			+	2013/05
29	Nico Lukac	Urban area- city level	1,5	Bottom up	White-Box	Rating	Flood-fill	no			+	+			2012/09
30	Aneta Strzalka	Multi-family houses within a	0,8	Bottom up	White-Box	modular	Raster	no	+		+	+		+	2012/01
31	M.C. Brito	Urban region	1,2	Top down	Grey-Box	Building of a	Solar Analyst	no	+	+	+	+			2011/11
32	Luca Bergamasco	The whole city of Turin- both	2	Top down	Grey-Box	Ortho-image	MATLAB,	no	+	+		+	+		2011/09
33	Luca Bergamasco	Piedmont Region	3	Bottom up	White-Box	Digital	MATLAB, ESI	no	+			+		+	2011/02
34	L.K. Wiginton	Residential buildings within urban	1,5	Bottom up	White-Box	Image	ArcGIS,Feat	no	+	+	+			+	2010/01
35	J. Ordóñez	Residential buildings	3,3	Top down	White-Box	Statistical	AutoCAD	no					+	+	2010/01
36	Salvador Izquerdo	Buildings within urban areas	3,3	Bottom up	White-Box	Statistical	Representati	no	+	+	+				2008/04

Table 6: : Comparison of the different methodologies analysed during the literal review

As we can visualise from the table below, we decided to analyse many different scientific papers coming from all over the world to have a wide look on which methodologies are used to investigate the solar energy photovoltaic potential. We decided to merge them from different points of view, the most interesting one is the classification of the geographic area covered by each of them on a scale from 1 to 5 in order to understand if they the analysed area consists in a neighbourhood (1), an entire district (2), city level (3), country scale(4) or at an international level (5); between this different limits there are some others geographical levels like small towns, big districts, or urban areas, anyway the analysis is always based on the km^2 of covered area.

The second interesting factor most affecting our analysis is the model which is used case by case on the different 36 scientific papers. More in detail we decided to classify them if they follow a top-down or bottom-up approach: they are information processing and knowledge management strategies used to analyse problematic situations and construct adequate hypotheses for the required solution.

The third factor analysed is the black, grey or white box. In this case our aim is to classify our literal review on the basis of the level of accessible knowledge and the access to the information contained within each single scientific paper. Through this approach it is possible to understand the potential level of data sharing, of the constraints and in general of the methodology used to develop the guidelines that lead to the solution of the problem in order to understand the efficiency of the of the open development

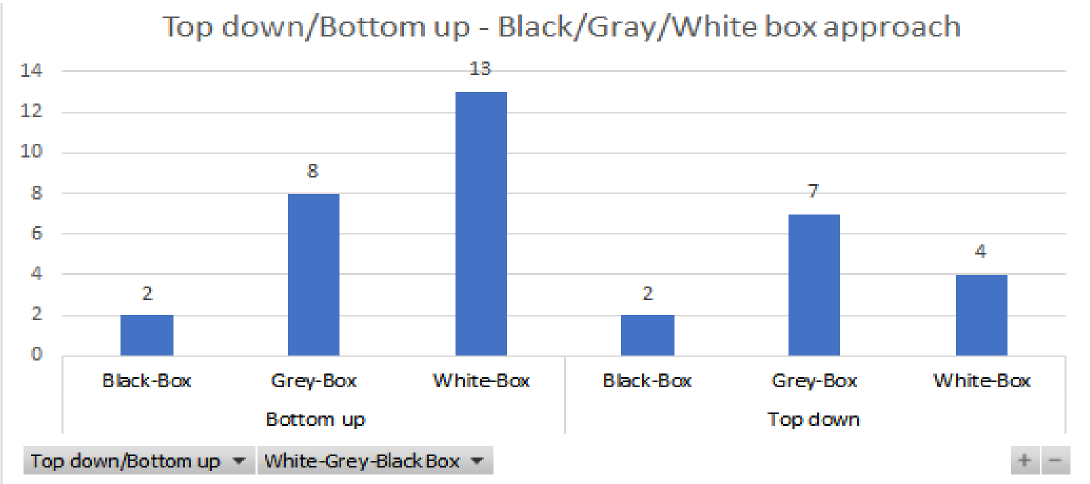

Figure 10: Pivot chart that relates the number of top-bottom method to the box approach type

In the end we wanted to normalise our classification from a more quantitative point of view and that's why an error band about the methodology was created. We tried to understand from every single paper the reason why there were uncertainties affecting the accuracy of the results and the reliability of the algorithm and tools discussed, providing a confidence interval on the quality of the analysis. Almost all of these showed inaccuracies mainly between the proposed method of analysis and real valuations using physical instruments available on the market.

These interesting factors are analysed in more detail in the following subchapter.

3.3.1 Top-down and Bottom-up approaches

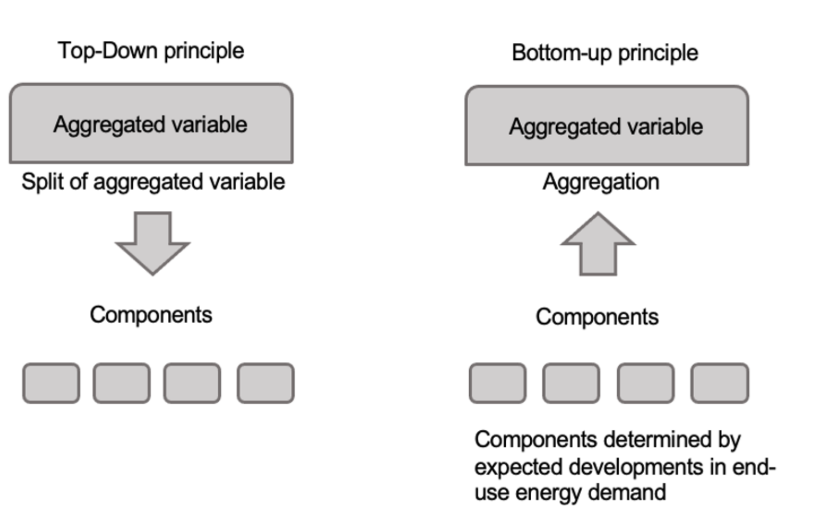

Figure 11: Aggregation level for relations determining energy demand

The main difference between top-down and bottom-up approaches can be simply explained by the direction they follow. Top-down approach goes from general to specific, and bottom-up starts from specific variables and ends up to the general. This leads two approaches to differ also in size of the project. Since the top-down approach surrounds a wide area and bottom-up is more narrowly focused. In other words, the top-down approach seeks the big picture with all variables included and bottom-up focuses on specific analysis locally and individually and combines them all to have a global result.

In the articles that were investigated during literature review, the traces of these two approaches have been detected, and each article has been assigned to one of those approaches as it was mentioned in the table above. It is seen that a major part of the current methodologies follows a bottom-up approach

(13/36). Since that research mainly focused on some specific factors to investigate the technical potential of rooftops and implement the methodologies in wide areas later. Shadowing factor and rooftop segmentation are the examples of these specific factors and this research take place in narrow areas for less buildings. On the other hand, a top-down approach has been implemented in research that focused on larger areas.

3.3.2 Black-box, Grey-box and White-Box approaches

Information that is recorded along with a geographic indicator is known as geospatial data. Vector and raster data are the two main types of spatial data. Data that is represented by points, lines, and polygons, such as houses, cities, highways, mountains, and water bodies, is known as vector data. A visual depiction employing vector data might, for instance, show dwellings as points, roads as lines, and entire towns as polygons. Raster data is made up of pixelated or gridded cells that are categorised by row and column. Images produced with raster data are significantly more complicated and include photos and satellite images.

Black-box approach is the approach that allows the user and also the analyst to be informed about input and output without giving any additional information of the process, methodology or the calculations related to the analysis. Methodologies containing neural network operation are included in the black-box approach.

White-box approach can also be called a glass box, open box or transparent box approach since it contains detailed information about internal logic and structure of the methodology. Full knowledge of the process should be known by the analyst. Most of the statistical approaches investigated during literature review are accumulated under white box approach title.

Grey-box approach is the combination of the two approaches. This approach requires limited knowledge of the internal work of the analysis and the knowledge of fundamental aspects of the process. Major part of the articles performed using QGIS only are classified as black-box approaches.

No	Black Box	Grey Box	White Box
1	Fundamental aspects	Partial knowledge of internal working	Full knowledge of internal working
2	Low granularity	Medium granularity	High granularity
3	User acceptance	User acceptance	Developers and testers
4	External exceptions	High level database diagrams and internal states	Internals are fully known
5	Least time consuming	Average	Most time consuming
6	By trial and error	Data domains and internal boundaries can be tested	Test better data domains and internal boundaries
7	Not suited for algorithm testing	Not suited for algorithm testing	Suited for algorithm testing

Table 7: Classification of the 3 approaches under different point of view

3.3.3 Input Data

In the papers investigated and summarised in chapter 3.2, it was seen that five different input data that are aerial images, LiDAR, 3D models, satellite maps and statistical data took place. These inputs are detailed in this subchapter. The major part of the research was using statistical data such as census data that consists of demographic, economic and population data for the desired research area and case study eventually. Census data includes basic population characteristics including age, sex, marital status, household composition and household size. Besides statistical census data, LiDAR data was used for an important number of research.

Light Detection and Ranging (LiDAR) data are high resolution raster of ground elevation with a delicate vertical accuracy. The technology that is used to obtain requires equipment mounted on an aircraft that are as follows: a laser scanner to transmit the pulses of light to the ground surface, a global positioning system (GPS) and an inertial navigation system (INS). (*Stoker, 2016*)

The third type of an input is 3D model data that is obtained by an onboard camera capturing hundreds of photos with intersecting aerial images. This data does not contain elevation data different from LiDAR data and they are mainly used to build a 3D map of the area of interest in order to better visualise and analyse the building characteristics. Satellite and aerial data are the other types of data that took place in current methodologies. Satellite images cover a wider area compared to aerial images, but proportionally includes less detail.

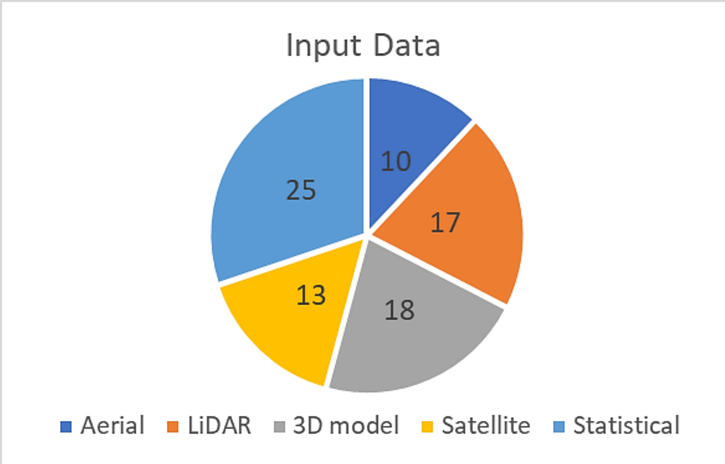

Figure 12: Different input data of our scientific papers analysed

3.3.4 Error Estimation

The most interesting output of the analysis is the capability to understand the accuracy of the estimation for what concerns the error related to the

Classification				Error estimation								
#	Author	Geographic are	Geo-scale	Error (%	Uncertainty	Uncertainty	Roof	Shadir	Solai	Numbe	Methoc	Year
36	Salvador Izquerdo	Buildings within	3,3	13,00	32,00%	4,160	+				+	2008/04
35	J. Ordóñez	Residential	3,3	10,00	2,53%	0,253	+					2010/01
34	L.K. Wiginton	Residential	1,5	15,00	60,00%	9,000					+	2010/01
33	Luca Bergamasco	Piedmont Region	3	1,70	110,63%	1,881		+				2011/02
32	Luca Bergamasco	The whole city of	2	10,00	18,81%	1,881					+	2011/09
31	M.C. Brito	Urban region	1,2	20,00	9,40%	1,881	+	+	+		+	2011/11
30	Aneta Strzalka	Multi-family	0,8	25,00	7,52%	1,881	+	+				2012/01
29	Nico Lukac	Urban area- city	1,5	2,60	72,33%	1,881			+			2012/09
27	J. Alstan Jakubiec	City level	1	4,45	40,88%	1,819			+			2013/05
28	Jinging Peng	Hotel and	1,6	12,50	15,05%	1,881			+			2013/05
26	Taehoon Hong	City level(due to	2,3	24,30	7,74%	1,881			+			2013/12
25	Nico Lukac	Large urban	1,5	12,00	19,63%	2,356					+	2014/01
23	Karoline Fath	Urban areas	1	15,00	12,54%	1,881		+	+			2015/03
24	Rhythm Singh	Buildings within	1,8	13,95	39,07%	5,450				+		2015/03
21	Dan Assouline	Urban areas	1	7,50	26,00%	1,950					+	2016/11
20	Romero Rodriguez	Federal State	1,8	13,00	14,47%	1,881			+			2017/04
19	Hong	Big district	1,5	15,00	12,54%	1,881			+			2017/05
18	Kai Mainzer	City level	2	10,00	30,00%	3,000	+	+				2017/07
17	Dan Assouline	Country scale	4	17,12	19,58%	3,352					+	2017/09
16	Mohajeri	City	2	13,00	34,00%	4,420					+	2018/06
14	Song	Small town	1,2	8,00	13,00%	1,040	+					2018/11
15	Mansouri	City	2	9,63	67,11%	6,466			+			2018/11
13	Huang	City	2,2	9,50	2,97%	0,282	+	+				2019/04
12	Alhammami	City	1,6	7,00	26,87%	1,881			+	+		2019/06
11	Bódis	European Union	5	15,00	1,80%	0,270					+	2019/07
10	Walch	Country	4	17,00	23,00%	3,910			+			2019/12
9	Lopez-Riuz	Capital city	2,4	13,00	3,48%	0,452						2020/03
8	Fina	Country	4	13,00	14,47%	1,881				+	+	2020/06
7	Schunder	City	3	9,38	20,06%	1,881					+	2020/07
6	de Vries	Several buildings	0,4	7,00	8,00%	0,560	+	+				2020/09
5	Gomez-Exposito	Andalucía, Pais V	3,8	1,00	188,07%	1,881					+	2020/10
4	Nelson and Grubesic	North part of the	1,25	15,76	40,93%	6,450					+	2020/10
2	Phap	Capital city (30	2,4	20,00	0,73%	0,145			+		+	2020/11
3	Reames	4 metropolies	3,2	7,00	4,29%	0,300					+	2020/11
1	Singh	13 cities, 79M	4	4,54	0,24%	0,011					+	2020/12

Table 8: Error estimation table

methodology which is used in every single paper. We focused our attention on the so-called absolute and relative uncertainty which is a good indicator to classify the kindness and reliability of the evaluations.

At the very end of the preliminary evaluation, we can identify an average value of the error estimation and its related uncertainty to give a quantitative evaluation of our review considering it as a milestone.

It should be noticed that there are only two uncertainty values which overwhelm the error estimation, marked in red. As can be seen they are over 100%, which should be physically improbable or a symptom of an error in the measurement estimation. For this reason, they seem quite strange and can be left aside and not considered. There are many statistical methods for establishing the reliability of a datum with respect to others, whether the datum is to be considered as an anomalous value rather than as only a suspect value. It can proceed by calculating the probability associated with the extraction of observations from the normal distribution furthest from the mean and if these are out of range they can be excluded.

We also decided to give a number to the standard deviation, which is a statistical dispersion index used to give an estimation of the variability of our data population around a position index, such as being in this case, the arithmetic mean of the error over the 36 scientific papers, therefore, it has the same unit of measurement as the observed values.

Error Estimation			
Average Value	11,77	Uncertainty	1,88
Standard Deviation	5,76	Confidence Level	90-95%
Final Result (abs.)	11,77 ± 1,88		
Final Result (rel.)	11,77 ± 16,0%		

Table 9: Error estimation and evaluation of uncertainty and confidence level

In statistics, when estimating a parameter, it is often not enough to identify a single value: it is therefore advisable to accompany the estimate with an interval of probable values for that parameter, defined as a confidence interval. The latter is an interval which indirectly characterises, in terms of probability, the amplitude of the value associated with. It's graphically equal to the area subtended by the probability distribution curve of the random variable in the considered interval and in our case this value is about 90-95% for every scientific paper analysed, which is a value we can trust to.

3.3.5 Error Band Estimation

Regarding the above explanation of which methods we used to analyse the estimation of the error, we decide to plot the outcoming results to better understand the actual situation of development regarding the methodologies used in our literal review.

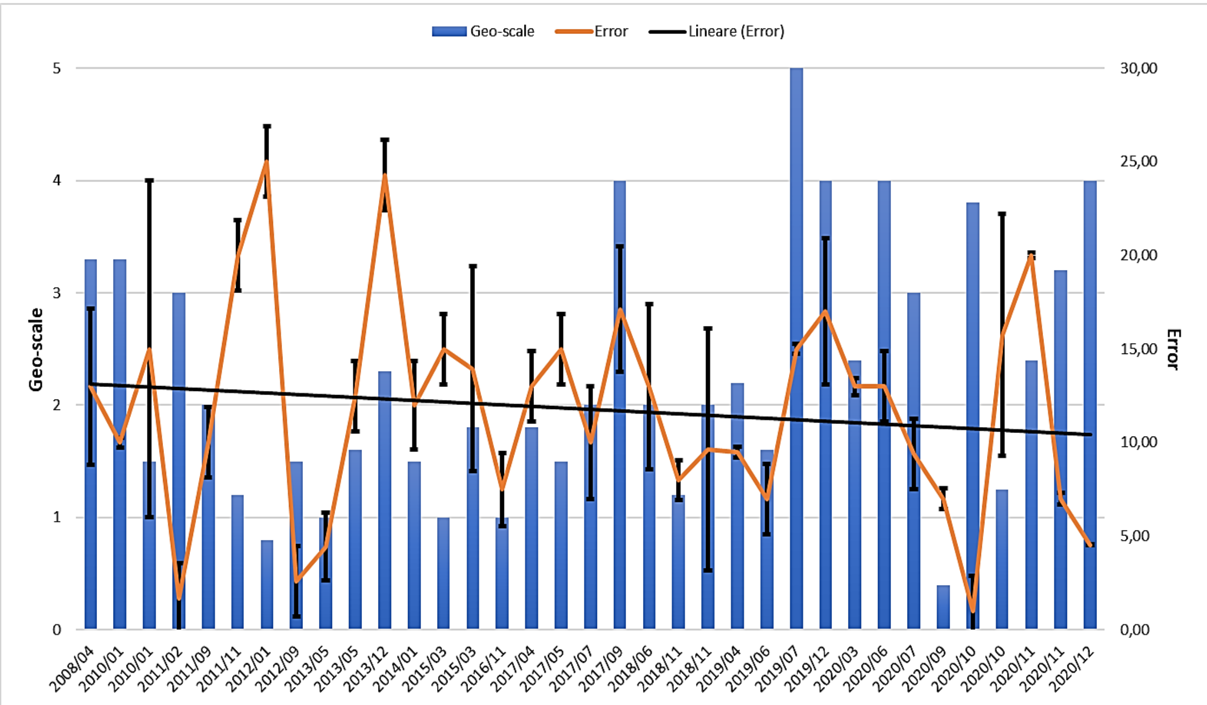

Figure 13: Trendline of the error estimation through the year for our scientific review

As can be seen on the x axis, we decided to follow the chronological order of the drawn-up date of papers in order to better visualise the trend line of the error band estimation. A linear error regression is used in order to statistically correlate in a functional way our data: we can see and understand through the graph that the uncertainty related to the accuracy of the calculations based on every specific methodology is decreasing during the years. This is an important output because it means that the quality of solar potential estimation is increasing year by year and we can base our knowledge on more and more reliable mathematical and logical estimations

Methodologies that use machine learning are adopted to support the decision-making model and to help researchers in calculations and simulation, in order to mitigate the problem related to other closed tools, such as ArcGIS and MATLAB, which limit the accessibility to the possible concept of open innovation[1], as a licence to be downloaded and used is strictly necessary.

[1] Open innovation is a way of managing innovation in stark contrast to the traditional management of company research laboratories, characterised by the secrecy of discoveries and therefore without sharing. It consists of a mentality, a way of conceiving research and development activities in the information age and globalisation based on openness.

4. Methodology

4.1 Preliminary data collection and preparation

The collection of the various data is the initial stage of the analysis which has a fundamental importance since the accuracy of the final outcome depends on the precision of the original data. In order to perform the analysis for achieving the purpose of the thesis, two main input data can be listed as the meteorological data and the territorial data. These two datasets are prepared in the QGIS environment to be used as input for the SEBE model eventually and for the spatial analysis of the available areas of the photovoltaic resource.

4.1.1 Meteorological data

Meteorological data is the input data for the SEBE model that needs observed hourly data of shortwave radiation for at least one year in length to obtain a detailed description of input forcing conditions. The data has been decided to refer to the Copernicus project (https://www.copernicus.eu/it) datasets provided by the "Ladybug tools" with the scope of making environmental design knowledge and tools freely accessible to every person, project and design process. [49]

The meteorological input data that are necessary to perform the calculations in the SEBE model are:

- Incoming/global software radiation [W/m^2]
- Diffuse software radiation [W/m^2]
- Direct software radiation [W/m^2]
- Air temperature [°C]
- Relative humidity [%]
- Time related variable (year, days of the year, hours, minutes)

The data mentioned above are provided by Italian datasets and they should be formatted by the UMEP pre-processing tool to be used with SEBE.

4.1.2 Territorial data

High quality territorial GIS data is crucial in order to calculate the PV technical potential. There are multiple public geoportals at national, regional levels of open sources service that provide geospatial data with high resolutions for Italian territory.

In order to estimate accurate values of the solar radiation on the roofs and ground surfaces the high-resolution DSMs are essential. They contain buildings and ground heights which are the main inputs for the SEBE model simulation. Firstly, it is important to distinguish between a DSM and a DTM (Digital Terrain Model), and this comparison will be done in the QGIS section in detail. In order to obtain the necessary DSMs, a formal request to the Ministry by mail and waiting for the credential to use for the download of the data were the required steps. The surveyed territory is organised in different raster which represent small land portions by several distinct DSMs.

Figure 14: Aosta city clipped overplayed by building footprints

Calculating the irradiance on the building walls it is necessary to generate the wall heights and wall aspects first, and that information are already included in the DSM raster datasets related function that is located in the urban geometry tool in the UMEP pre-processor plugin.

4.2 Tools and Software

In this section will be analysed in detail all the tools and the software we used in order to perform the simulation of the irradiance and the to apply the spatial analysis methodology, the most important of them are: QGIS with its relative plugins and the Machine Learning algorithm based on Python language and programming.

4.2.1 QGIS

4.2.1.1 Description and functionality

QGIS is a software that is an Open-Source Geographical Information System (GIS). In order to solve real world problems, the GIS tool is used by displaying, creating and analysing spatial information. Core library contains all functionality of GIS and analysis library that is built on the top of core library helps to perform spatial analysis on vector and raster data. QGIS aims to be user friendly via its easy-to-use graphical user interface (GUI) and it synthesises maps with data tables related attributes. Project can be viewed as the combination of vector and raster data that are respectively 2D and 3D. while various vector data formats are as follows: GeoPackage, ESRI shapefile, SDTS, GML; on the other hand, some raster data formats can be listed as GeoTIFF, ERDAS IMG, ArcInfo ASCII GRID. [50]

4.2.1.2 WMS and WFS

Online spatial data is obtained as OGC Web Services such as Web Mab Services (WMS) and Web Feature Services (WFS). WMS servers are used to obtain image files mostly in TIFF format. In contrast, vector files are provided via WFS servers that can be used to create base map layers and points, polygons and polylines would be imported. Therefore, they provide features including geometry and attributes to be used in geospatial analysis [51]. We used this service to upload Google Satellite on the maps in order to better visualise and adjust the projection of our raster maps and to check if everything is aligned in the correct way.

4.2.1.3 DTM, DSM, and DEM

Digital elevation models (DEM) are the spatial datasets that can be explained as digital sets of ground elevation for terrain demonstration. They can be classified in three main titles that are digital elevation model (DEM), digital surface model (DSM), and digital terrain model (DTM). DEM is a surface model without non-ground objects such as buildings and trees. DSM is an elevation model which includes ground and everything on the ground while DTM is a generic model that consists of one or more types of terrain information and it is a subset of DTMs. [52]

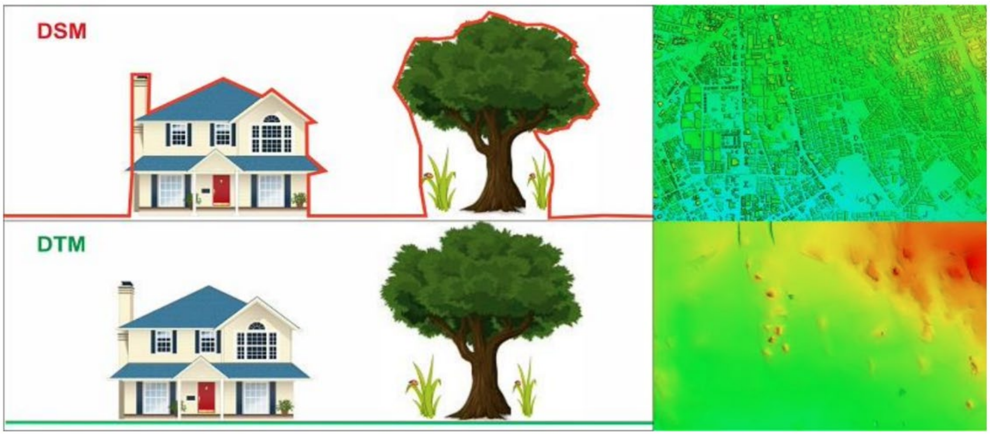

Figure 15: Graphical difference from DSM and DTM

4.1.2.4 Graphical modeler

In QGIS, the majority of analytical tasks are integrated into a process rather than performed individually. The graphical modeller can be used to merge that collection of processes into a single process, making it as straightforward and useful to run that process later on with a different set of inputs. [51] The steps in order to create a model and how to run it will be detailed in the following subchapters.

4.2.2 UMEP

UMEP (Urban Multi-scale Environmental Predictor), a urban-based climate service tool, combines models and sensitive planning essentially for climate simulations. There are many interesting applications, for example the impact of green infrastructure on the urban context, the effects of buildings on people's thermal stress, the impact of human activities on heat emissions, also mostly used for applications related to outdoor thermal comfort, urban energy consumption and climate change mitigation, in order to undertake simulations, consider scenarios, and to compare and visualise different combinations of climate indicators. As an open-source tool totally available on GitHub (also the developer version), UMEP is created to be easily updated as new features are created, and to be freely accessible to researchers, decision-makers and practitioners.

UMEP is an QGIS open-source plugin that is used to improve the modelling capabilities and it has been developed by the community in order to avoid the purchase of any kind of right or licence. The main function is to allow users to link the spatial information aiming to define model parameters.

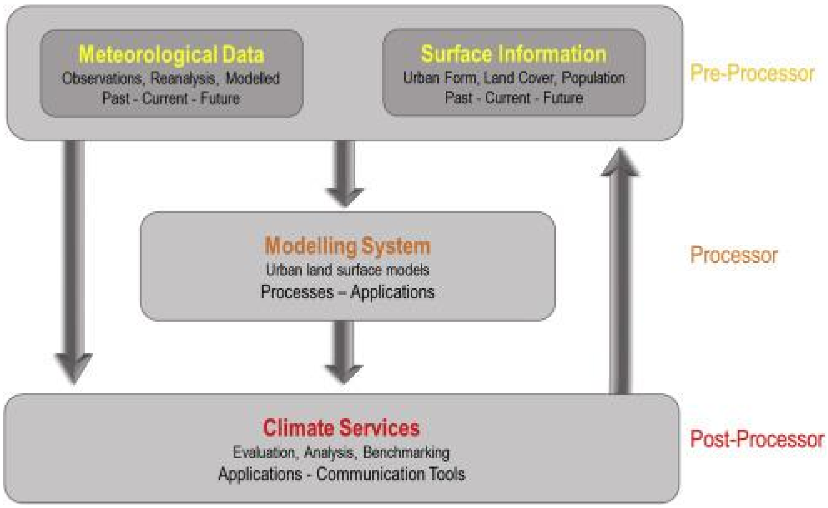

Figure 16: Pre-processing, processing and post-processing are the parts of the UMEP plugin

Pre-processing aims to prepare the spatial and meteorological data as inputs to the system. Processor feature covers all the main models for calculations. In order to plot the results and provide quick looks, a post-processor feature is used.

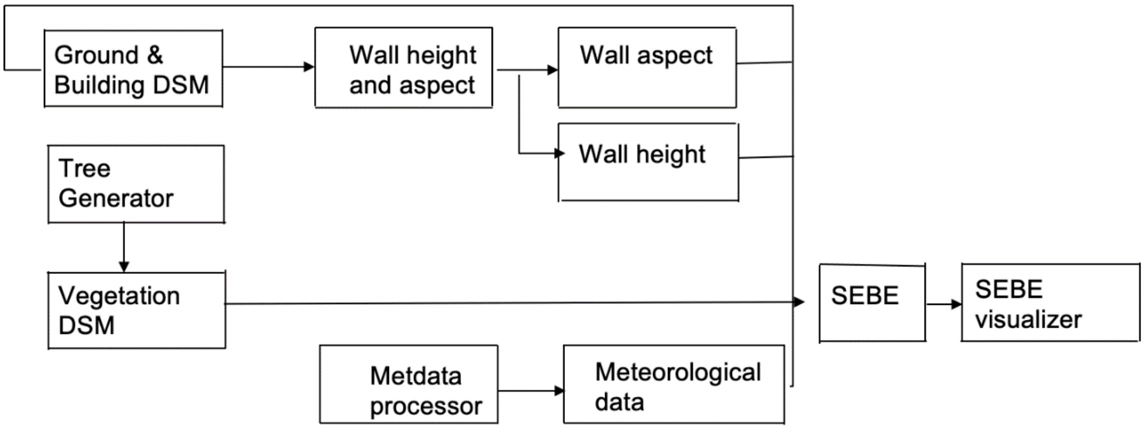

Figure 17: Workflow of solar irradiance analysis on building envelopes using SEBE model with necessary geodata

4.2.2.1 Solar Energy on Building Envelopes (SEBE)

The most useful UMEP plugin feature is used to map potential solar energy production, and it's called SEBE (Solar Energy on Building Envelopes). It can calculate irradiances at pixel resolution, so pixel wise potential solar energy on building roofs and walls using a 2.5-dimensional model using ground and building digital surface model (DSM). Summation of direct "IwS", diffuse "DS" and reflected "G(1-S)α radiation helps to calculate the total irradiance for a roof pixel "R".

$$R = \sum_{i=0}^{p}[(IwS + DS + G(1 - S)\,\alpha\,)] \qquad (1)$$

The shadow "S" calculated for each pixel is evaluated as:

$$S = Sb - (a - Sv(1 - \tau) \qquad (2)$$

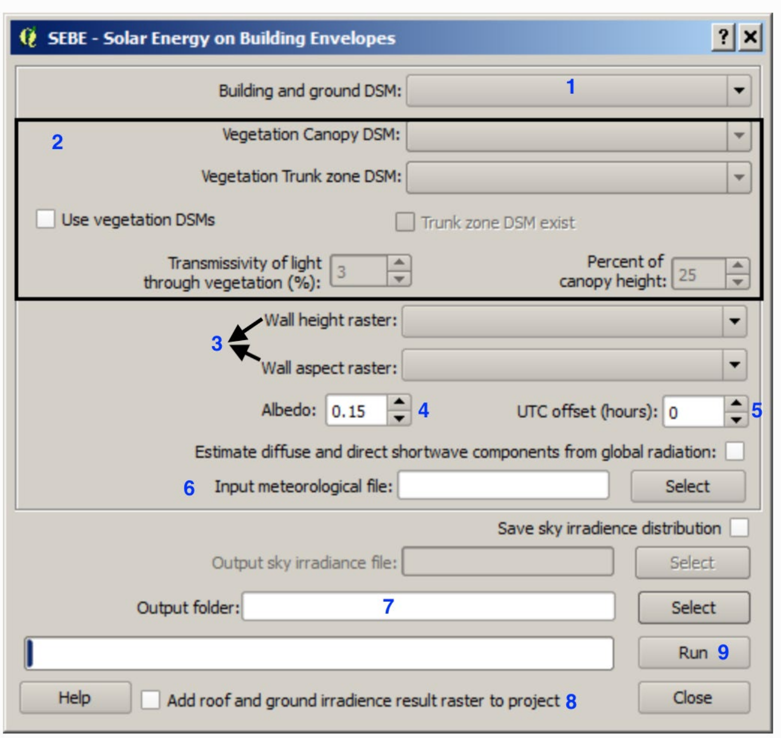

Figure 18: SEBE dialog box

Building shadows “Sb” and vegetation shadows “Sv” are represented as presence “0” or absence “1”, and transmissivity of shortwave radiation is

represented by “τ”. SEBE is focused on hourly radiation so this may result in underestimation or overestimation of the shortwave radiation fluxes. [51]

Starting from 1 to 6, the box contains the introduced input data, and after the 7th step the user needs to assign the output folder where they want to keep the output files and how to use the output files.

- **Building and ground DSM:** The box assigned as number 1 is used to select the DSM file with ground and building heights. The selected DSM file also includes the latitude and longitude used for Sun position calculation.
- **Use vegetation DSMs:** Two vegetation DSMs are necessary when this box is selected. The first vegetation DSM describes the top of the vegetation (Vegetation Canopy DSM) and the second one for the bottom that is underneath the canopies (Vegetation Trunk Zone DSM). This section is optional and not considered for this thesis work as an

assumption Vegetation data shadows ground, walls and roofs that causes a reduction in potential solar energy. Therefore, not considering vegetation data can cause an overestimation about solar irradiance.

- **Wall height and Wall aspect raster:** Two raster files are prepared by the high-resolution DSM that is stated in the first box above. Those two rasters contain wall heights and wall aspects of the buildings, relatively. It is mandatory to obtain these two rasters for calculating the irradiance on building walls.
- **Albedo:** The reflectivity of shortwave radiation of all surfaces that are specified in the vegetation DSM section as ground, roofs, walls and vegetation is determined by albedo box. The albedo is set to 0,15 and it is an average value used for all surfaces.
- **UTC offset (hours):** The position of the sun needs to be specified accurately, and UTC offset is used to achieve this objective. UTC is zero if the ERA5 dataset is used since it is related to the meteorological forcing data.
- **Input meteorological file**: The meteorological file that is in the right format to be used in UMEP is selected and added to the project here. At least one year in length dataset which contains hourly time resolution should be used for SEBE as it has been stated in the introduction part of preliminary data collection and preparation.
- **Output folder:** A folder specifically created for users where results are saved should be selected in this box. There would be two output files after this process that are one raster file showing irradiance on ground and building roofs and the other one for wall irradiance as a text file.
- **Add roof and ground irradiance result raster to the project:** By selecting this box the result is automatically added to the QGIS map canvas interface right after the process is performed.
- **Run:** This box starts the calculations that are computationally intensive with respect to the resolution and extension of the DSM.

Three mandatory datasets are saved as outputs if the model runs successfully:

- The geoTIFF **Energyyearroof.tif** shows the pixel wise total irradiance in kWh, both on ground and roofs
- The first and second columns of the text file, **Energyyearwall.txt,** represent the modelled grid, and the remaining columns display the irradiance values for each wall voxel starting from the ground and moving upwards as going right in each row.
- A second file, **Vegetationdata.txt**, which contains details on the height and location of the vegetation, is also stored if the vegetation DSMs were included. This vegetation file is in use of the SEBE visualisation plugin.

4.2.2.2 Graphical modeller

Most analytical activities are part of a chain of operations rather than being separate in QGIS. That series of processes can be combined into a single process using the graphical modeller, making it just as simple and practical to run that process later on with a different set of inputs, optimising the whole process while reducing the working time. A model is executed as a single algorithm regardless of how many steps it contains, saving time and effort, especially for larger models. Therefore, graphical modeller enables the users to create complex models via its easy and user-friendly interface. [51]

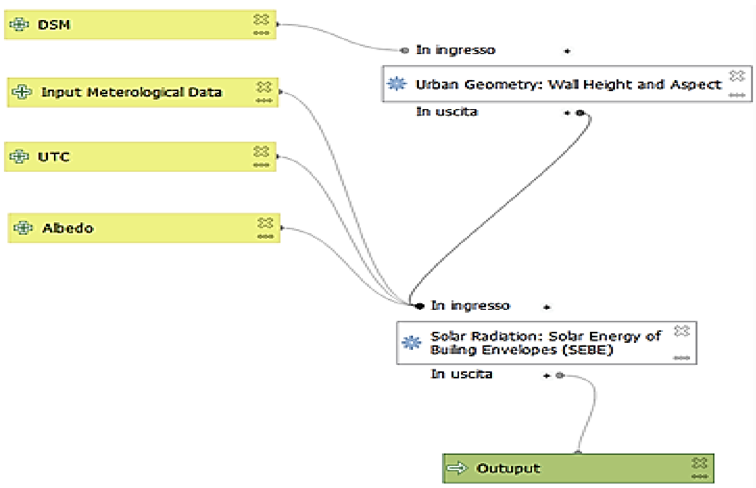

Figure 19: SEBE graphic modeler workflow

For our purposes the creation of the model involves two basic steps:

- Definition of necessary inputs: these inputs will be added to the parameters window, which will be generated automatically, so that the user can set their own values during the execution of the model, as for all the algorithms available in the processing framework.
- Definition of the workflow: Using the model input data, the workflow is defined by adding algorithms from the processing toolbox and selecting how they use the defined inputs or outputs generated by others in the model and how they are linked and related.

4.2.3 Machine learning

Machine learning is a field that leads computers to be able to learn without the necessity of being explicitly programmed. It helps to teach machines how to handle big data more effectively in order to interpret the exact information from the data and extract the most interesting features and results. Machine learning is the right study that should be applied; the objective of ML is to learn from the data. [53]

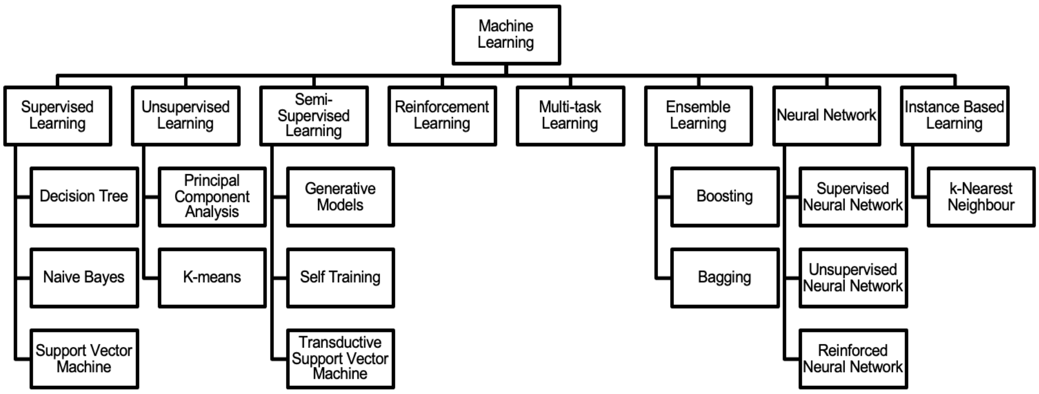

Figure 20: Major machine learning techniques

4.2.3.1 Supervised learning

The main objective in supervised learning is to generate a general rule for mapping inputs to outputs. In this method they are presented with examples given by a teacher: training dataset consists of a set of trained examples. Supervised learning algorithm's function can be simply summarised as analysing the training data and producing an inferred function, each pattern is paired including an input object and a desired output value. [54]

4.2.3.2 Unsupervised learning

In unsupervised learning, an "expert" intervention is not needed and the hidden structure of the inputs of a model can be found without the knowledge of its outputs. In other words, it can be said that unsupervised learning technique is the contrary of supervised learning. [55] In statistics, density estimation has several similarities with unsupervised learning. Unsupervised learning is integrated into several other techniques that are seeking to summarise and explain the key features of the data. It also includes several methods that are based on data mining used to pre-process data.

4.2.3.3 Neural network

Neural network is based on algorithms built in order to understand the underlying relations in a dataset based on a process that simulates the way the human brain operates. Therefore, it refers to a system of neurons that can be either organic or artificial in nature. These algorithms can adapt according to varying inputs for the network to generate the best achievable result without the need of output criteria redesign. Artificial neural networks act in a similar way that works in three layers as follows: input layer assigned for input, hidden layer processing the input, and output layer to send the calculated output. [53]

4.2.4. Research project on GitHub

GitHub is a hosting service for software projects used mainly by developers, who upload the source code of their programs and make it downloadable by users. They can interact with the developer through a system of problem tracking, pull requests, and comments that allows them to improve the repository code by fixing bugs or adding features. In addition, GitHub elaborates detailed pages that assume how developers work on the various versions of the repositories. [56]

Chapter 5

Results and Discussion

5.1 Photovoltaic potential assessment workflow

The photovoltaic energy potential is achieved proceeding through an original and scalable methodology. The assessment requires the evaluation of the useful solar radiation (physical potential), suitable surface (geographical potential) and PV system efficiency (technical potential). A GIS-based approach is used, and the Urban Multi-scale Environmental Predictor (UMEP) plug-in is implemented.

At the very end of the analysis, we will compare our results with the help of some machine learning techniques, in particular two of them, found on GitHub and managed to be implemented in our case study in order to extract some useful features and to give some ideas of future developments based on this type of technology.

The methodology is divided in three fundamental steps:

1. Collecting the geospatial input data from the public geoportal of Aosta city [57] and data pre-processing using the QGIS toolbox and plugin of UMEP.

2. Estimating the solar irradiation on roofs and ground surface, making use of high and low resolution of the digital surface models (DSMs) and some spatial analysis manipulation in order to correctly proceed with the simulation, while querying rooftop available segments with the solar irradiance map.

3. Estimating the solar irradiance and evaluating the technical PV potential on the suitable areas considering the system technical specifications.

The summary workflow methodology for the solar energy assessment is shown in the following figure:

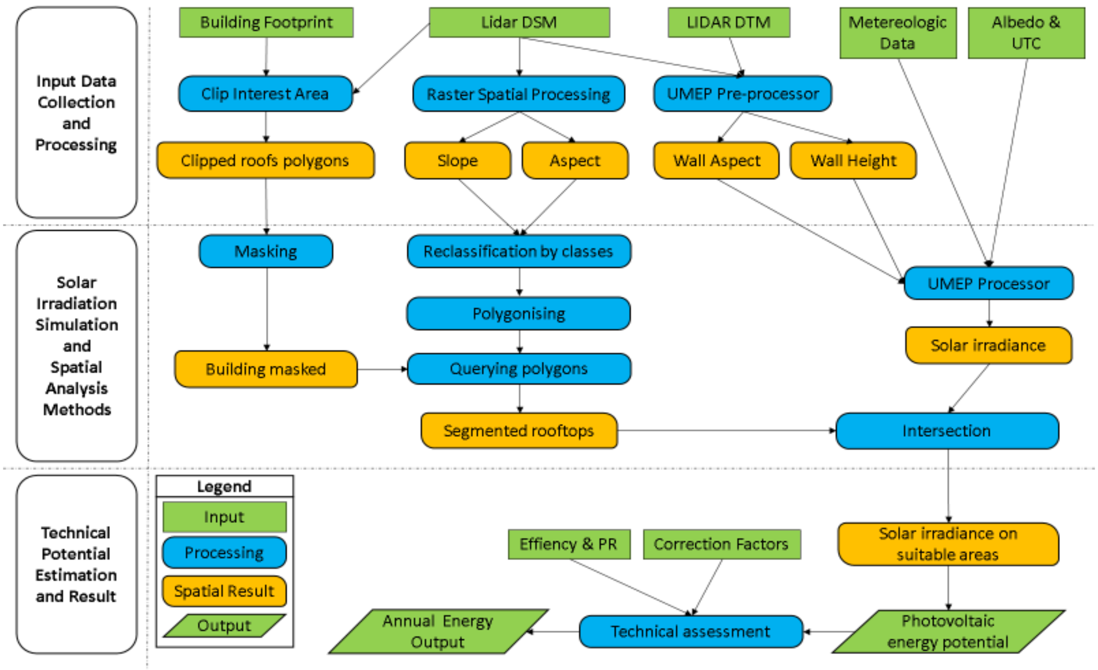

Figure 21: Photovoltaic technical potential assessment workflow

The scope is to honestly follow this workflow trying to explain as better as possible every single step with the help of intermediate representations. Firstly, we will start with the simulation applied to the Aosta data with a 2.0 x 2.0 metres of resolution and then a focus on a smaller area in order to not increase too much the computational effort of the simulation, as it's time-consuming running it for the whole 0.5 x 0.5 metres resolution.

5.1.1 Evaluation of the technical potential

Based on the calculated annual solar radiation per unit surface, the total usable area, and the efficiency of the PV technology, the annual potential of solar power generation at a chosen location can be approximated. The annual technological potential for photovoltaic energy is calculated using the equation below:

$$EAC = Hg \cdot APV \cdot \eta STC \cdot PR \tag{3}$$

"Hg" represents the global in-plane solar irradiation [kWh/m^2/year].

- The total area for PV electricity generation is represented as "APV".
- "ηSTC" is the rated efficiency of PV modules at STC.
- Performance Ratio is referred to "PR". It compares the energy actually produced with respect to what is produced under the same amount of irradiation, but under ideal no-losses conditions. It is the ratio of the final system yield to the reference yield and it's evaluated by the calculation of many different coefficients.

5.2 Sebe for Aosta DSM 2.0

Going through the simulation, the stages relative to the pre-processing and processing SEBE plugins are skipped, because they were explained in detail before. Parallel to this process, there is the Raster Spatial Processing: using the in-built QGIS tools, the DSM was used to create an aspect raster and a slope raster which assigns each pixel an angle. In the case of the aspect raster, there is an angular direction or bearing where 0° represents the North then each degree is assigned through changing color until wrapping back around to represent 360°; in this way 180° is the south direction. For the slope raster, each pixel is assigned an angle with 0° representing flat rooftop and 90° representing a vertical one. Colours are totally arbitrary, but used in the smartest way to make them understandable as much as possible.

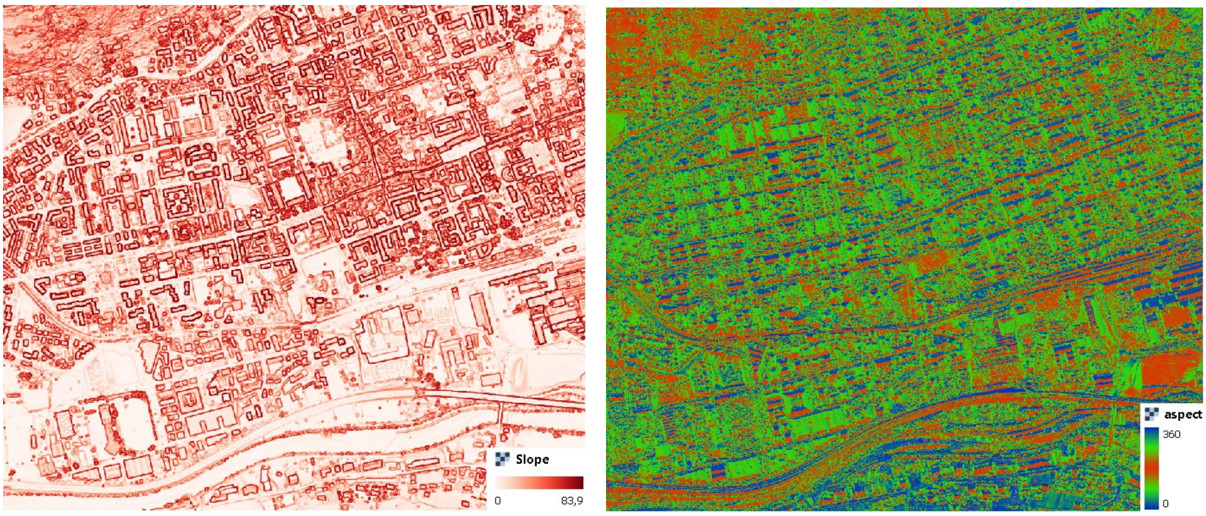

Figure 22: Slope (left) and Aspect (right) result from processing tool analysis

From these rasters, roof planes can be distinguished as they usually face in a common direction and have a common slope across the entire surface of the roof plane. For the slope map on the left, a red gradient was used to highlight the shape contours of the upper part of the building and its inclination with respect to the horizon. For the aspect map, a more colour band was used, to identify the directions to the cardinal points, as we are interested in the South (an exposure of 0 means that the slope faces north, 90 turns east, 180 turns south and 270 turns west), we should focus our attention on the red color, between 150°-180°.

Other details such as the ridges of the roof can also be seen as red lines dividing the roofs, because it means that the slope is rapidly changing in degrees meaning that they are not suitable to place photovoltaic systems.

In order to identify the roof planes, the aspect and slope rasters were first reclassified through the "r.reclass" function, so that the full range of aspect and slope values could be narrowed down to 4 and 9 classes respectively. Classes are ranges of values over which pixels in rasters have similar values. This allows to group areas with a common exposition and slope. The 4 aspect classes were 315 ° - 45 °, 45 ° - 135 °, 135 ° - 225 ° and 225 ° - 315 ° which represented North, East, South and West, trying to pay attention to the pixels with a null (zero) value, which are not to be considered towards the North but towards the South as they are flat but in any case, favourable to the installation of PV panels. The 9 slope classes were in groups of 10° each from 0° to 90°.

After the reclassification, only the pixels belonging to the south direction and the flat ones were taken into consideration, while the points with slopes greater than 60° were removed as not suitable for PV installation. They were also clipped using building footprints from the software to remove data outside the building boundaries, such as data from trees and roads. The sum of these processes can be more easily understood graphically below.

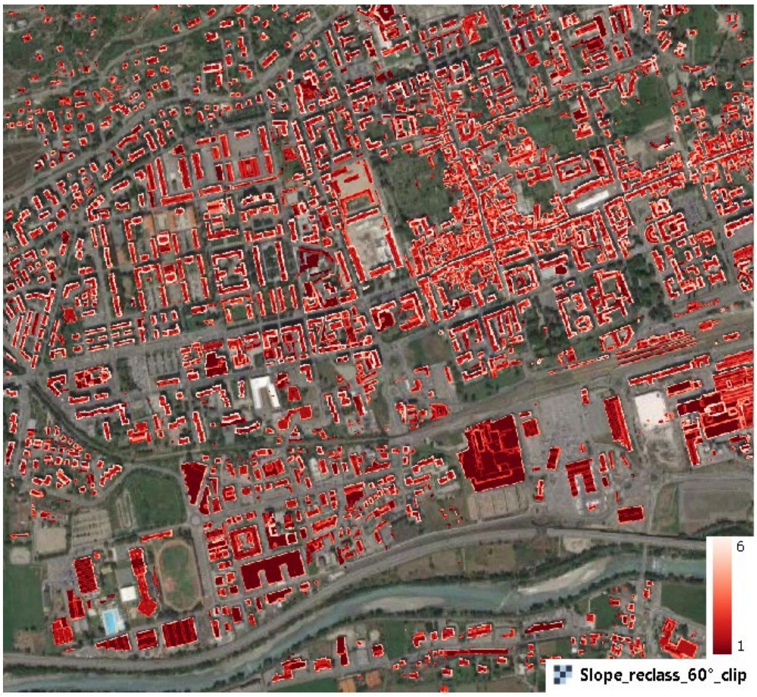

Figure 24: Slope reclassification and querying until 60°

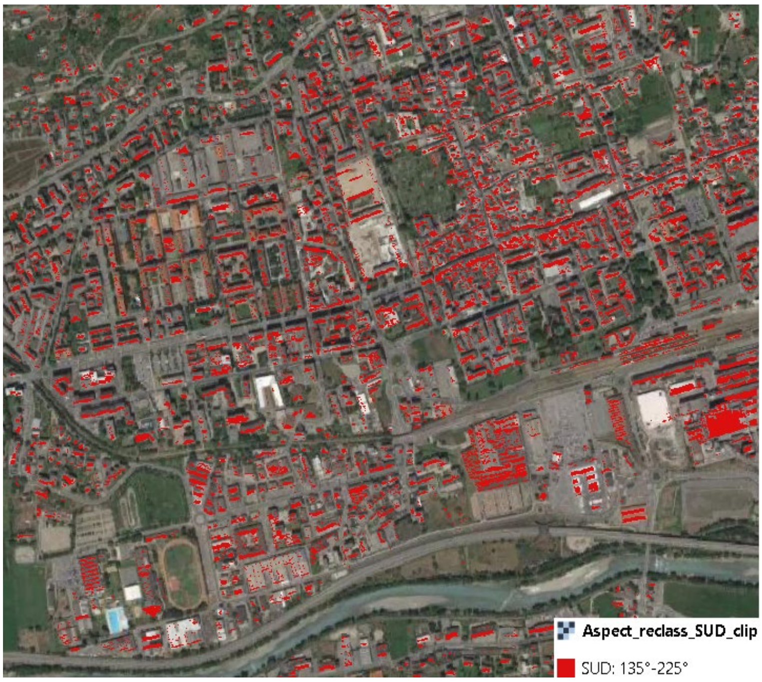

Figure 23: Aspect reclassification and querying SUD and "flat" direction

The aspect and slope rasters were then converted to vector polygons using the QGIS "Polygonise" tool. This method allows to group adjacent pixels with the same value together and creates polygons to represent that pixel area. This format is much more useful as polygons, such as lines or dots, can be exported as a shapefile allowing to store the shape, height, width and position of each polygon as well as any other attributes, such as slope and aspect which can be saved as part of the database file.

As roof planes have a common aspect and a common slope across their surface, they can be identified by taking the intersection of the aspect polygons (south direction + flat roofs) and the slope polygons (below 60°). The result of this process is a new set of shapes with one aspect value and one slope value, they thus represent roof planes. In this way we can fully visualise the best total area for installing PV panels.

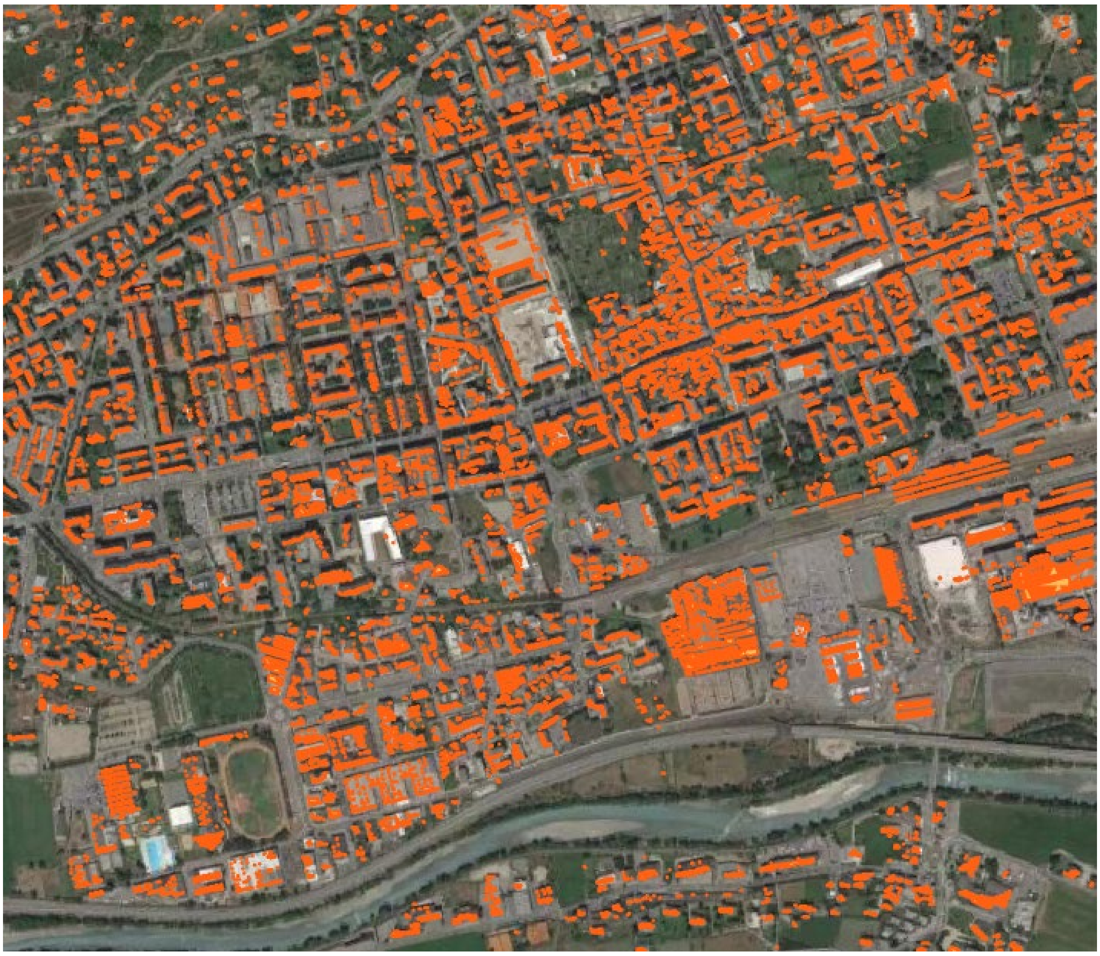

Figure 25: Available rooftop polygon querying for solar irradiance estimation

Many of the polygons identified in the intersection of the vectors are only a few pixels in size or have a strange shape. Some post-processing is needed to get rid of these weird roof areas and keep the actual roof plans. This was achieved through a series of buffering processes that involve expanding or shrinking the shapes by a certain amount through the filters in order to smooth out the shapes. The result of these post-processing steps can be seen in the figure, after considering that it is needed at least 7 m^2 for every kWp.

The aspect and slope of each plane of the roof were found using the "Zonal Statistics" tool in the QGIS processing bar which allows for statistical calculations involving a vector layer, such as the ability to measure the area of each extracted polygon. Now that spatial analysis simulation has been managed, the focus of our attention goes to the irradiance simulation.

An irradiation simulation is required as the diffuse component of solar irradiance depends highly on the surrounding structures, such as buildings and trees, as some component of the light bounces and scatters off of them. Each pixel has a colour representing the annual irradiation on exactly that point with colours from blue to red representing low to high irradiation. The units in the output raster are in kWh, and each pixel represents an area of 2 m^2.

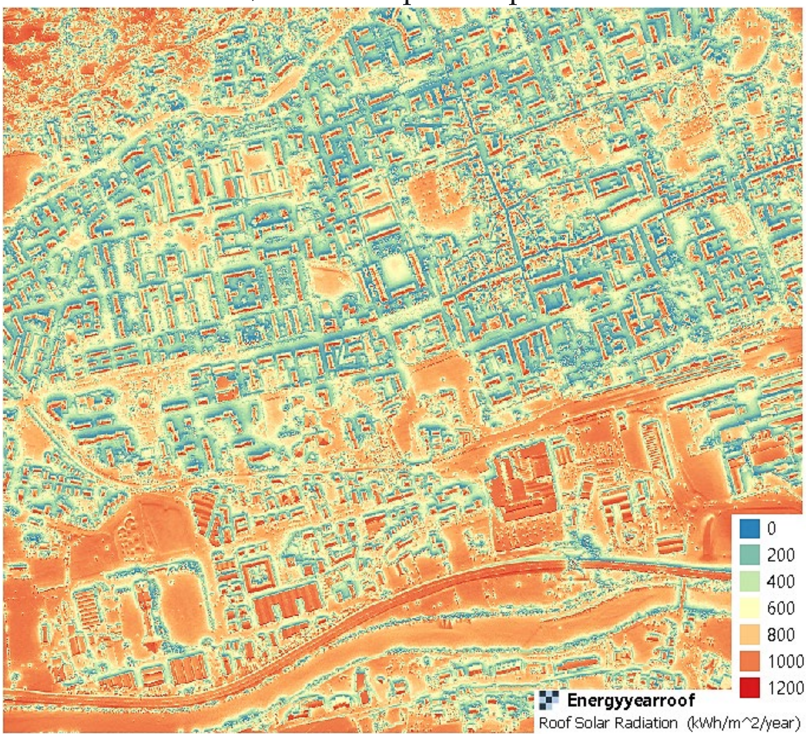

Figure 26: Solar irradiance simulation of the city using SEBE

This fact was exploited using the "Zonal Statistics" tool in QGIS by finding some key information such as the mean or median of the irradiance values within each polygon. This results in the calculation of the average annual energy across polygons in kWh/m^2, a quantity that could be used to find the solar potential energy output for a solar PV system and it's also capable of understanding which is the best spot to install it.

Figure 27: Irradiance output form SEBE, clipped with the building footprint

5.3 Sebe for Aosta DSM 0.5

All the steps just performed should be faithfully repeated also for the DSM of the city of Aosta with a resolution of 0.5 m^2. The problem is that with the current instrumentation at our disposal the processing times would increase exponentially as the grid available for the simulation would have many more values to consider. With this consideration we have decided to limit the extension of the studied area in order to have the ability to bear this load in terms of computational time of our IT resources.

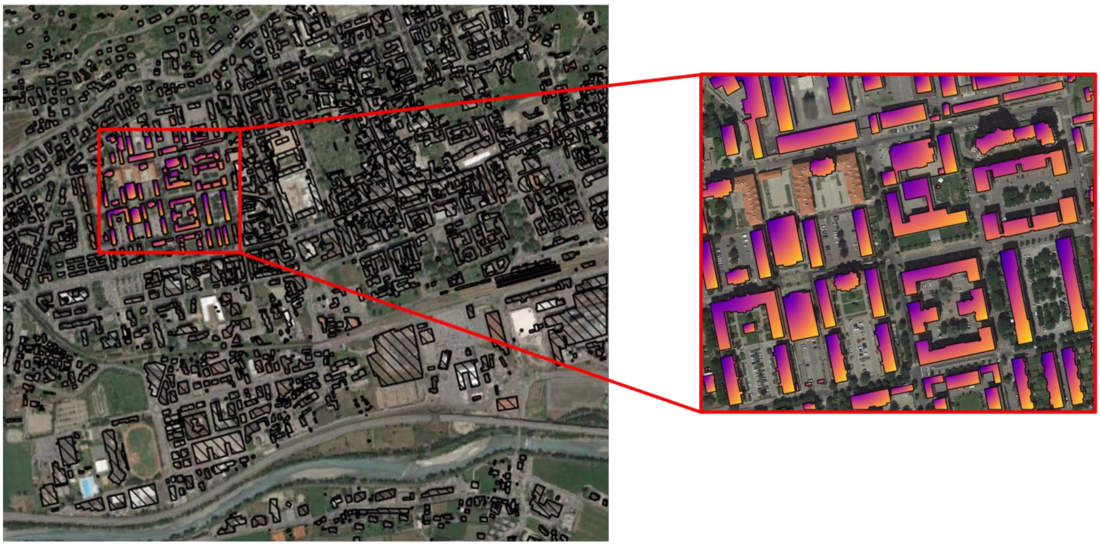

Figure 28: focus on the selected map area to simulate irradiance with 0.5x0.5 DSM model

It's not the intention of the thesis showing again the passages to follow, as they are the same as before, so only the results about irradiance and rooftop area suitability will be shown here below and the related clipped layer.

Figure 29: Example of SEBE output solar radiation visualization for Aosta 0.5

Last step, determine the area after removing the wall area (which is assumed to have a slope greater than 65) from the buildings opening the raster calculator. To visualise where to place solar panels, the amount of energy received needs to be cost effective. As irradiance below 900 kWh is considered to be too low for solar energy production pixels lower than 900 can be filtered out (Figure below). Changing transparency allows you to make only points above a threshold of interest visible.

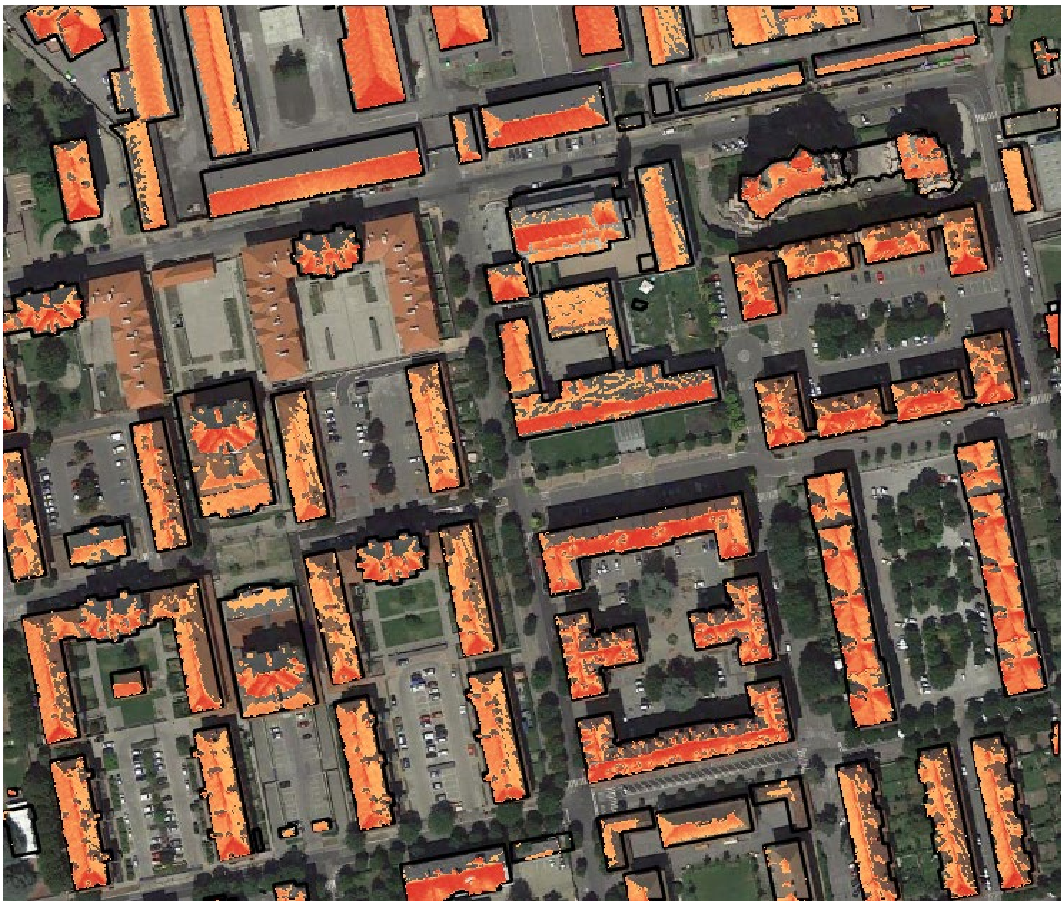

Figure 30: Solar radiation on Aosta urban centre roofs and filtered by threshold minimum irradiance

5.4 Indirect method

5.4.1 Reduction coefficient

Estimating the roof size is a crucial component in understanding a building's shape and solar thermal potential. The available roof area must be calculated after considering a number of factors. The Aosta Valley geoportal downloads the reference data that identifies every single building and its footprint and manipulates it on QGIS, but only the civil, social, and administrative buildings, as well as the commercial and industrial ones, are taken into consideration for this study. Since little is known about roofing characteristics, it is necessary to assume a representative roofing typology and conduct an empirical analysis of it based on a visual review of Google Earth images. This lack of information will force us to estimate the rooftop topology and it will increase the error related to the available area estimation; the usable roof area that receives solar radiation for the PV facility is defined as the geographic potential for solar photovoltaic installation, as we explained before.

5.4.2 Suitable area calculation

The mutual-shading effect is taken into consideration in this analysis since all building roofs are regarded as flat roofs. The inter-array distance design is not straightforward and depends on a variety of variables that are easily adaptable on a case-by-case basis. The covering index coefficient C_C, which is assumed to be equal to 0,5 and indicates the ratio of module surface divided by the total roof surface available, is included in order to account for these gaps; we will refer to the only papers related to Italian country [20], [21]

Solar collectors are not very effective and practical economically at that latitude the climatic and meteorological conditions are to be understood as mountainous, there are few numbers of them on the rooftops of Aosta Valley. Due to the possibility of solar-thermal systems occupying 20% of the roof surface, it is thought that this area may not be used for installation of solar panels. The corrective solar-thermal coefficient C_{ST} is therefore considered to be equal to 0,8. As a precaution, it is also considered that chimneys, aerials, roof terraces, heating, ventilation, and air conditioning (HVAC) systems, or other functions, occupy 35% of the roofs' area. As a result, the corrective feature coefficient, or C_F, of value 0,65 is introduced. The total corrective coefficient for roof-top PV plants, C_{RedF}, is produced by adding all the coefficients stated above. [19] It limits the total roof area that can be used for solar panels, hence limiting the global PV potential. It shows what percentage of the roof's surface can be covered by solar panels:

$$CRedF = CC * CST * CF = 0{,}5 * 0{,}8 * 0{,}65 = 0{,}26 \qquad (4)$$

5.5 Aosta Graphical Model

So here the idea is to combine a series of processes into a single process using the graphical modeller, optimising the computational time, making it just as simple as practical and to run that process every time we want with different set of inputs.

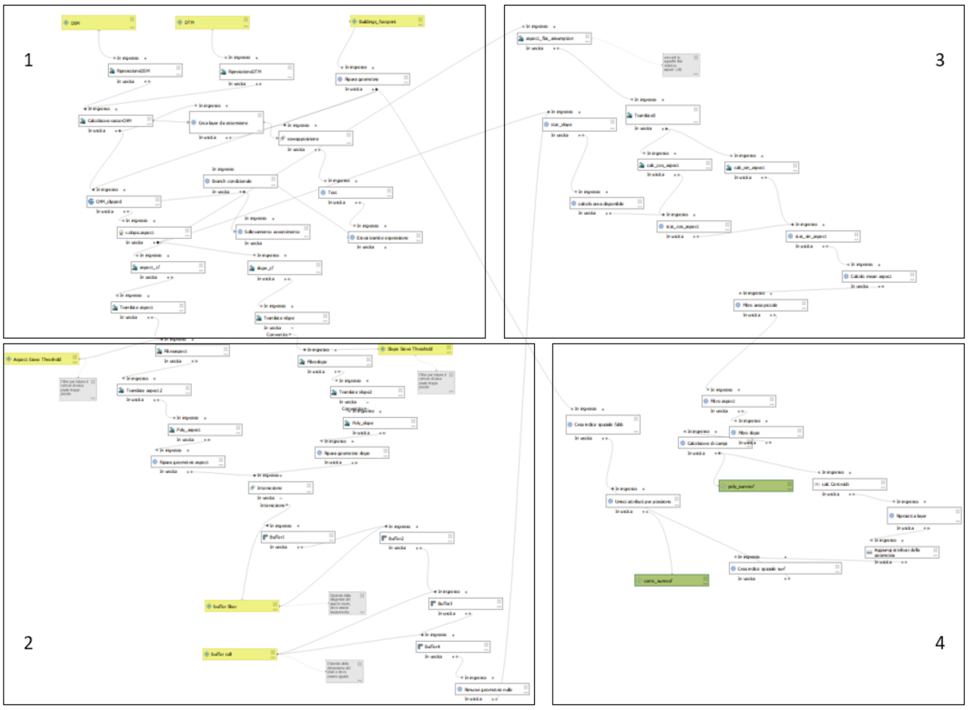

Figure 31: SEBE graphical model

This SEBE model kindly lent by PhD' students at the EC[2] and slightly modified and adapted to our case study. It is mainly composed by 4 parts:

1. Reprojection of the input raster data to get the right coordinates, than obtaining the difference between two rasters of DSM and DTM as well as checking if there are available buildings for selected areas.
2. Buffering the obtained layers by considering the length of the diagonal of each pixel that is for the observation of a zone around a geographic feature containing locations that are within a specified distance of that feature. The aim is to reduce and put together small or strange pieces of rooftop area, while eliminating the isolated one.
3. Calculating statistics such as sin and cos of the orientation of the building in order to obtain correctly slope and orientation. Reprojection of the buildings and correction of wrong topologies.
4. Extracting centroids and calculating the available rooftop areas considering buffering and planes statistics.

5.5 Machine Learning Algorithms

As previously mentioned, we will take advantage of some of the most interesting projects present on GitHub, always riding the wave of open innovation, and we tried as much as possible to integrate that machine learning algorithms to our case study in order to verify if our thesis at the beginning will be demonstrated. Once again, our aim is no to fully substitute a methodology at the expense of others (none other than the human mind), but to find the best combination between different tools to find the best solution. The project has been developed and tested with python 3.6: the required libraries are numpy, Pytorch, sklearn, OpenCV and the library for visualisation is matplotlib.

[2] EC is a research centre, located in Turin and belonging to the Polytechnic (based on the themes of energy, sustainability and innovation with the aim of launching a series of actions and projects that will provide support and strategic advice to local authorities, national and transnational bodies, on energy policies and technologies to be adopted)

5.5.1 "Detecting available rooftop area from satellite images to install photovoltaic panels" project

The repository contains the code for the Machine Learning 2020 course (CS-433) project 2 at the EPFL, based in Switzerland, is on the most important European scientific and technological institutions in collaboration with LESO-PB Lab and is also the code of basis for the research project: "Quantification of the area suitable for the installation of solar panels on the roof from aerial images using convolutional neural networks" and it belong to Riccardo Cadei and his colleagues, also called contributors. [58], [59]

The original project target was to segment in aerial images of Switzerland (Geneva) the area available for the installation of rooftop photovoltaics (PV) panels, namely the area we have on roofs after excluding chimneys, windows, existing PV installations and other so-called "superstructures". The task is a pixel-wise binary-semantic segmentation problem and we are interested in the class where pixels can be classified as "suitable area" for PV installations. We tried and managed to integrate these scripts to our case study, here below there is a brief description of the different processes and the data needed to run the algorithm. To fully explore this solution, a "Jupyter Notebook"[3] where we will present the entire pipeline to train a U-net model from a desired data set, evaluate the results and visualise the predictions. We present multiple ways to initialise and train a U-net. [60]

<u>Data:</u>

- The input aerial images are RGB aerial images in PNG form and each image has size 250×250×3 with pixelwise 0.25×0.25 m^2.
- All the images in the dataset are labelled
- The labelled images are a binary mask with 1 for pixel in PV area, and 0 otherwise.

[3] It is a community run project with a goal to "develop open-source software, open-standards, and services for interactive computing across dozens of programming languages, to support interactive data science and scientific computing. Jupyter will always be 100% open-source software, free for all to use.

- The original input images are transformed with saturation and classic normalisation before training.
- A real-time data argumentation is applied only on the training set by randomly flipping images horizontally or vertically or rotating in ninety degrees, in order to increase the volume of the dataset.
- The output of our model is again a binary image, where the pixel is one, if its probability of being in the PV area is bigger than a fixed threshold.
- Train/Validation/Test dataset Ratio: 80/10/10 %

First, we load the data set that we will use for training. Each sample is an image with its mask (label). An image is represented as a 3x250x250 array with each of the 3 colour channels being 250x250 pixels. The associated mask is a 250x250 array. Note that we already split the images in train/validation/test 80/10/10%, exactly they are 420/53/52, in advance to make our reproducibility as clear as possible. Please note that we should have labelled our Aosta buildings one by one manually. In order to avoid this very time-consuming procedure and heavily susceptible to human error, we decided to run the model on data already well labelled but with the difference that they belong to the Geneva city and not the Aosta one. We decided to keep them as the two cities are quite close to each other and also the climatic characteristics are quite similar.

We perform data augmentation and transformation on the training set to counter the low number of images in our data set. However, in the validation set and test set, we only perform transformation and no augmentation.

<u>Methods</u>

- A Convolutional Neural Network (CNN) model based on U-net has been used. IoU and Accuracy are computed to evaluate the performances.
- We trained our model firstly on the whole dataset, then we focused only on a specific class of images, residential area

Then we train the U-net model and we may compute the mean and standard deviation of the train loader: this is used either to check if the data loader is normalised, or to compute the mean and the standard deviation for the normalizer in the data loader, these variable are tensors and their values are:

- tensor _mean ([-0.0078, -0.0093, -0.0102])
- tensor_ std ([0.9968, 0.9970, 0.9968])

There were a certain number of training methods, they all initialised a new U-net model from scratch and trained it. For simplicity we decided to use the regular training: we can tune the number of epochs, the learning rate and the parameter of the loss function.

The following steps are quite straightforward considering that the project structure was very clear and delineated. Now it's time to explain better the most important parameters that affect our simulation, as they are key features for the training, the testing and the validation of our model.

- Loss function: is an optimization function that maps an event, or values of one or more variables, to a real number which intuitively represents the minimization of a "cost" associated with the event. In other contexts, we may be dealing with an objective function or its negated function that must be maximised. It is typically used to estimate parameters and is a function of the difference between predicted and actual values for a data instance.
- Number of epochs: iteration period
- IoU: the intersection over the union, also known as the Jaccard similarity coefficient, is a statistical index used to compare the similarity and diversity of sample sets. It is defined as the size of the intersection divided by the size of the union of the sets of samples; is a term used to describe the extent to which two boxes overlap. The larger the overlap region, the greater the IOU.
- Accuracy: the fraction of predictions our model got right used to evaluate classification models

Result

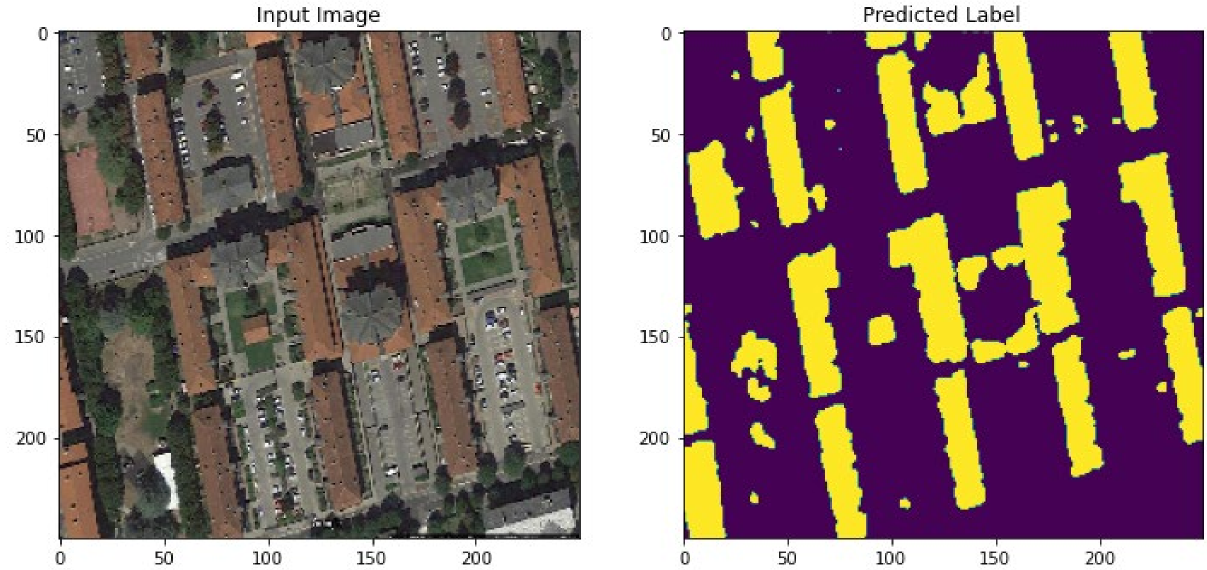

Figure 32: Image binary classification from ML method

We are able to automatically detect with satellite images the available rooftop area at pixel level with performances comparable to state-of-the-art seen in the literal review. In particular, taking a snapshot of our area of interest, which is characterised only by residential buildings, we got on the test set an accuracy of about 0.87 and an Intersection over Union index of 0.77 using only 244 images for training.

5.5.2 "Rooftop detection using Python"

This second repository of GitHub contains a project written in Python to detect rooftops from low resolution satellite images and calculate area for solar panel installment. Individual rooftops of each and every house are identified and segmented out, this is the most difficult part meet during our dissertation. There were a lot of paper showing a well and complete image segmentation using machine learning techniques, but none of them really explaining in detail how those algorithms were written and how they really worked.

The images can be imported from Google Satellite with a quality of 20 zoom levels and even if the resolution of them is so poor that the boundaries can't be even figured out by, this script makes all the job. [61]

Methods

The most important features of image segmentation are:

- Hough Transform: It is used to localize shapes of different types of rooftops. When applied to the image, it gives very less true positives. The main problem was to set threshold parameter of Hough Transform as it's used to detect exact shapes like squares and rectangles. The main limitation of this method was that it won't work for other structures if not perfectly square or a rectangle present in the image.
- Adaptive Canny Edge: Applying auto canny on the low-quality image of rooftop results in exact edge detection of rooftops. Contour Area localization and then applied threshold to detect rooftop.
- Watershed Segmentation: Segmentation on the images from maps to count the number of buildings and to plot rooftop area of each building present in the image.
- Edge Sharpening: Due to the poor quality of the image, to mark the rooftop area edge sharpening of the image is to be done.
- Active Contours: It can be applied on the rooftop area to extract the optimal area for the solar panel.
- Hough Transform: Hough Transform can be used to analyse the shape of the rooftop, it outlines the rooftop and obstacle boundaries.
- Pixel wise Polygon filling: Applying this method on the rooftop and moving around the contour in a clockwise direction each pixel and its surroundings was marked as rooftop area.
- Region Based Polygon filling: After applying Hough Transform in combination with K-Means clustering, the rooftop area can be divided into different regions. Checking the intensity of different patches, the area was marked as a rooftop area or not.

Result

So, trying to unify some of these features, we manage to analyse a small image of Aosta rooftop, as we noticed that increasing the width of the photo, also the computational time will increase exponentially; an advice for the next development of this technology must be for sure the improvement of the available resources as running this algorithm on a more performing pc, will save the day.

In the figures below we can see the original image and the final one also containing the solar PV panels, which were trying to be placed by the algorithm with a quite high accuracy, as can be seen by the avoidance of the obstacles (superstructure, chimneys, windows). This process passes through the sharpening and the edging of the selected image and the segmentation of the roof thresholds in order to understand where the solar panels can be placed.

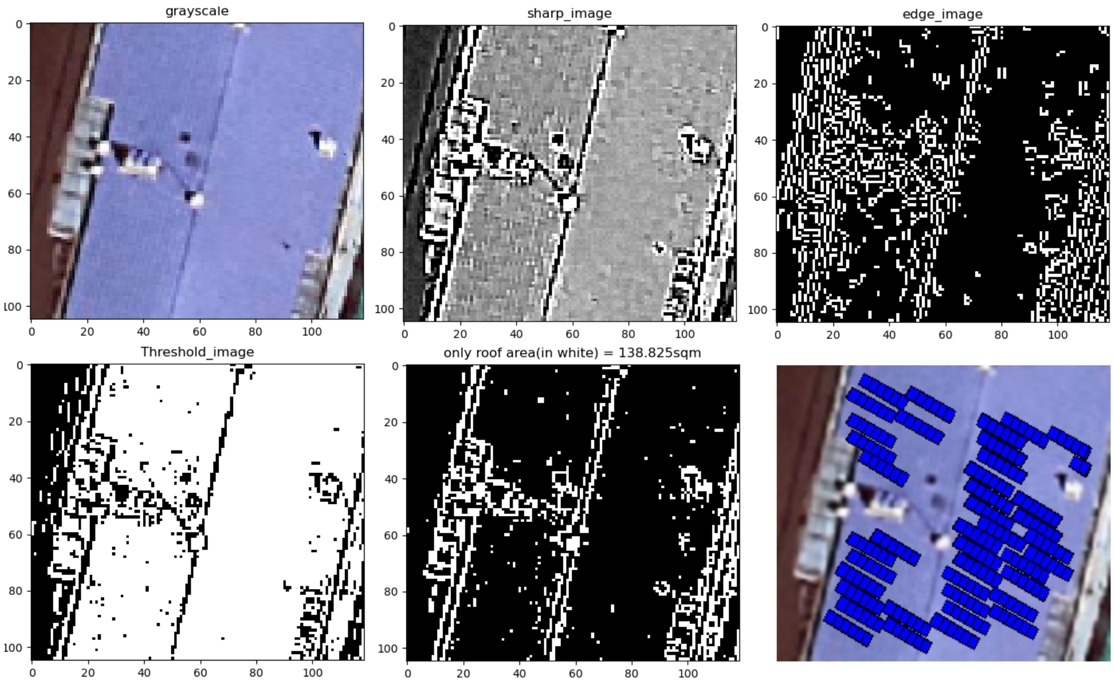

Figure 33: Optimal Rooftop Area for Solar Panels

As mentioned earlier, the idea behind the implementation of these projects is not to find a universal and a definitely working solution, but more than anything else a real help to scholars, researchers, students or even amateurs, to find a good compromise in terms of lack of data, accuracy, computational time and the ability to share knowledge.

For sure, these algorithms have to be well tuned in order to achieve our working objective, but many of the times their existence is unknown (see the poor presence of machine learning techniques over the past 15 years in our literary review), but the but the potentials of this tool are very high, almost unlimited.

5.6 Comparison of different methods

As a final part of the analysis, the presented multiple methodology is applied to the Aosta Valley to estimate the technical potential of solar photovoltaic resources, which are recognized as the most commercially convenient and developed renewable technology. It's possible to estimate for the amount of electricity potentially generated by the photovoltaic panels installed on roofs; it has been also provided an estimation of the nominal power to install to produce the calculated potential energy.

The self-sufficiency is also estimated and this parameter is defined as a percentage of annual electricity demand covered by the photovoltaic technical potential divided by the actual electricity production from fossil fuels, therefore it represents the amount of fossil energy that can be potentially replaced by the renewable energy produced by PV panels on roofs.

The numerical results of each method are show by means of a table, a little exception has to be made for the machine learning techniques. Even if they can be easily tuned to specific parameters for the identification of factors useful for our research, perhaps thanks to the help of external skills such as data science and analyst, at this moment they are limited to a qualitative analysis. As a result, there won't be displayed kilowatt hours per year or m^2 of rooftop building availability, but the accuracy of the methods used, the segmentation ratio and the ability to locate the photovoltaic panels in the first place, which QGIS or other software seen in this thesis cannot do, except with a paid license.

Here following the factor affecting the analysis and their values:

- Hg = 1,588 [kWh/m^2/year] [62]
- Efficiency = $\eta STC \cdot PR$ = 0.73 * 0.16 = 0.1186
- Equivalent hours = 992 hours/year
- Reduction factor = 0,26

Method	**Resolution**	**Suitable PV surface**	**Annual Electricity Production**	**Potential PV power**	**Self-Sufficiency**
	[m^2]	[m^2]	[GWh/year]	[MW]	%
Direct method	0.5 x 0.5	3168	0,60	0,60	14%
Direct method	2.0 x 0.5	5160	0,97	0,98	23%
Grafical modeler	0.5 x 0.5	5570	1,05	1,06	25%
Grafical modeler	2.0 x 2.0	9270	1,74	1,76	41%
Indirect Method	-	12408	2,33	2,35	55%
		IoU	Accuracy		
ML	0,25 x 0,25	0,93	0,87		

Table 10: Photovoltaic technical potential result on Aosta roofs

In the table above the result of our dissertation are shown. We can immediately notice that the annual electricity demand and so the potential power change a lot considering a method instead of another. As we could expect, the indirect method, which has been used in many articles we analysed, is the one which approximates by excess and overvalue the estimations, as well as the one that does not differentiate the resolution of one method rather than another, in short, it is the last resource when very few data are available.

In contrast, we can say that the other estimations are very close to each other and seems to be very accurate. For what concern the direct method, we can see that the suitable PV surface for installing solar panels, geographical potential, is 3,2 km^2 for 0.5 x 0.5 and 5,2 km^2 for 2.0 x 2.0 resolution. This difference is due to the fact that working at a higher resolution, the finding of more obstacles that may not be detected by lower resolutions cause the reduction of the whole total available rooftop area.

Instead, considering the graphical modeler, we can say that it seems to be more precise (this assertation can be proved or denied by the machine learning techniques) as it has the ability of filtering and buffering the roof area as explained in the relative chapter. It can potentially unify and put together littler areas (below 7 sqm), which in other cases are eliminated, in order to create bigger spot without obstacle; for the model of 0.5 x 0.5 resolution the found area is increased of about 176% while for the 2.0 x 2.0 resolution it increases of about 180%.

Regarding the applicability of the ML models, it can be said that they are very accurate for solar irradiation simulations of terrain models, where shading is generally less extreme. It is worth noting that both proprietary and open-source GIS platforms have their own tools that can calculate solar irradiation for DEM and DSM. An advantage of the trained ML models is that they are platform-agnostic, hence, they can be incorporated in existing GIS packages to dramatically decrease calculation times, since less input are required. In this case, we managed to find only some of the required input and qualitative evaluated some of them, cause the tuning should be improved with the help of more competent people. As one of the biggest macro sector trend, machine learning has the ability to be used almost by everyone but manipulated at most by very few people.

Anyway, a quantitative evaluation can be done, as the IoU index is in some way a performance value, meant as a capability of the model to recognize and correctly mask an image. A value of 0.93, which can go up to 97 according to the developer tests, is an estimation of the "real" area identified by the algorithm meaning that the 93% of the rooftop area is really suitable, so the 93% of the annual electricity production and potential power can be reached: leading to and error on 7%, much lower than the 11,8 found in the literal review and that was our hot spot to be finally demonstrated.

Chapter 6

6. Conclusion and future developments

In this section the possible improvements and recommendations for future work are mentioned. This thesis is focused on Aosta Valley and for it all the data are public and can be accessible directly from geoportals. However, if the location of the desired research area changes, the administrative procedure to get data might change as well and it should be formally requested for the research purpose. It takes time to receive the exact data one is looking for, and even if the data reliability is high, some improvements can be necessary in assistance.

It was seen that increasing the data resolution enhanced and improved the accuracy of the solar irradiance results in the last section. Results of finer resolution 0.5m are clearly more reliable with respect to the coarse resolution of 2m raster results. Most of the papers read during the literature review mentioned high resolution difference and its effect on the results. So, this improvement has been done and the next step might be to apply this method for larger areas. Right technical equipment should be supplied for this purpose since computational time is too high for doing it on our personal computers, such as some more complex project found on GitHub.

Machine Learning techniques and, in particular, roof segmentation can be done more precisely and the module placement can be optimized as a future work. Based on this improvement, the economic potential of the rooftops can be estimated as well, since with the methodology we used, we observed only the physical, geographical and technical potentials of Aosta Valley.

Another key point of improvement is the estimation of the energy output of a rooftop solar PV system, considering the number of panels which could fit on each roof had to be estimated. We manage to do it but with a very low-resolution projects as there were just the first drafts of the script. This is a very difficult problem as it essentially requires finding the maximum number of rectangles that can fit inside any given shape. There are no easy ways to implement a solution within QGIS to find an accurate estimate for every rooftop area and so a more statistical approach is strongly needed.

However, there is still significant potential for future research to create ML models and hybrid models that accelerate calculations for daytime lighting, microclimate, thermal comfort and energy use in the building; this was only one of the possible applications as the biggest challenge is not to meet the annual production of electricity, but rather to do it in an efficient, reliable, sustainable and secure manner, trying to exploit local resources aiming to the self-sufficiency energy point of view.

Bibliography:

[1] I. Ernmenta and L. P. A. Nel, *Climate Change 2014 Synthesis Report*. .

[2] C. Change, *Climate Change 2022 Mitigation of Climate Change Summary for Policymakers*. 2022.

[3] UNITED NATIONS, *CLIMATE CHANGE 2021*. .

[4] IEA, "World Energy Outlook 2020," 2020.

[5] Global Initiative of the United States, "The Sustainable Development Goals," no. December, 2015.

[6] European Commission, *Proposed Mission : A Climate Resilient Europe*. .

[7] Eurpoean Market Observatory for Energy, "Quarterly Report 2021, on European Electricity Markets," vol. 13, no. 4, 2020.

[8] D. general for E. European Commission, "Annual Activity Report," 2022.

[9] G. S. E.- GSE, "Copertina RAPPORTO STATISTICO 2017."

[10] Gestore Servizi Energetici - GSE, "RAPPORTO STATISTICO 2020 ENERGIA," 2020.

[11] Gestore Servizi Energetici - GSE, "Rapporto statistico Solare Fotovoltaico 2021."

[12] M. M. Hoogwijk, *ON THE GLOBAL AND REGIONAL POTENTIAL OF RENEWABLE ENERGY*, no. november 1974. 2004.

[13] J. A. D. Deceased and W. A. Beckman, *of Thermal Processes Solar Engineering*. .

[14] S. Izquierdo, M. Rodrigues, and N. Fueyo, "A method for estimating the geographical distribution of the available roof surface area for large-scale photovoltaic energy-potential evaluations," vol. 82, pp. 929–939, 2008.

[15] A. Walch, R. Castello, N. Mohajeri, and J. Scartezzini, "Big data mining for the estimation of hourly rooftop photovoltaic potential and its uncertainty," *Appl. Energy*, vol. 262, no. November 2019, p. 114404, 2020.

[16] K. Mainzer, S. Killinger, R. Mckenna, and W. Fichtner, "Assessment of rooftop photovoltaic potentials at the urban level using publicly available geodata and

image recognition techniques," *Sol. Energy*, vol. 155, pp. 561–573, 2017.

[17] T. Hong, M. Lee, C. Koo, and J. Kim, "Estimation of the available rooftop area for installing the rooftop solar photovoltaic (PV) system by analyzing the building shadow using Hillshade analysis," *Energy Procedia*, vol. 88, pp. 408–413, 2016.

[18] E. Jadraque, J. Alegre, G. Martı, and J. Ordo, "Analysis of the photovoltaic solar energy capacity of residential rooftops in Andalusia (Spain)," vol. 14, no. 2010, pp. 2122–2130, 2020.

[19] L. K. Wiginton, H. T. Nguyen, and J. M. Pearce, "Computers , Environment and Urban Systems Quantifying rooftop solar photovoltaic potential for regional renewable energy policy," *Comput. Environ. Urban Syst.*, vol. 34, no. 4, pp. 345–357, 2010.

[20] L. Bergamasco and P. Asinari, "Scalable methodology for the photovoltaic solar energy potential assessment based on available roof surface area : Application to Piedmont Region (Italy)," *Sol. Energy*, vol. 85, no. 5, pp. 1041–1055, 2011.

[21] L. Bergamasco and P. Asinari, "Scalable methodology for the photovoltaic solar energy potential assessment based on available roof surface area : Further improvements by ortho-image analysis and application to Turin (Italy)," *Sol. Energy*, vol. 85, no. 11, pp. 2741–2756, 2020.

[22] M. C. Brito, N. Gomes, T. Santos, and J. A. Tenedo, "Photovoltaic potential in a Lisbon suburb using LiDAR data," vol. 86, pp. 283–288, 2012.

[23] A. Strzalka, N. Alam, E. Duminil, V. Coors, and U. Eicker, "Large scale integration of photovoltaics in cities," *Appl. Energy*, vol. 93, pp. 413–421, 2012.

[24] "Lukacˇ et al. - 2013 - Rating of roofs' surfaces regarding their solar po.pdf." .

[25] J. Peng and L. Lu, "Investigation on the development potential of rooftop PV system in Hong Kong and its environmental bene fi ts," *Renew. Sustain. Energy Rev.*, vol. 27, pp. 149–162, 2013.

[26] J. A. Jakubiec and C. F. Reinhart, "A method for predicting city-wide electricity gains from photovoltaic panels based on LiDAR and GIS data combined with hourly Daysim simulations," *Sol. Energy*, vol. 93, pp. 127–143, 2013.

[27] T. Hong, C. Koo, J. Park, and H. S. Park, "A GIS (geographic information system) -based optimization model for estimating the electricity generation of the rooftop PV (photovoltaic) system," *Energy*, vol. 65, pp. 190–199, 2014.

[28] "Lukač et al. - 2014 - Buildings roofs photovoltaic potential assessment .pdf." .

[29] R. Singh and R. Banerjee, "Estimation of rooftop solar photovoltaic potential of

a city ScienceDirect Estimation of rooftop solar photovoltaic potential of a city," *Sol. ENERGY*, vol. 115, no. May, pp. 589–602, 2015.

[30] K. Fath, J. Stengel, W. Sprenger, H. Rose, F. Schultmann, and T. E. Kuhn, "ScienceDirect A method for predicting the economic potential of (building-integrated) photovoltaics in urban areas based on hourly Radiance simulations," *Sol. Energy*, vol. 116, pp. 357–370, 2015.

[31] D. Assouline, N. Mohajeri, and J. Scartezzini, "Quantifying rooftop photovoltaic solar energy potential : A machine learning approach," *Sol. Energy*, vol. 141, pp. 278–296, 2017.

[32] T. Hong, M. Lee, C. Koo, K. Jeong, and J. Kim, "Development of a method for estimating the rooftop solar photovoltaic (PV) potential by analyzing the available rooftop area using Hillshade analysis," *Appl. Energy*, vol. 194, pp. 320–332, 2017.

[33] D. Assouline, N. Mohajeri, and J. Scartezzini, "Large-scale rooftop solar photovoltaic technical potential estimation using Random Forests," *Appl. Energy*, vol. 217, no. February, pp. 189–211, 2018.

[34] N. Mohajeri, D. Assouline, B. Guiboud, and A. Bill, "A city-scale roof shape classi fi cation using machine learning for solar energy applications," *Renew. Energy*, vol. 121, pp. 81–93, 2018.

[35] F. Mansouri, K. James, B. Dan, J. Locke, and S. Paul, "Evaluating solar energy technical and economic potential on rooftops in an urban setting : the city of Lethbridge , Canada," *Int. J. Energy Environ. Eng.*, vol. 10, no. 1, pp. 13–32, 2019.

[36] Song, "An Approach for Estimating Solar Photovoltaic Sensing Images," pp. 1–14.

[37] Z. Huang, T. Mendis, and S. Xu, "Urban solar utilization potential mapping via deep learning technology : A case study of Wuhan , China," *Appl. Energy*, vol. 250, no. May, pp. 283–291, 2019.

[38] H. Alhammami and H. An, "Techno-economic analysis and policy implications for promoting residential rooftop solar photovoltaics in Abu Dhabi , UAE," *Renew. Energy*, vol. 167, pp. 359–368, 2021.

[39] K. Bódis, I. Kougias, A. Jäger-waldau, N. Taylor, and S. Szabó, "A high-resolution geospatial assessment of the rooftop solar photovoltaic potential in the European Union," *Renew. Sustain. Energy Rev.*, vol. 114, no. August, p. 109309, 2019.

[40] H. G. Lopez-ruiz, J. Blazquez, and M. Vittorio, "Assessing Residential Solar Rooftop Potential in Saudi Arabia Using Nighttime Satellite Images : A Study

for the City of Riyadh," no. June, pp. 1–24, 2019.

[41] B. Fina, H. Auer, and W. Friedl, "Cost-optimal economic potential of shared rooftop PV in energy communities : Evidence from Austria," *Renew. Energy*, vol. 152, pp. 217–228, 2020.

[42] T. Schunder, D. Yin, S. Bagchi-sen, and K. Rajan, "Remote Sensing Applications : Society and Environment A spatial analysis of the development potential of rooftop and community solar energy," *Remote Sens. Appl. Soc. Environ.*, vol. 19, no. June, p. 100355, 2020.

[43] T. N. C. De Vries, J. Bronkhorst, M. Vermeer, J. C. B. Donker, and S. A. Briels, "A quick-scan method to assess photovoltaic rooftop potential based on aerial imagery and LiDAR," *Sol. Energy*, vol. 209, no. June, pp. 96–107, 2020.

[44] A. Gomez-exposito, A. Arcos-vargas, and F. Gutierrez-garcia, "On the potential contribution of rooftop PV to a sustainable electricity mix : the case of Spain University of Seville , Spain," vol. 1, pp. 1–37.

[45] J. R. Nelson and T. H. Grubesic, "The use of LiDAR versus unmanned aerial systems (UAS) to assess rooftop solar energy potential," *Sustain. Cities Soc.*, vol. 61, no. June, p. 102353, 2020.

[46] T. G. Reames, "Energy Research & Social Science Distributional disparities in residential rooftop solar potential and penetration in four cities in the United States," *Energy Res. Soc. Sci.*, vol. 69, no. June, p. 101612, 2020.

[47] V. M. Phap, N. Thi, T. Huong, P. T. Hanh, P. Van Duy, and D. Van Binh, "Assessment of rooftop solar power technical potential in Hanoi city , Vietnam," *J. Build. Eng.*, vol. 32, no. June, p. 101528, 2020.

[48] R. Singh, "Approximate rooftop solar PV potential of Indian cities for high-level renewable power scenario planning," *Sustain. Energy Technol. Assessments*, vol. 42, no. April, p. 100850, 2020.

[49] "index @ www.ladybug.tools." .

[50] "information-technology-gis @ www.oswegoil.org." .

[51] F. Lindberg *et al.*, "Environmental Modelling & Software Urban Multi-scale Environmental Predictor (UMEP): An integrated tool for city-based climate services," vol. 99, 2018.

[52] "13ce01176315bcb90859b1f7be659666bdd7d3cb @ 3dmetrica.it." .

[53] B. Mahesh, "Machine Learning Algorithms - A Review," no. January 2019, 2020.

[54] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, “Solar forecasting methods for renewable energy integration,” *Prog. Energy Combust. Sci.*, vol. 39, no. 6, pp. 535–576, 2013.

[55] Badescu, “Modeling Solar Radiation at the Earth ’ s Surface Recent Advances.”

[56] “index @ github.com.” .

[57] “index @ geoportale.regione.vda.it.” .

[58] “photovoltaic-detection @ github.com.” .

[59] “riccardocadei @ github.com.” .

[60] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” pp. 1–8.

[61] “rooftop-detection-python @ github.com.” .

[62] SolarGis, “Yield assessment of the photovoltaic power plant,” vol. 51, no. February, pp. 1–4, 2018.