Analysing the location choice of solar fields in the Netherlands

by

Luuk Verdonk

2702088

Bachelor thesis

Aarde, Economie & Duurzaamheid

Faculty of Science

Vrije Universiteit Amsterdam

Supervised by dr. Eric Koomen

30-06-2023

Abstract

The Netherlands is currently shifting from fossil fuels to renewable energy sources. This transition is needed to mitigate climate change and maintain the country's position as an energy hub. Solar energy is a key component of this transition, and the government has plans to implement solar panels not only on roofs but also on agricultural fields and unused industrial estates. This research aims to determine which factors influence the location choice of solar fields.

This is done through a literature review and a logistic regression analysis. In the literature review, three categories of factors have been determined: environmental, technical, and socio-economic. The environmental factors include solar radiation and the land use type. Technical factors are the distance to roads and to the electricity grid. The socio-economic factor used in this research is the distance to urban areas. Region fixed effects and land use fixed effects are included as well.

To find out how these factors affect the suitability of sites, a logistic regression has been conducted. The factors determined in the literature review are the independent variables. The distance to the electricity grid has the highest influence on the location choice, followed by the global horizontal irradiance, the distance to urban areas, and the distance to roads. Zeeland has the highest chance of occurrence due to region fixed effects. Semi-built up land use has the highest chance due to the land use fixed effect.

With the regression results, a probability map is made showing where solar fields have a higher chance of being developed. Policymakers can use this map to see where landscapes might change and where policies are needed.

Table of contents

1. Introduction

The landscape in the Netherlands is constantly changing. About 1,000 years ago, these changes came through deforestation and peat extraction to provide energy for the population. Later on, the transition to wind energy and later fossil fuels caused new landscape changes (De Jong & Stremke, 2020). Currently, the landscape in the Netherlands is changing again. This is because the Netherlands is going through a new energy transition: from fossil fuels to renewable energy. This new transition is becoming more important, as shown in the recently published IPCC synthesis report (IPCC, 2023). The IPCC is an intergovernmental body that provides scientific knowledge about climate change. The report shows that the possibilities of limiting global warming to 1.5 degrees are rapidly decreasing. Furthermore, if the 1.5-degree climate target is not met, it is important to limit further warming as much as possible; every tenth of a degree makes the consequences less severe (IPCC, 2023).

For this reason, the Dutch government has stated in the coalition agreement that it wants to pursue the Paris Climate Agreement (Rijksoverheid, 2021). To achieve this, the Dutch government has made plans to switch from fossil fuels to renewable energy faster. This is not only important for the climate but also for the hub position the Netherlands currently has due to, for example, location and infrastructure (Netbeheer Nederland, 2023). With this hub, the Netherlands contributes to an affordable and reliable energy supply for Europe. From an economic perspective, this position is important to maintain. Because of the plans of the Dutch government, the role of fossil fuels is rapidly diminishing. If the country wants to maintain its hub position, it has to make choices to develop new networks and energy sources. One way to make such a switch is by using solar energy. The Dutch government wants to implement solar panels not only on roofs but also on agricultural fields and unused industrial estates, so-called solar fields.

The implementation of these solar fields will change the land use and landscape of the Netherlands. These changes happen not only in the field itself but also in other places, e.g., the power grid. There are many similar suitable places for solar fields, but only a few are actually used for solar fields. On the one hand, this is due to limited demand for solar fields; if the Netherlands were covered entirely in solar fields, the total energy demand would be exceeded six times (Zonne-energie, z.d.). On the other hand, which field will be chosen can depend on factors like distance to powerlines and visual pollution. To understand where the solar field will be developed, it is important to know the effect these factors have on the suitability of sites. This is relevant for policymakers because they can make a more accurate prediction of where the solar fields are likely to be developed.

To find out the effect these factors have on the suitability of solar fields, this research aims to answer the following research question: *Which spatial factors determine the location of solar fields in the Netherlands?* To answer the research question, there are four sub-questions:

- What are the environmental factors that determine the suitability of locations for solar fields?
- What are the technical factors that determine the suitability of locations for solar fields?
- What are the socio-economic factors that determine the suitability of locations for solar fields?
- What is the relative importance of these factors at the sites where solar fields were developed?

This research is conducted through a literature review, a data analysis using a geographical information system (GIS), and a logistic regression analysis. To answer the first three sub-questions, there will first follow a background on solar fields through a literature review in Chapter 2. Chapter 3 contains the methodology for the GIS research and regression analysis. In Chapter 4, the results are shown. Chapters 5 and 6 contain, respectively, the discussion and the conclusion.

2. Background

The first solar field in the Netherlands was built in 2004 (Dröes & Koster, 2021). In 2019, there were more than eighty solar fields that produced more than 1 MW (separately) (Van Dam et al., 2019). More than 200 solar fields are scheduled to be built in the coming years (Zon op Kaart, 2023). Figure 1 shows the number of solar fields producing more than 15 kWp in the Netherlands over the years.

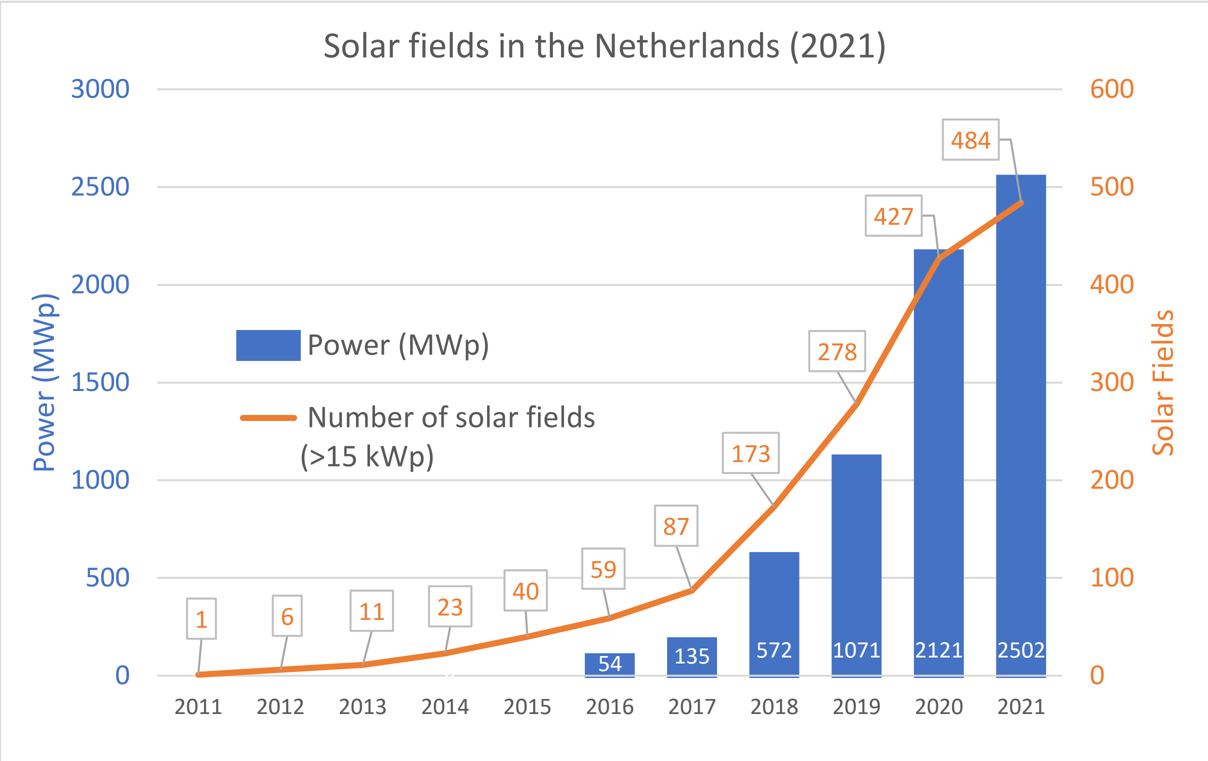


Figure 1. *Solar fields in the Netherlands that have a higher power generation than 15 kWp (until September 2021) (Rolf, 2021). Orange shows the number of solar fields, and blue shows the power generation.*

As shown in Figure 1, there has been rapid growth in the past ten years. To know where solar fields are likely to be developed in the future, it is important to look at the ones that already exist. This chapter consists of a literature review about the factors that determine the locations of solar fields. Based on research from Mierzwiak & Calka (2017), the factors are divided into three subgroups: environmental factors in Section 2.1, technical factors in Section 2.2, and socio-economic factors in Section 2.3.

2.1 Environmental factors

There are several environmental factors that determine the suitability of sites for solar fields. First, the environmental factors that make a site more suitable will be mentioned. This type of factor is called a proximity factor. Then, the factors that form a restriction will be mentioned.

Research from Elboshy et al. (2021) shows that solar radiation is the most important factor because it determines the electricity generation from the solar panels. More solar radiation makes a site more suitable. This research uses the global horizontal irradiance (GHI). GHI means all solar radiation that reaches the earth’s surface, which is relevant for electricity generation by solar panels (The World Bank, 2020).

Another factor is the slope (Perpiña Castillo et al., 2016). Variability in elevation, surface orientation, and shadows caused by topographic features can have a strong impact on the amount of solar irradiance reaching the panels. Therefore, an orientation towards the north is unsuitable. The maximum slope that is considered suitable is 30 percent. However, the maximum slope in the Netherlands is 22 percent (AHN, 2023), and these slopes are only found locally in Limburg. Thus, slope and orientation are not used in this research.

Because a solar field requires a large location, it may have negative effects on the biodiversity and landscape, making some places unsuitable. Thus, solar fields cannot be placed in areas with monuments of World Heritage, archaeological zones, areas with landscape protection, Natura 2000 areas, or protected forests (Baltas & Dervos, 2012). There are some solar fields on inland waters, e.g., on drinking water reservoirs and small lakes (RVO, 2023). Inland waterbodies that are unprotected are therefore considered suitable places.

Another criterion is the current land use. Research from Perpiña Castillo (2016) shows that the installation of a solar field is not possible in areas that are currently being used as built-up areas or areas with a marshy soil type (wetlands). Wetlands have limited accessibility and unstable ground. Built-up areas have a high degree of existing development, meaning there is no place for solar fields. The ideal location is an undeveloped area with short vegetation types to avoid the vegetation blocking the solar radiation. Thus, a forest can be a suitable location, but parts of it might have to be cut down. Industrial areas are suitable when there are no buildings on them.

2.2 Technical factors

The first technical factor is the accessibility of the solar field. It is important to look at the existing infrastructure because the solar field has to be constructed and maintained. A maximum of three thousand metres from existing main roads is considered suitable (Carrión et al., 2008; Perpiña Castillo et al., 2016).

Another technical factor is the electricity grid. A solar park needs to be connected to a substation to transport the generated electricity to the users. The maximum distance to the existing high voltage electricity grid should be 50 kilometres (Elboshy et al., 2021). For the distances to roads and the electricity grid, it is more suitable when the distance is smaller. However, solar fields cannot be placed on the road or railway itself.

2.3 Socio-economic factors

The social factor used in this research is the population density. Solar fields cause visual pollution for people living nearby. The average size of solar fields has increased from two hectares for the first solar fields to twenty hectares in 2019 (Van den Berg & Tempels, 2022), making them more visible for inhabitants. Therefore, sites further away from urban areas are more suitable. This way, the local community will have less resistance to the construction of a solar field near their living areas (Perpiña Castillo et al., 2016).

An economic factor is the productivity of the land. When a site has high agricultural productivity, it is not desirable to develop a solar field on these locations (Baltas & Dervos, 2012). However, research from Spruijt (2015) shows that solar fields can be even more profitable than croplands (e.g., winter wheat and potatoes). Agricultural land is therefore not excluded from this research.

3. Methodology

This chapter describes the methodology of the statistical research. It consists of the methodology for collecting and preparing the data (Sections 3.1 and 3.2). The preparation is needed for the logistic regression analysis. The method for the regression analysis is explained in Section 3.3. After the regression is finished, a probability map can be made using GIS. The methodology for this map is explained in Section 3.4.

3.1 Data collection

After determining the factors in Chapter 2, datasets containing map layers can be collected and prepared for the regression analysis. The preparation will be done in QGIS, a software programme to transform spatial data into maps.

Keeping the factors from Chapter 2 in mind, solar fields in the Netherlands will be built mainly on what is now agricultural land. Because of restrictions in natural areas and because the Netherlands is densely populated, most of the suitable land is agricultural land. If this is the case, fewer processes have to be conducted in GIS, and fewer steps in the regression have to be taken because only one layer would be needed. To check whether this is true, the current solar fields in the Netherlands are compared with land use in 2017. This way, it can be investigated whether solar fields are indeed placed on agricultural land or whether other types of land use are relevant as well.

The source of the data used for this analysis is the 'Basisregistratie Topografie' from the Cadastre. The Cadastre has a dataset that contains all existing solar fields in the Netherlands. The dataset was created in February 2023. Until this date, the total area of solar fields was approximately 2,750 hectares. The solar fields are on average 6.2 hectares in size. The smallest solar field has a size of 0.0238 hectares (238 m^2). The biggest solar field is 93.38 hectares. A map with the existing solar fields is shown in Figure 2 on the next page. The layer with land use in 2017 is made by CBS. Both layers have the same resolution and projection (28992 – Amersfoort). More details about the data are shown in Appendix A.1.

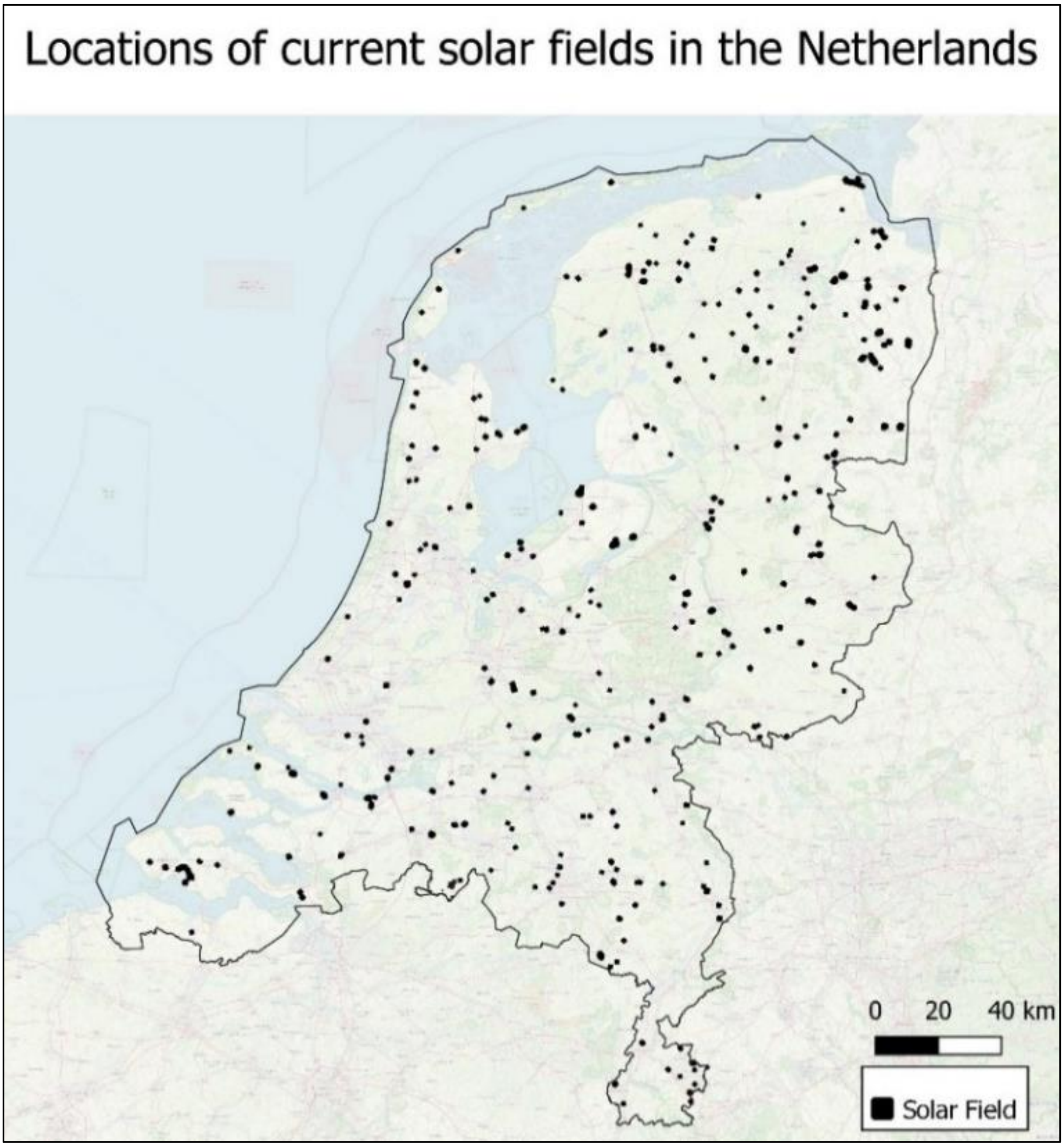


Figure 2. Locations of the current solar fields in the Netherlands. The size of the locations is enlarged on the map to enhance visibility (Kadaster, 2023).

After collecting the data, the layer with solar fields can be overlaid with a land-use map to infer the most common type of land they are located on. This results in an attribute table showing the land use on which each solar field is placed. The results are shown in Figure 3 on the next page.

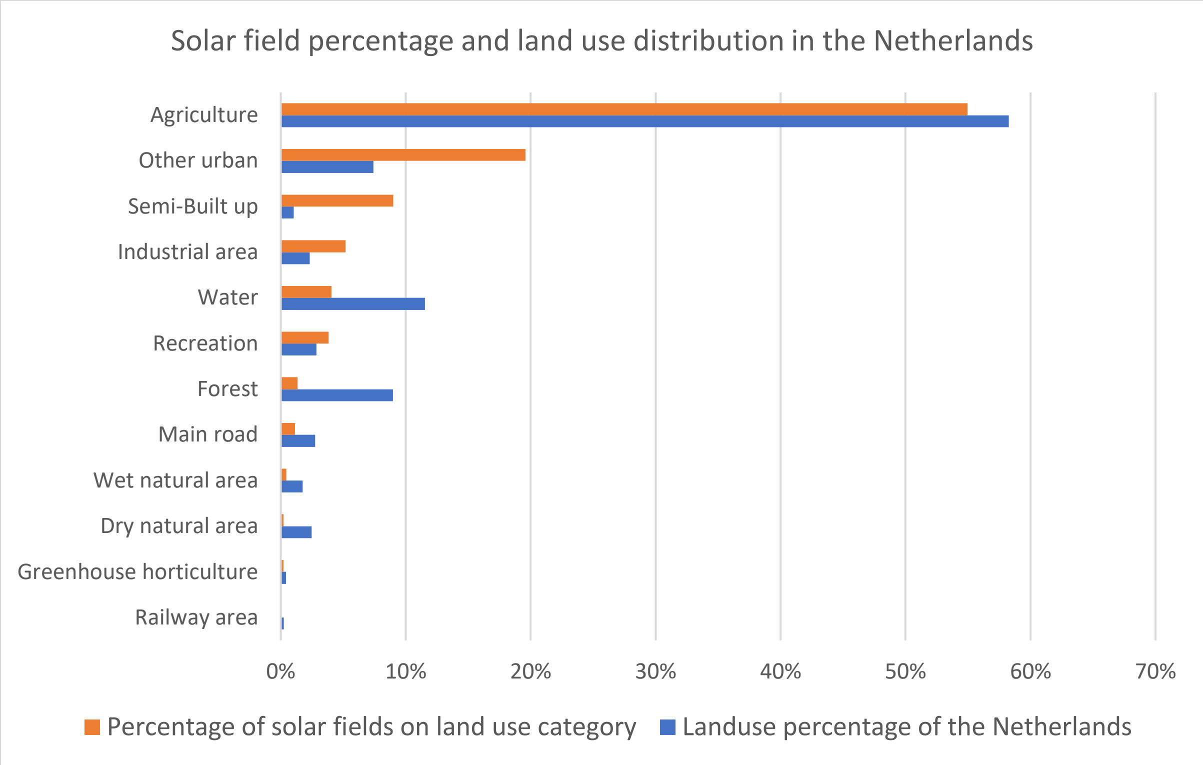


Figure 3. *The percentage of solar fields per land use category in 2017 (orange) and the percentage of land use in the Netherlands (blue) (Kadaster, 2023; CBS, 2017).*

Figure 3 shows that the largest percentage of solar fields is developed on agricultural land. However, this is also the largest category of land use. The categories 'other urban', 'semi-developed', 'recreation', and 'industrial' will be used as suitable categories as well, because there is a relatively larger percentage of solar fields on these categories. Figure 3 also shows that there are some solar fields in natural areas. Therefore, these categories will be used as well. However, not all sub-categories will be used. For example, a sub-category of semi-built up is 'construction site', which is not suitable for solar fields.

The category 'other urban' consists of residential buildings, public amenities, social-cultural facilities, and hospitality buildings. From this category, only the sub-category public amenities will be used. This is because solar fields that are constructed before 2017 are classified as public amenities. When excluding this sub-category, the solar fields before 2017 would be deleted from the dataset, and they would not contribute to the outcome of the regression analysis. Appendix A.2 shows the (sub)categories of the CBS dataset that are used in this research. Appendix A.2.1 shows the categories that are excluded.

To exclude restricted areas (e.g., Natura 2000 areas and large waterbodies), other datasets are used because the CBS does not distinguish between protected and non-protected natural areas. In addition, for the land use 'water', another dataset is used. The percentage of solar fields on water is low because all types of water are used in the calculation (including sea, sea-arms, and the Ijsselmeer). Therefore, a more detailed dataset is used for this category in the regression analysis. Similarly, another dataset is used for 'main roads'. The high number of solar fields is caused by wrong classification; when solar fields are located between two roads, they are sometimes classified as main road.

For the proximity factors (irradiance and distances to the electricity grid, urban areas, and roads), other datasets will be used as well. A more detailed description of the other datasets is shown in Appendix A.3. The table shows the source, year, and type of data used for each factor.

3.2 Data preparation

After the necessary data has been collected and cleansed, it needs to be further prepared to be usable for the regression. The methods for the preparation are shown in flowcharts. A flowchart shows the steps taken in the GIS. The flowcharts are shown in Appendix B.1.

The first step is to make sure every layer is projected with the right resolution and projection (28992 – Amersfoort). This ensures that layers can be overlaid correctly. Then, the data needs to be converted into a CSV table to be usable for the logistic regression. This conversion requires some other steps first.

The vector layers need to be converted to raster layers because GIS tools for this research often only work with raster formats. The raster layers will have grid cells of 10 m^2. Then, for the proximity factors, three maps that show the distances to roads, urban areas, and the electricity grid will be made. These maps are called proximity maps. The proximity maps are shown in Appendix C.

The next step is to make pixel points for each pixel inside the study area. By generating pixel points inside polygons, the polygon data can be converted into a point-based representation. These points will form the observations for the table used in the regression analysis. The polygon will first be a province of the Netherlands to test the tool with a small batch of data. Hereafter, the polygon will be the border of the Netherlands for making observations for the whole country.

In the next step, the pixel points will get the attributes of the factor layers. These layers will be the proximity factors and the solar fields layer. A layer with provinces of the Netherlands will be added as well to account for the region fixed effects. Region fixed effects account for spatial heterogeneity that is not included in the dataset. The regional-specific effects may impact the relationship between the independent variables and the dependent variable. Different regions may have unique characteristics, policies, or socioeconomic factors that can affect the occurrence of solar fields. Including the region fixed effects increases the reliability of the regression analysis. The region fixed effects are calculated per province. The chance of occurrence will be calculated compared to Drenthe, as that is the base province.

In addition, the land use type of each observation is added to the dataset to account for the fixed effects of the land use. As mentioned in Chapter 2, land use types have different effects on the occurrence of solar fields. Adding the fixed effects will explain this relationship and make the analysis more complete. The base land use type for this analysis is industrial. As mentioned in Section 3.1, solar fields developed before 2017 are classified as a sub-category of 'other urban'. This can lead to a bias; more solar fields in this category will increase the coefficient. However, Figure 1 shows that the number of solar fields before 2017 is minor compared to the following years. Hence, the potential bias is small.

The result of the steps above is a new attribute table with each pixel point, which contains information about every factor. This table can be used for the logistic regression analysis. The variables that are used in the regression are described in Table 1.

Table 1. *Descriptive statistics of the independent variables that will be used in the logistic regression analysis (The World Bank, 2020; CBS, 2017; PDOK, 2019).*

Independent variable	**Minimum**	**Maximum**	**Mean**	**Standard deviation**
Global Horizontal Irradiance (kWh/m^2)	997.498	1126.431	1045.487	24.65
Distance to the electricity grid (km)	0	40.80	4.13	4.257
Distance to the nearest road (km)	0.107	15.09	0.49	0.479
Distance to the nearest urban area (km)	0	16.72	1.10	0.848

3.3 Logistic regression analysis

After creating the observation table with GIS, a logistic regression analysis will be conducted to identify the relative importance of the factors at the sites where solar fields were developed. A logistic regression is a statistical technique used to model the relationship between a binary dependent variable and one or more independent variables. Binary means that the variable can either be yes or no (1 or 0). In this case, the binary dependent variable is the presence or absence of a solar field at a given location. The independent variables are the factors that may affect the placement of solar fields, as identified in Chapter 2. In short, this form of regression analysis estimates the probability of the occurrence of a certain phenomenon, in this case, the occurrence of solar fields.

The data from the CSV table will be used to develop a logistic regression model that will determine the coefficients for each independent variable. The coefficients will indicate the strength and direction of the variables' influence on the placement of solar fields. However, to compare the variables, the standardised coefficients need to be used. A standardised coefficient indicates how many standard deviations the occurrence of solar fields changes for a change of 1 standard deviation in the independent variable when all other variables stay constant. The standardised variables are needed to compare the relative influence of different variables, regardless of their units.

The logistic regression analysis will be conducted using Stata, a software programme for data analysis and statistical modelling. The analysis will be performed in several steps, including data preparation, model building, and model evaluation. The data preparation phase is explained in Section 3.2. The model-building phase consists of the logistic regression itself: estimating the coefficients for each independent variable. In the model evaluation phase, the model's accuracy will be evaluated. The flowchart for steps taken in the logistic regression analysis is shown in Appendix B.2.

3.4 Probability map

After conducting the regression analysis, a probability map can be made with the coefficients of each of the independent variables. Based on the location of solar fields that have been developed in the past, a probability map can be made that shows where solar fields are likely to occur in the future. To do this, the coefficients and the constant will be filled in using the formula shown in Appendix B.3. With the formula, the probability of the occurrence of a solar field will be calculated based on the independent variables and the fixed effects. The restricted and unsuitable places are used for the probability map as well. To make suitable land use 1 and unsuitable land use 0, the land use categories will be reclassified first. With the calculated probability, the map can be made using GIS. The flowchart for the probability map is shown in Appendix B.3.

4. Results

In this chapter, the results of the logistic regression are shown in Section 4.1. The results will show the relative importance each factor has on the location choice of solar fields (sub-question 4). The probability map that can be made with the results is shown in Section 4.2.

4.1 Logistic regression analysis

Table 2 shows the results of the regression. The table shows the coefficients of the independent variables. The coefficients indicate the strength and direction of the influence that the factors have on the occurrence of solar fields. The standardised coefficients are shown as well. Additionally, the coefficients of the provinces and land use types are shown to account for the region and land use fixed effects.

Table 2. *Results of the logistic regression on the occurrence of solar fields in the Netherlands based on four independent variables and fixed effects. The used sources are shown in Appendix A.3.*

Occurrence of solar fields	Coefficient	Standardised coefficient	(Standard error)
Independent variables			
Global Horizontal Irradiance (kWh/m^2)	-0.0160 ***	-0.395	(-27.72)
Distance to the nearest urban area (km)	-0.220 ***	-0.186	(-22.71)
Distance to the electricity grid (km)	-0.108 ***	-0.461	(-45.74)
Distance to the nearest road (km)	-0.191 ***	-0.092	(-8.05)
Region fixed effects			
Drenthe (base)	0		
Flevoland	0.576 ***		(15.84)
Friesland	-0.169 ***		(-6.16)
Gelderland	-0.479 ***		(-16.43)
Groningen	0.586 ***		(27.42)
Limburg	-0.0985 *		(-2.15)
Noord-Brabant	-0.375 ***		(-10.30)
Noord-Holland	0.0205		(0.50)
Overijssel	-0.382 ***		(-13.64)
Utrecht	-0.354 ***		(-8.60)
Zeeland	0.772 ***		(14.29)
Zuid-Holland	0.0816		(1.84)
Land use fixed effects			
Industrial (base)	0		
Semi-built up	2.464 ***		(54.44)
Agricultural	0.0435		(1.27)
Waterbodies	0.745 ***		(14.19)
Nature	-1.595 ***		(-31.73)
Other urban	1.907 ***		(45.08)
Recreational	0.166 ***		(3.86)
Constant	10.25 ***		(17.52)
Observations	33,065,741		
Pseudo-R^2	0.0632		

Notes: *$p < 0.05$, ***$p < 0.001$

The independent variables all have a negative coefficient. This means that when they increase, the chance that a solar field will be developed decreases. For example, when the GHI increases by 1 kWh/m^2, the occurrence of solar fields decreases by 0.0160. This also applies to the distances to the nearest urban area, electricity grid, and road. When the distances to these factors increase by 1 kilometre, the chance of the occurrence of solar fields decreases by, respectively, 0.220, 0.108, and 0.191. The three stars next to the coefficients indicate that there is a significant relationship between the independent variables and the occurrence of solar fields. The p-value is less than 0.001. Thus, the probability that the relationship between the occurrence of solar fields and the independent variables is random is less than 0.1%.

The standardised coefficients show how many standard deviations the occurrence of solar fields changes for a change of 1 standard deviation in the independent variables. When the distance to the electricity grid increases by 1 standard deviation, the occurrence of solar fields decreases by -0.461. The GHI, distance to the nearest urban area, and distance to the nearest road follow with, respectively, -0.395, -0.186, and -0.092.

The coefficients of the region fixed effects show the occurrence of solar fields compared to the base province (Drenthe). In Flevoland, Groningen, and Zeeland, the chance is higher because the coefficient is positive. Zeeland has the highest coefficient (0.772). Zuid-Holland and Noord-Holland also have a positive coefficient, but these provinces are not statistically significant. This means that the occurrence of solar fields is not significantly higher in these provinces than in Drenthe. The other provinces have a lower chance than Drenthe, with Gelderland having the lowest (-0.479). Limburg has a negative coefficient as well, but a lower significance level of 0.05.

The land use fixed effects show that solar fields are more likely to be built on semi-built up sites than on industrial sites. Other urban, waterbodies, and recreational land use types also have a higher chance than industrial sites. Agricultural land use does not have a significantly higher chance than industrial land use. Nature has a negative coefficient, implying that the chance of solar fields being constructed on natural land use types is lower than on industrial sites.

Without the fixed effects, the pseudo-R-squared of this research is 0.0364. With the fixed effects included, the pseudo-R-squared is 0.0632. This means that the independent variables and the fixed effects collectively explain 6.32% of the variance in the occurrence of solar fields.

4.2 Probability map

Figure 4 shows the probability map that is made with the regression results. The map shows where solar fields are most likely to be developed in the future.

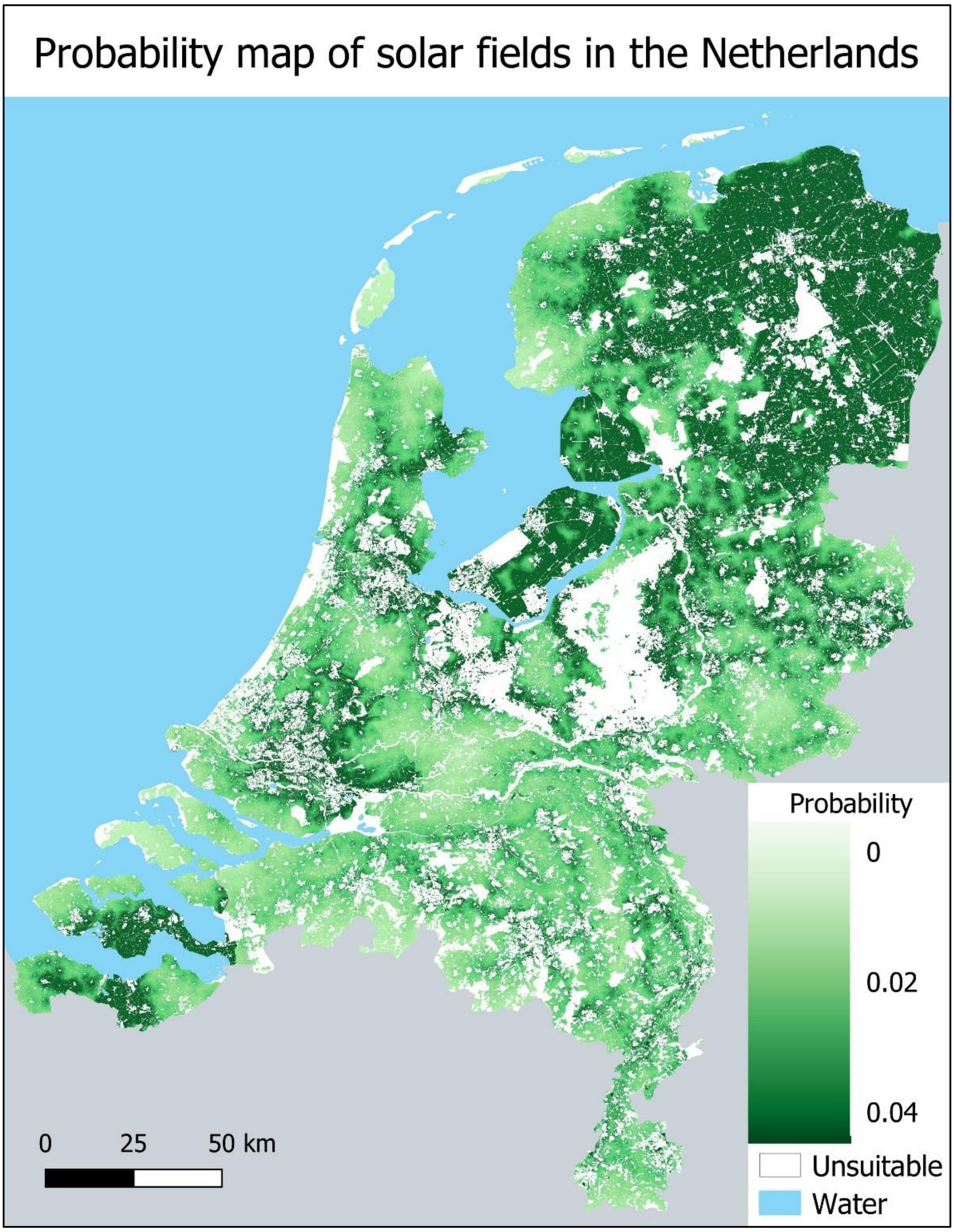


Figure 4. *Probability map showing the chance that sites will be used for solar fields. The map is based on the regression analysis. The blue colour indicates unsuitable water areas (chance = 0). The green colour indicates the probability. The white colour shows unsuitable places.*

The light green colour indicates a lower chance of solar fields than the dark green colour. Figure 4 shows that Groningen, Flevoland, Drenthe, and parts of Zeeland and Friesland have the highest chance of occurrence. The other provinces have a lighter colour and thus a lower chance. The white parts on the map are unsuitable or restricted areas. Examples of restricted areas are protected nature areas such as the Veluwe, the Wadden islands, and the Oostvaardersplassen. Unsuitable areas can be, for example, residential areas or rivers.

5. Discussion

In this chapter, the results will be interpreted in Section 5.1. Furthermore, the limitations and implications of the research are described in Section 5.2.

5.1 Interpretation

The standardised coefficients of the regression show that the distance to the electricity grid has the highest influence on the occurrence of solar fields, as this variable has the highest absolute standardised coefficient. It implies that sites further away from the electricity grid are less suitable. The global horizontal irradiance has the second highest influence on the occurrence of solar fields, followed by the distance to the nearest urban area and the distance to the nearest road.

The coefficients of distance to roads and the electricity grid are in line with the findings in Chapter 2. However, the coefficients of the GHI and the distance to urban areas are different than expected based on Chapter 2. A higher GHI means that a place can generate more electricity and is thus more suitable. However, the negative coefficient shows that solar fields are less likely to be developed in places with a higher GHI. An explanation is the population density in the Netherlands. The GHI is highest in the western part of the Netherlands (The World Bank, 2020). However, there is less space to build solar fields in the western part because there the population density is the highest (CLO, 2020). Thus, despite the high GHI, fewer solar fields are built in this part of the country. The northern part of the Netherlands has a lower population density but also a lower GHI, resulting in a negative coefficient.

The second striking result is the coefficient of the distance to the nearest urban area. As mentioned in Chapter 2, sites further away from urban areas are more suitable because there is less visual pollution. However, the negative coefficient implies the opposite. Research from Elboshy et al. (2021) shows that sites in urban areas are indeed not suitable for solar fields. However, being closer to urban areas makes the transportation costs lower. This makes sites near urban areas more suitable. This can be an explanation for why the coefficient for this factor is negative (-0.220).

The probability map shows that the parts of the Netherlands with agricultural land are light green, which is in line with the land use fixed effects. The darker green parts on the map are caused by the other variables.

The probability map shows that the chance of solar fields being developed is highest in Groningen, Drenthe, Flevoland, and Zeeland. The region fixed effects are visible in the map: the coefficients of these provinces are positive, making the probability higher. Gelderland, Limburg, Noord-Brabant, and Utrecht have a lower probability due to the region fixed effects. Although Zeeland has the highest region fixed effect, it is not entirely dark green. An explanation is that the distance to the electricity grid is larger in the northern part of the province, making the probability there lower.

The distance to the electricity grid has the highest absolute standardised coefficient of the four variables. This causes Friesland to be dark green at places where the distance to the electricity grid is small, despite the negative region fixed effect in this province. The GHI also affects the probability in the north of the Netherlands because it has the second highest standardised coefficient. The GHI is lowest in the north, making the probability higher and the map in the north greener.

The probability is higher near urban areas, caused by the negative coefficient of the distance to urban areas. Thus, near urban areas, the map is darker green. However, cities are unsuitable, creating the white spots on the map. The other white parts on the map show unsuitable places, such as protected nature areas and rivers. This is visible in the Veluwe, the Wadden Islands, the Oostvaardersplassen, and other protected areas.

5.2 Limitations and implications

The coefficient of ‘other urban’ could be high and positive because solar fields are classified as ‘public amenities’, a subcategory of ‘other urban’. The coefficient might be high because the subcategory 'public amenities' is not excluded from the dataset. This would change the outcome of the coefficients of the independent variables. Although the number of solar fields before 2017 is low, it is recommended to determine on what types of land these solar fields are placed to determine the land use fixed effects more accurately.

A limitation of this research is the availability of sources for the literature review. The existing literature is often focused on other countries. This can cause biases in the analysis, as other countries can have different weather patterns, policies, landscapes, et cetera. To keep any potential bias as small as possible, the sources that are used in this research are as recent as possible and preferably peer reviewed. Sources sometimes contain different information about the same topics. For example, when identifying the optimal distance from solar fields to urban areas, Carrión et al. (2008) and Perpiña Castillo et al. (2016) have identified different distances. In this case, the average distance is used to decrease the chance of biases.

Another implication of this research is the low pseudo-R-squared (0.0632) obtained from the regression analysis. The low value suggests that the independent variables explain only a small proportion of the variance in the occurrence of solar fields. This means that other factors may play a significant role in influencing the occurrence of solar fields. Identifying these factors can increase the pseudo-R-squared and thereby the accuracy of this research. In this research, the region fixed effects and land use fixed effects were added to increase the pseudo-R-squared. The fixed effects were added to the regression to account for differences in provinces (e.g., policies) and differences in land use suitability. Although the fixed effects increased the pseudo-R-squared, it is still low.

6. Conclusion and recommendation

In this chapter, the main research question and the sub-questions will be answered in Section 6.1. Furthermore, recommendations about the topic will be given in Section 6.2.

6.1 Conclusion

This research is conducted to analyse the location choice of solar fields in the Netherlands. The aim was to answer the following research question: *Which spatial factors determine the location of solar fields in the Netherlands?* To answer the research question, four sub-questions were formulated.

To answer the first three sub-questions, the environmental, technical, and socio-economic factors that determine the suitability of sites are determined through a literature review. The environmental factors that were determined are the global horizontal irradiance and the unsuitable land use categories. The unsuitable land use categories are excluded from the probability map. The technical factors that were identified are the distance to roads and to the electricity grid. The socio-economic factor used in this research is the distance to urban areas. These factors formed the independent variables in the logistic regression analysis.

To answer the fourth sub question, a logistic regression is conducted to determine the relative importance the factors have on the suitability of solar fields. The distance to the electricity grid has the highest influence on the occurrence of solar fields, followed by the global horizontal irradiance. The distance to the nearest urban area and the distance to the nearest road have, respectively, a lower influence. The coefficients of these factors are all negative and statistically significant.

The sub-questions collectively answer the main research question. The location of solar fields in the Netherlands is primarily determined by the distance to the electricity grid. Due to the region and land use fixed effects, locations in Zeeland and semi-built up locations have the highest chance of being used for solar fields.

6.2 Recommendations

To conduct more reliable research, it is recommended to determine more factors that might influence the location choice of solar fields. An example is the effect of solar fields on house prices. Research from Dröes & (2021) shows that solar fields decrease house prices nearby, which might influence the location choice. Another factor that might influence the location choice is the acceptance of solar fields by the local community (Van den Berg & Tempels, 2022). The determination of extra factors can either be done with field research or with a more detailed literature review. It can also be beneficial to use more detailed resolutions in the data. It is, for example, an option to use more pixels on the raster layers or to use more observations in the dataset. The fixed effects showed that there are differences per province and land use type. For further research, it is recommended to explain what causes these fixed effects. This can result in a more reliable regression analysis.

It is important to know the factors and the relative importance they have on the suitability of sites. The climate is changing, and the need for renewable energy is growing. The use of solar energy will have to increase in the future to pursue the Paris Climate Agreement. The Dutch government has plans to develop more solar fields, which will change the landscape of the Netherlands. With this research, policymakers can make predictions about the land use changes these solar fields will bring with them. That is important in the densely populated country that the Netherlands is. Solar fields do not only mitigate climate change; they also have an impact on the landscape. Solar fields are more likely to be developed near cities, so they will, in any case, impact the Dutch society.

7. References

AHN. (2023, march 27th). *AHN Viewer*. Actueel Hoogtebestand Nederland. Accessed April 25th, 2023, from https://www.ahn.nl/ahn-viewer

Baltas, A.E. and Dervos, A.N. (2012). Special framework for the spatial planning & the sustainable development of renewable energy sources. *Renewable Energy, 48*(2012), 358-363. http://dx.doi.org/10.1016/j.renene.2012.05.015

Carrión, J.S., Espín Estrella, A., Aznar Dols, F., Zamorano Toro, M., Rodríguez, M., Ramos Ridao, A. (2008). Environmental decision-support systems for evaluating the carrying capacity of land areas: optimal site selection for grid-connected photovoltaic power plants. *Renewable and Sustainable Energy Reviews*, *12*(9), 2358–2380. https://doi.org/10.1016/j.rser.2007.06.011

CLO (Compendium voor de Leefomgeving). (2020, October). *Bevolkingsgroei, 2015-2020.* Rijksoverheid. https://www.clo.nl/indicatoren/nl2102-bevolkingsgroei-nederland-

De Jong, J., Stremke, S. (2020). Evolution of Energy Landscapes: A Regional Case Study in the Western Netherlands. *Sustainability, 12*(11), p.2-9 https://doi.org/10.3390/su12114554

Dröes, M., & Koster, H. R. (2021). Wind turbines, solar farms, and house prices. *Energy Policy, 155*, 112327. https://doi.org/10.1016/j.enpol.2021.112327

Elboshy, B., Alwetaishi, M., Aly, R., Zalhaf, A. (2021). A suitability mapping for the PV solar farms in Egypt based on GIS-AHP to optimize multi-criteria feasibility. *Ain Shams Engineering Journal. 13*(3), 101618. https://doi.org/10.1016/j.asej.2021.10.013

IPCC (Intergovernmental Panel on Climate Change). (2023). *Synthesis report of the IPCC sixth assessment report (AR6).* https://report.ipcc.ch/ar6syr/pdf/IPCC_AR6_SYR_LongerReport.pdf

Kadaster. (2023). BRT_20230323_Zonneparken [Dataset]. In *Basisregistratie Topografie*.

Mierzwiak, M., & Calka, B. (2017). Multi-Criteria Analysis for Solar Farm Location Suitability. Reports on Geodesy and Geoinformatics, *104*(1), 20–32. https://doi.org/10.1515/rgg-2017-0012

Netbeheer Nederland. (2023). *Het energiesysteem van de toekomst: de II3050-scenario's (integrale energiesysteemverkenning 2030-2050).* https://www.netbeheernederland.nl/_upload/Files/Rapport_II3050_Scenario's_274.pdf

Perpiña Castillo, C., Batista E Silva, F., & Lavalle, C. (2016). An assessment of the regional potential for solar power generation in EU-28. *Energy Policy*, *88*(2016), 86–99. https://doi.org/10.1016/j.enpol.2015.10.004

Rijksoverheid. (2021, December 15th). *Omzien naar elkaar, vooruitkijken naar de toekomst (coalitieakkoord).* https://open.overheid.nl/documenten/ronl-f3cb0d9c-878b-4608-9f6a-8a2f6e24a410/pdf

Rolf. (2021, October). Zonneweides in Nederland. Zelf Energie Produceren. Accessed May 16th, 2023, from https://www.zelfenergieproduceren.nl/nieuws/zonneweides-in-nederland/

RVO (Rijksdienst voor Ondernemend Nederland) (2023). *Nieuwe ruimte voor de zon* (RVO-086 2023/RP-DUZA). Ministerie van Economische Zaken en Klimaat. https://ce.nl/wp-content/uploads/2023/04/CE-Delft_220389_Inspiratiegids_nieuwe_ruimte_voor_de_zon_Def.pdf

Spruijt, J. (2015). *Wat levert een zonneweide per ha op?* ARCRES – Wageningen UR. https://edepot.wur.nl/336567

The World Bank (2020). *Global Solar Atlas 2.0.* Accessed May 23th, 2023, from https://globalsolaratlas.info/download/netherlands

Van Dam, F., Van den Broek, L., Breedijk, M. (2019). *Zorg voor landschap. Naar een landschapsinclusief omgevingsbeleid.* Planbureau voor de Leefomgeving. https://www.pbl.nl/sites/default/files/downloads/pbl-2019-zorg-voor-landschap-3346_0.pdf

Van den Berg, K., Tempels, B. (2022). The role of community benefits in community acceptance of multifunctional solar farms in the Netherlands. *Land Use Policy, 122*, 106344. https://doi.org/10.1016/j.landusepol.2022.106344

Zonne-energie. (z.d.). Wise. Accessed June 29, 2023, van https://wisenederland.nl/artikel/zonne-energie/?cn-reloaded=1

Zon op Kaart (2023, February). *De ontwikkeling van zonneparken in kaart.* Accessed April 18, 2023, from http://www.zonopkaart.nl/

8. Appendices

Appendix A: Data for GIS

A.1: GIS data for existing solar fields analysis

Table A.1. Sources for the data used in GIS to analyse on what land use categories the existing solar fields are placed on.

Name	Year	Source	Data type	Criteria applied on this dataset
OpenStreetMap	2023	https://www.openstreetmap.org/#map=7/52.154/5.295	Raster	To create a map
BRT_Zonneparken	23-3-2023	Kadaster. (2023). BRT_20230323_Zonneparken [Dataset]. In *Basisregistratie Topografie*.	Vector	To show where the current solar fields are located
National_Border_NL	29-12-2021	https://service.pdok.nl/kadaster/bestuurlijkegrenzen/atom/v1_0/bestuurlijke_grenzen.xml	Vector	To create a polygon of the Netherlands to demarcate the study area

A.2: Land use categories

Table A.2. Land use categories from the CBS dataset used for the regression analysis.

Category	Subcategory	Name	Area (km^2)
Other urban	22	Public amenities	120.62
Industrial	24	Industrial Area	882.08
Semi-Built up	30	Dump	19.35
Semi-Built up	31	Wreckage repository	4.29
Semi-Built up	33	Mineral extraction site	31.51
Semi-Built up	35	Semi paved other terrain	41.31
Recreational	40	Park	329.91
Recreational	41	Sport terrain	363.04
Recreational	42	Allotment garden	36.09
Recreational	43, 44	Recreational	115.44; 240.68
Agricultural and other agricultural	51	Agricultural	22,217.95
Nature	60	Forest	3,414.49
Nature	61	Dry natural area	938.85
Nature	62	Wet natural area	671.13
Water	73	Bordering lakes	155.16
Water	74	Reservoir	12.40
Water	75	Recreational water	111.16
Water	76	Water for mineral extraction	29.76
		Total	**28,975.62**

Table A.2.1. Deleted land use categories from the CBS dataset

Category	Subcategory	Name	Area (km²)
Transport terrain	10, 11, 12	Railway, road, airport	1.155,81
Other urban	20, 21, 23	Residential, cultural, hospitality	2.699,65
Semi-built up	32, 34	Graveyard, building lot	296.955
Agricultural	50	Greenhouse horticulture	157.71
Water	70, 71, 72, 77, 78, 80, 81, 82, 83	Ijsselmeer & Markermeer, enclosed estuary, Rijn & Maas, flow & sludge, other inland water, Waddenzee, Eems & Dollard, Oosterschelde, Westerschelde, Noordzee*	4,087.24
Abroad	90	Abroad*	-
		Total	**8,397.37**
		*Noordzee and abroad land not used in area calculation	

A.3: GIS data for regression analysis

Table A.3. Sources for the data used in GIS. The layers are used for making the CSV table used for the regression.

Name	Year	Source	Data type	Criteria applied on this dataset
bbg2017 bestand_bodemgebruik_2017_bodemgebruik	2017	https://service.pdok.nl/cbs/bestandbodemgebruik/2017/atom/bestand_bodemgebruik_2017.xml	Vector	- Suitable land use categories - Distance to urban areas.
BRT_Zonneparken	23-3-2023	Kadaster. (2023). BRT_20230323_Zonneparken [Dataset]. In *Basisregistratie Topografie*.	Vector	To show where the current solar fields are located
Global Horizontal irradiance	3-4-2020	© 2020 The World Bank, Source: Global Solar Atlas 2.0, Solar resource data: Solargis https://solargis.com/maps-and-gis-data/download/netherlands	Raster	Irradiance in NL
Monuments – Protected Site	21-4-2023	https://service.pdok.nl/rce/ps-ch/atom/v1_0/ps-ch.xml	Vector	Protected sites (unsuitable)
National parks – Protected Site	23-8-2022	https://service.pdok.nl/rvo/beschermdegebieden/natura2000/atom/natura2000.xml (natura 2000 gebieden)	Vector	Protected sites (unsuitable)
Natura2000 – Protected Site	21-6-2022	https://service.pdok.nl/rvo/nationaleparken/atom/nationaleparken.xml (nationale parken)		Protected sites (unsuitable)
National_Border_NL	29-12-2021	https://service.pdok.nl/kadaster/bestuurlijkegrenzen/atom/v1_0/bestuurlijke_grenzen.xml	Vector	To create a polygon of the Netherlands to demarcate the study area
Province boundaries	2021	https://service.pdok.nl/kadaster/bestuurlijkegrenzen/atom/v1_0/index.xml	Vector	To account for the region fixed effect in the regression
Roads	2017	https://service.pdok.nl/cbs/bestandbodemgebruik/2017/atom/bestand_bodemgebruik_2017.xml	Vector	Distance to Main roads
Swimmingwater **Waterways**	18-2-2022 17-4-2023	https://service.pdok.nl/provincies/zwemwater-provinciaal-rijkswateren/atom/v1_0/am-lzr.xml https://service.pdok.nl/rws/nwbvaarwegen/atom/nwb_vaarwegen.xml	Vector	To exclude recreational water and waterways from suitable areas.
Tennet hoogspanningskabels	1-5-2019	https://www.pdok.nl/introductie/-/article/liander-elektriciteitsnetten-1	Vector	Distance to the electricity grid

Appendix B: Flowcharts

B.1: Flowcharts for data preparation using GIS

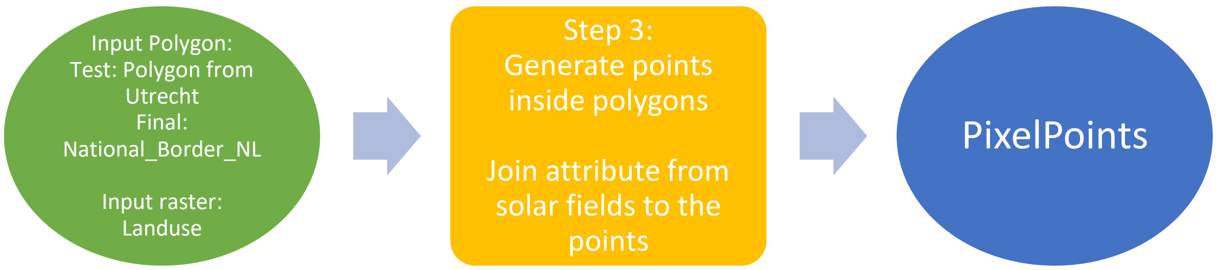
This appendix contains the flowcharts to show the tools that have been used in GIS to prepare the data for the regression analysis. To explain what colour represents what process or layer, an example is shown below.

Example:

Step 1: rasterize the vector layers. The green patch shows the input layers that are collected directly from the internet. The yellow square shows the tool that has been used for this step. The blue patch shows the processed layers that can be used for the next step.

Step 2: the proximity factors need to be converted to layers that contain the distance from the factor to the solar fields. For example, the distance from the solar field to the electricity grid. The urban raster layer is created from the land use layer.

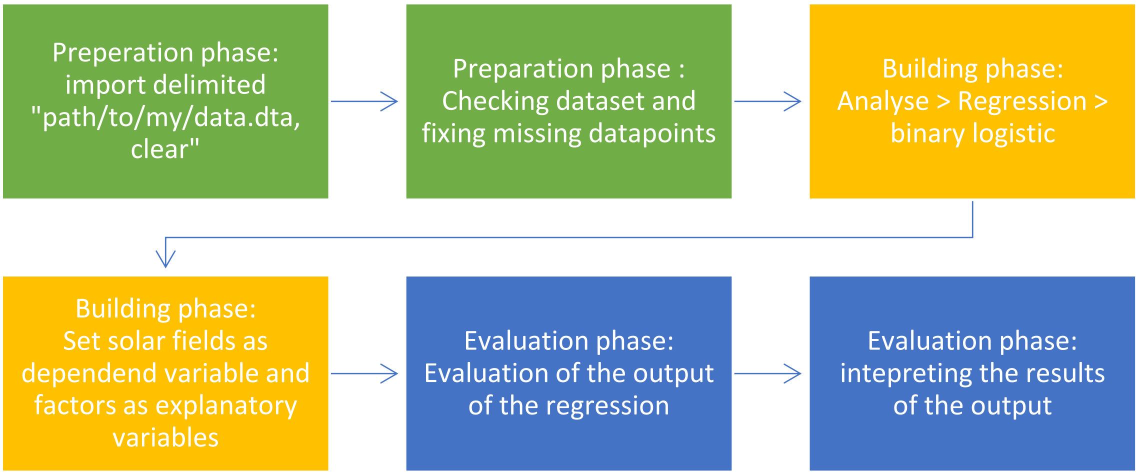
Step 3: this step uses 'generate points inside polygons'. This will be done with a smaller test polygon first (the province of Utrecht) and later with the border of the Netherlands.



Step 4: the point sampling' tool is used to make a csv table for the regression. The output is a table that can be used for the logistic regression analysis.

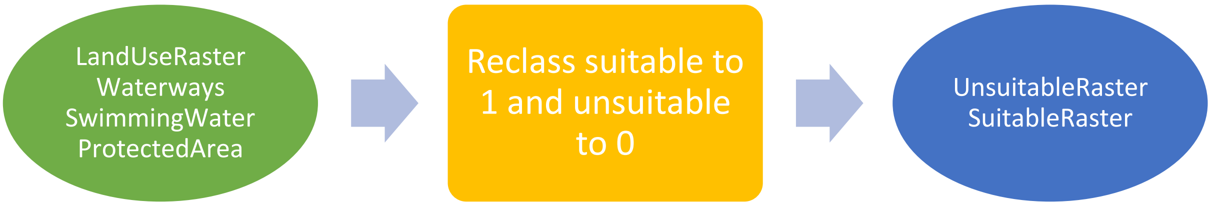
B.2: Flowchart for logistic regression analysis

Table B.2. Flowchart for the regression analysis. There will also be a log file from the regression to see the exact commands that are used in Stata.

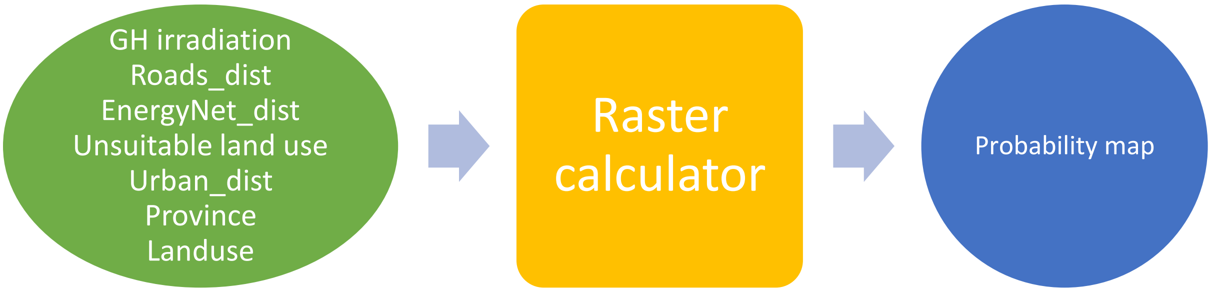


B.3: Flowchart for the probability map using GIS

For the probability map, some land use categories mentioned in chapters 2 and 3 need to be reclassified to 0 if unsuitable and 1 if suitable. This is done with the reclass tool.



Then, the layers can be used for the raster calculator.



The formula used in the raster calculator is as follows:

$$P = e^{a+\beta*X}/(1+e^{a+\beta*X})$$

In which:

P is the probability of occurrence of solar fields.

e is the basis for the natural logarithm (2.718).

a is a constant retrieved from the regression analysis

ß is a vector of estimation parameters for all variables X.

X is a set of location factors (explanatory variables).

Appendix C: Proximity maps

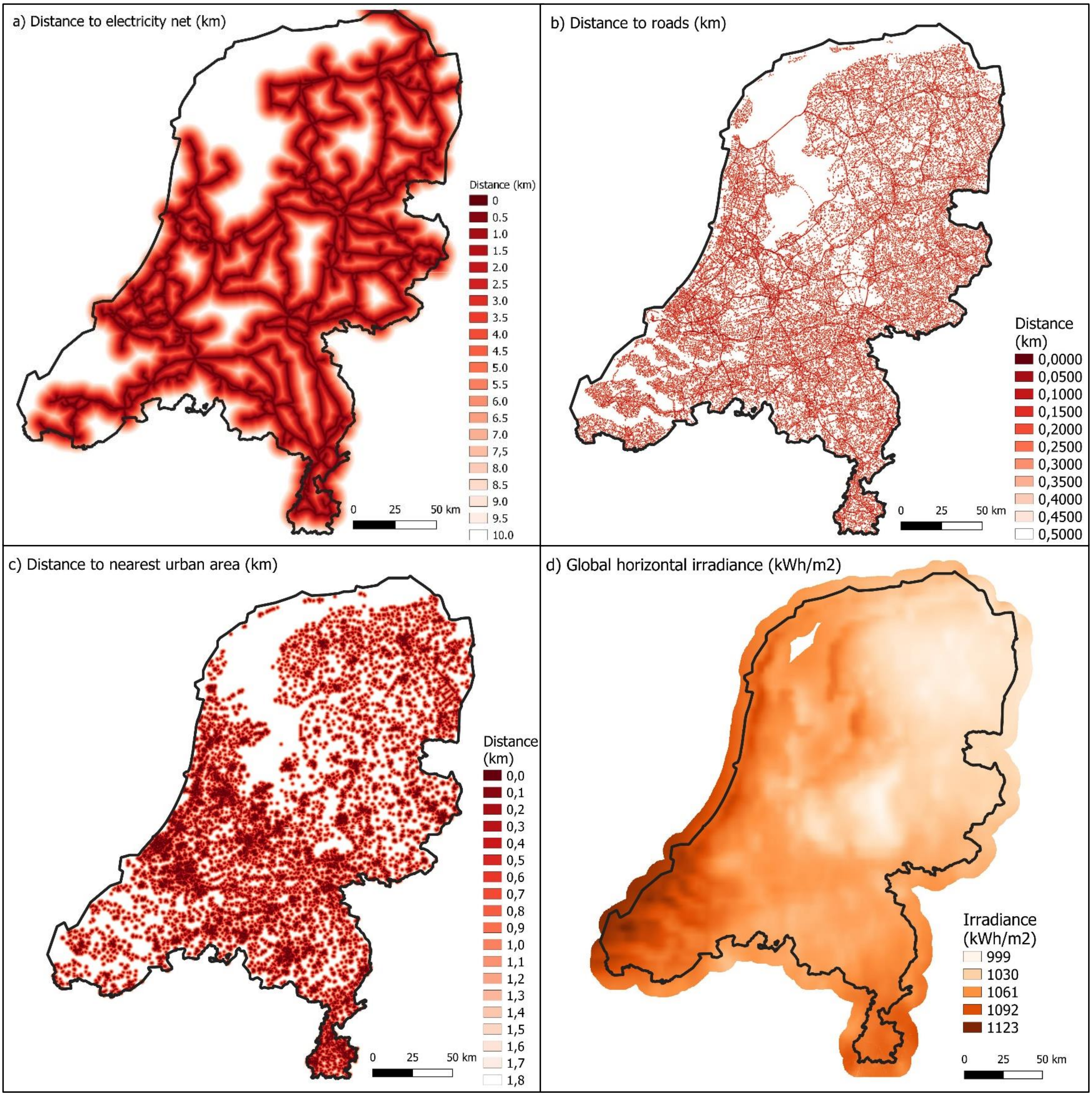


Figure C.1. *Proximity maps made in QGIS to be used for the logistic regression analysis. For maps a, b, and c, a darker colour implies being closer to the factor. In map d, a darker colour means a higher GHI. Based on data shown in appendix A.3.*