

Modelling the Spatial Distribution of Installed Solar Photovoltaic Capacity



Hussah Alghanem

Supervisor: Alastair Buckley

A thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

in the

Department of Physics and Astronomy

October 8, 2025

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Prof. Alastair Buckley, for his exceptional mentorship throughout this journey. His clarity of thought, generosity with his time, and steady encouragement made an enormous difference at every stage of this research. He taught me to see the world through a quantitative lens and patiently guided me through the craft of academic writing. I am especially thankful for the intellectual freedom he offered, the confidence he placed in me from the very beginning, and the many opportunities he created for me to connect with others in the field and engage with the wider research community. I feel truly privileged to have worked with him; it has been nothing but a pleasure.

I am grateful to the Sheffield Solar group for their support and collaborative spirit during the course of my research, especially Julian Briggs, Andrew Richards, and Jamie Taylor. I also thank Prof. David Lidzey and the Electronic and Photonic Molecular Materials (EPM) research group, as well as Prof. Solomon Brown and the Brown Group, for providing a stimulating and supportive academic environment.

I thank Dr. Aljawharah Alnaser for her generous support and insightful advice on pursuing both a PhD and an academic career. Her encouragement made a lasting impact, and I'm grateful for the joy her daughter Lolia brought into my life.

I am deeply thankful to my parents and siblings for their unwavering support. My mother, Aljawharah, was both my first teacher and my first science experiment supervisor—unpaid, and remarkably tolerant. My father, Maher, showed me that even the hardest things can be met with humour, and that life is often too strange not to laugh at. My siblings—Nora, Abdulaziz, Mohammed, and Hammad—have been a source of steady support and the kind

of honest feedback that makes peer review feel familiar. I thank Nora especially for listening to me go on about my work and for reading it, too—even when neither made much sense.

I couldn't have done this without the encouragement, humour, and well-timed distractions from my friends, whose presence made this experience lighter and more joyful.

Finally, I gratefully acknowledge Imam Abdulrahman Bin Faisal University for funding my doctoral studies.

Abstract

The global transition toward decarbonized energy systems has intensified the need for accurate, spatially resolved data on installed solar photovoltaic (PV) capacity. However, inconsistencies in reporting, limited geographic granularity, and varying measurement standards pose challenges for planning and assessment. This thesis addresses these challenges by developing spatial models to estimate, benchmark, and forecast installed PV capacity across global, regional, and subregional scales.

Structured as a thesis by publication, the work comprises three core studies. The first develops a global model of installed PV capacity at the national level, identifying key geographic and socioeconomic drivers. The second estimates regional capacity across 36 European countries, including those lacking official regional data. The third focuses on Great Britain, modelling subregional capacity. Collectively, the models disaggregate national capacity, benchmark deployment, and forecast where future capacity is likely to be installed—supporting efforts to monitor generation, reduce connection delays, plan grid expansion, and address land-use conflicts by identifying areas where solar development may compete with other uses, including agriculture.

The models exhibit strong performance across spatial scales. The Global Model estimates annual capacity additions with a global error of 9.7%. The European Model estimates cumulative capacity at the NUTS 2 level and achieves a national error of 19.5% when applied across all countries. In countries with available regional data—including the UK, Italy, Spain, Belgium, Germany, and France—the error falls to 2.5%. The GB Model achieves a national MAPE of 5.4% at the NUTS 3 level.

Across spatial scales, a shift in deployment drivers emerges. National capacity is shaped

by socioeconomic factors, while regional and subregional deployment is driven by land-use characteristics, with artificial surfaces and agricultural areas as strong predictors. While solar irradiation is often assumed critical, the models show that structural and socioeconomic conditions are more influential, particularly in developed markets.

Publications Included in This Thesis

- Alghanem, H., & Buckley, A. (2024). *Global Benchmarking and Modelling of Installed Solar Photovoltaic Capacity by Country*. **Energies**, 17(8), 1812. <https://doi.org/10.3390/en17081812>
- Alghanem, H., & Buckley, A. (2025). *Modelling Regional Solar Photovoltaic Capacity in Europe: A Data-Driven Approach for Disaggregation, Benchmarking, and Forecasting*. **Energy Reports**, 14, 1283–1302. <https://doi.org/10.1016/j.egyr.2025.07.010>
- Alghanem, H., & Buckley, A. (2025). *Modelling Regional Solar Photovoltaic Capacity in Great Britain*. **arXiv** preprint arXiv:2502.19243. <https://doi.org/10.48550/arXiv.2502.19243>

Contents

Glossary	xxii
List of Abbreviations	xxiv
1 Introduction	1
1.1 The Evolution of Solar Photovoltaics: From Scientific Discovery to Global Deployment	1
1.2 Significance and Challenges of Tracking Solar PV Capacity	8
1.2.1 Tracking National PV Capacity	8
1.2.2 Tracking Regional and Subregional PV Capacity	11
1.2.3 Significance of Tracking Solar PV Capacity	15
1.2.4 Challenges of Tracking Solar PV Capacity	20
1.3 Capacity Definitions and Scope of Use	23
1.4 Research Focus and Overview	23
2 Literature Review	27
2.1 Factors Associated With Solar PV Deployment	27
2.1.1 Household and Subregional Factors	27
2.1.2 Regional Factors	29
2.1.3 National Factors	31
2.2 Methods for Modelling Solar PV Capacity	32
2.2.1 Statistical Models	32

2.2.2	Vision-Based Models	35
2.3	Solar PV Policies	43
2.3.1	Feed-in Policies	43
2.3.2	Net Metering	48
2.3.3	Renewable Portfolio Standards and Certificate-Based Mechanisms . .	49
2.3.4	Other Policies	51
2.3.5	Future Policies	53
2.4	Knowledge Gap	54
3	Global Model	56
3.1	Abstract	56
3.2	Introduction	57
3.3	Literature Review	59
3.4	Methodology	63
3.4.1	Determining Key Features Associated with Global Solar PV Capacity Additions	63
3.4.2	Feature Screening and Selection	73
3.4.3	Model Building	75
3.4.4	Model Application	77
3.5	Results and Discussion	78
3.6	Conclusions	88
3.7	Appendix	89
3.7.1	Data Availability for Correlation Analysis	89
3.7.2	Summary of PCA Cluster Characteristics	90
3.7.3	Descriptive Statistics of Model Features	90
3.7.4	Pearson Correlation Matrix of Model Input Features	91
3.7.5	Feature Selection: Stepwise Regression	91
3.7.6	Feature Importance, Correlation, and Variance Explained by Combined Model Features	94

3.7.7	MLR Model	95
3.7.8	Polynomial Model	96
4	European Model	97
4.1	Abstract	97
4.2	Introduction	98
4.3	Methodology	100
4.3.1	Data and Data Processing	101
4.3.2	Feature Selection	102
4.3.3	Model Training	104
4.3.4	Model Applications	107
4.4	Results and Discussion	107
4.4.1	Feature Analysis and Selection	107
4.4.2	Model Performance Evaluation	111
4.4.3	Model Interpretation and Applications	117
4.4.4	Limitations	123
4.5	Conclusion	124
4.6	Appendix	126
4.6.1	Full Feature Analysis Tables	126
4.6.2	Input Features and Predicted Outputs of the Absolute Model	135
4.6.3	Input Features and Predicted Outputs of the Normalized Model	136
4.6.4	Feature Importance Based on SHAP Values	137
4.6.5	Regional PV Capacity Predictions for 2023	140
4.6.6	Forecast Validation	158
5	Great Britain Model	161
5.1	Abstract	161
5.2	Introduction	162
5.2.1	Literature Review	164

5.3	Methodology	167
5.3.1	Data and Data Processing	167
5.3.2	Feature Selection	168
5.3.3	Model Training	169
5.3.4	Model Applications	170
5.4	Results and Discussion	171
5.4.1	Model Applications	175
5.5	Conclusion	179
5.6	Appendix	181
5.6.1	Normalized data analysis	181
6	Concluding Discussion	188
6.1	Future Research	194

List of Figures

1.1	Historical deployment of solar photovoltaic capacity for the top 10 countries by installed capacity: China, USA, Japan, Germany, India, Brazil, Australia, Italy, Spain, and the Netherlands. Global cumulative capacity is shown as a black dashed line. Data sourced from IRENA [17]. In the early 2000s, Japan, the United States, and Germany led global solar PV deployment, with installed capacities significantly higher than China's. Japan's capacity in 2000 was nearly ten times greater than China's, while the United States and Germany had fivefold and threefold higher capacities, respectively. However, by 2015, this pattern had reversed: China overtook all three countries to become the world's leading market for cumulative installed solar PV capacity.	3
1.2	Historical evolution of solar photovoltaic module prices (USD/W) against cumulative global installed PV capacity from 1975 to 2023, demonstrating the consistent decrease in module prices with increased deployment. Data sourced from IRENA [19].	4

3.1	Relationship between number of features and model performance metrics. (a) Number of features versus coefficient of determination, and (b) number of features versus root mean squared error after fitting a linear model. Features are numbered as follows: (1) electricity net consumption, (2) tertiary education, (3) agricultural land area, (4) previous year's PV cumulative capacity, (5) population, (6) electricity net generation, (7) total unemployment, (8) land area, (9) GDP, (10) primary education, (11) value added by industry (including construction), (12) labour force, (13) rural land area, (14) value added by manufacturing.	79
3.2	Linear and logarithmic actual vs. predicted solar photovoltaic capacity additions for the combined model. Blue points represent the training set and orange points represent the test set.	81
3.3	Average cumulative solar photovoltaic capacity and error per year for the combined model. The yearly error was calculated from the entire dataset by dividing the mean absolute error per year by the average cumulative capacity that year.	81
3.4	Country error of the combined model, determined by dividing the mean absolute error per country by the respective mean cumulative photovoltaic capacity. Countries without available data are represented in white.	82
3.5	Mean cumulative photovoltaic capacity across various countries from 2001 to 2018. Countries without available data are represented in white.	83
3.6	Solar PV deployment index (SPVDI) for different countries during the period from 2010 to 2018. Countries are ranked based on their solar photovoltaic deployment compared to other countries with similar social, economic, and land-use factors. A positive value indicates that a country has more capacity than expected, while a negative value means less capacity than expected from similar countries.	84

3.7	Actual versus estimated solar photovoltaic capacity additions for Italy, the United Kingdom, Mexico, and Spain for years between 2001 and 2018. Blue points represent actual capacity, while orange points represent estimated capacity.	86
4.1	Actual regional photovoltaic capacity for 150 NUTS 2 regions in 2023. The capacity data is sourced from the United Kingdom, Italy, Spain, Belgium, Germany, and France.	103
4.2	Actual versus predicted regional photovoltaic capacity for 150 NUTS 2 regions using the absolute and normalized XGBoost models under the mixed hold-out configuration (2010–2023). Blue points represent training data and orange points represent test data. The actual capacity data are sourced from the United Kingdom, Italy, Spain, Belgium, Germany, and France.	116
4.3	Predicted regional solar photovoltaic capacity for 36 European countries and 333 NUTS 2 regions in 2023 using the scaled XGBoost models trained under the mixed hold-out configuration. The corresponding regional capacity values are provided in Appendix 4.6.5. (a) The scaled absolute model is more effective for allocating unknown national capacity to geographic locations in the training countries, which include the United Kingdom, Italy, Spain, Belgium, Germany, and France. (b) The scaled normalized model is better suited for estimating PV capacity in countries where regional data is unavailable. Although the model predicts capacity as a percentage of the national total (i.e., normalized output), these percentages are subsequently multiplied by national PV capacity to produce final capacity estimates in megawatts.	119

4.4	The solar PV deployment index results for the year 2023 derived from the XGBoost models trained under the mixed hold-out configuration. Positive values indicate regions where the actual capacity exceeds the expected capacity, while negative values indicate regions where the actual capacity is less than expected. (a) The absolute model benchmarks deployment compared to similar regions in Europe. (b) The normalized model benchmarks deployment compared to similar regions within a country. The regions included are Brandenburg (DE40), Castilla-La Mancha (ES42), Oberbayern (DE21), Niederbayern (DE22), Mecklenburg-Vorpommern (DE80), Stuttgart (DE11), Schleswig-Holstein (DEF0), Weser-Ems (DE94), Oberpfalz (DE23), Extremadura (ES43), Murcia (ES62), East Flanders (BE23), and Limburg (BE22).	120
4.5	Forecasted regional solar PV capacity in Germany for 2030, based on the national target of 215 GW [261]. The distribution across regions is estimated using the scaled normalized XGBoost model trained under the mixed hold-out configuration, which allocates national capacity proportionally according to regional characteristics. The forecast is based on 2018 CORINE land-cover data, with climatic variable values set to those observed in 2023.	123
4.6	Flow chart of the absolute XGBoost model trained under the mixed hold-out configuration, showing all input features, model configuration, and predicted output.	135
4.7	Flow chart of the normalized XGBoost model trained under the mixed hold-out configuration, showing all input features, model configuration, and predicted output.	136

5.1	Actual versus predicted regional photovoltaic capacity using the XGBoost model for 168 NUTS 3 regions in Great Britain, covering the years 2010 to 2023. Blue points represent training data, while orange points represent test data. Panel (a) shows the unscaled model; panel (b) shows the scaled model, where predictions are adjusted to match national totals.	173
5.2	(a) Actual regional solar photovoltaic capacity across 168 NUTS 3 regions in Great Britain for the year 2023. (b) Predicted regional solar photovoltaic capacity across 168 NUTS 3 regions in Great Britain for the year 2023 using the scaled XGBoost model. (c) Solar PV Deployment Index (SPVDI) calculated using the XGBoost model predictions for the years 2010 to 2023. Positive values indicate regions where the actual capacity exceeds the expected capacity, while negative values indicate regions where the actual capacity is less than expected. The regions included are Barnsley, Doncaster and Rotherham (UKE31) which under deployed by 985 MW and Cambridgeshire CC (UKH12) which over deployed by 780 MW. <i>Note: Map lines delineate study areas and do not necessarily depict accepted national boundaries.</i>	176
5.3	Unallocated solar photovoltaic capacity for 168 NUTS 3 regions in Great Britain from the year 2010 to 2023.	176
5.4	SHAP analysis for the unscaled XGBoost model, showing the mean contribution of factors to solar PV capacity percentage between 2010 and 2023 for (a) Cambridgeshire CC (UKH12) and (b) Barnsley, Doncaster, and Rotherham (UKE31). In the SHAP framework, $E[f(x)]$ represents the model's average prediction, while $f(x)$ is the predicted value, calculated as the sum of the contributions of the features and the model's average prediction.	177

List of Tables

3.1	Features considered for modelling solar photovoltaic capacity additions. The definition, category, and availability of the data are shown. The correlation and coefficient of determination (R^2) with solar photovoltaic capacity additions are calculated. Principal component analysis (PCA) is performed, and similar features are clustered together. Finally, the literature that used the same or similar features is linked.	64
3.2	Comparison between the results of the different models. MLR is the Multiple Linear Regression model, PR is the Polynomial Regression model, NN is the Neural Network model, and CM is the Combined Model. The training and test sets used in each model were identical.	80
3.3	Number of data points used in the analysis presented in Table 3.1 to examine the correlation between each factor and solar PV capacity additions.	89
3.4	Summary of the characteristics of principal component analysis (PCA) clusters, including the number of variables within each cluster, the coefficient of determination (R^2) obtained when fitting cluster variables into a linear model with PV capacity additions, and the corresponding number of data points used in the fitting process.	90
3.5	Descriptive statistics of the features used in the models. This table shows the count, mean, standard deviation, and distribution quartiles of the selected features.	90

3.6	Pearson correlation coefficients among selected variables used as inputs for the models.	91
3.7	Fit statistics for stepwise regression models under different criteria and directions. SSE is the sum of squared errors, DFE is the error degrees of freedom, RMSE is the root mean square error, R^2 is the coefficient of determination, Adj. R^2 is the adjusted coefficient of determination, Cp is Mallows' Cp statistic, p is the number of parameters (including the intercept), AICc is the corrected Akaike's Information Criterion, and BIC is the Bayesian Information Criterion.	92
3.8	Predictors selected by stepwise regression.	93
3.9	Feature importance, correlation, and variation explained by the combined model features. The main effect measures the contribution of the feature alone, while the total effect measures the contribution of the feature alone and in combination with other features.	94
3.10	Parameter estimates for the MLR model.	95
3.11	Parameter estimates for the polynomial model.	96
4.1	Top 15 features with the highest correlation to regional solar photovoltaic capacity in absolute terms. Displayed are the coefficient of determination (R^2), Pearson correlation, Spearman correlation, and the average of the two correlations. Land cover features are preceded by their CORINE Land Cover (CLC) classification codes (e.g., 1.2.1 for industrial or commercial units).	109
4.2	Top 15 normalized features with the highest correlation to the percentage of regional solar photovoltaic capacity. Features are normalized relative to national values and expressed as percentages. Displayed are the coefficient of determination (R^2), Pearson correlation, Spearman correlation, and the average of the two correlations. Land cover features are preceded by their CORINE Land Cover (CLC) classification codes (e.g., 1.2.1 for industrial or commercial units).	110

4.3	Regional error metrics for the absolute XGBoost models trained with different objectives and scaling approaches under the mixed hold-out configuration. Models were trained using either a Gaussian (squared-error) or Tweedie objective, with results reported both before and after national scaling. Metrics include the coefficient of determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), reported separately for training and test sets in megawatts (MW).	111
4.4	Regional error metrics for the benchmark multiple linear regression (MLR) models. The Absolute Model metrics are reported in megawatts (MW), while the Normalized Model metrics are reported in both megawatts (MW) and percentage points (%). Metrics include the coefficient of determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The MLR models were trained using the full dataset without cross-validation.	112
4.5	Regional error metrics for the XGBoost models under the mixed hold-out configuration, reported separately for training and test sets in both megawatts (MW) and percentage points (%). For the scaled models, regional predictions were adjusted to ensure that their sum matched the corresponding national total. Metrics include the coefficient of determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).	113
4.6	Regional error metrics for the XGBoost absolute model (MW) under the hold-out country configuration. For each iteration, the model was trained on all countries except one, which was used for testing. Metrics are reported separately for training and test sets.	114

4.7	Regional error metrics for the XGBoost normalized model (MW) under the hold-out country configuration. For each iteration, the model was trained on all countries except one, which was used for testing. Metrics are reported separately for training and test sets.	114
4.8	Regional error metrics for the XGBoost models under the hold-out year configuration, reported separately for training and test sets in both megawatts (MW) and percentage points (%). The models were trained on data from 2010–2020 and tested on 2021–2023. Metrics include the coefficient of determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).	115
4.9	National-level error metrics for the absolute and normalized XGBoost models under the mixed hold-out configuration. Metrics compare the actual national capacity with the estimated national capacity, obtained by summing predicted regional values. Results are shown for both the training countries and the full European dataset. Training countries include the United Kingdom, Italy, Spain, Belgium, Germany, and France. Scaling refers to adjusting regional capacities so that their totals match the observed national capacity.	117
4.10	Absolute data analysis. Features considered for modelling regional solar photovoltaic capacity. The data availability, coefficient of determination (R^2), Pearson correlation, Spearman correlation, correlation average are shown. Relevant literature that explores similar features is cited. Land cover features are preceded by their CORINE Land Cover (CLC) classification codes (e.g., 1.2.1 for industrial or commercial units).	126

4.11	Normalized data analysis. Features considered for modelling the percentage of regional solar photovoltaic capacity. All features presented are normalized relative to national values and expressed as percentages. The data availability, coefficient of determination (R^2), Pearson correlation, Spearman correlation, correlation average are shown. Relevant literature that explores similar features is cited. Land cover features are preceded by their CORINE Land Cover (CLC) classification codes (e.g., 1.2.1 for industrial or commercial units). . . .	130
4.12	SHAP value contributions of features in the absolute XGBoost model trained under the mixed hold-out configuration. The mean SHAP value represents the average absolute SHAP value for each feature across all observations. The percentage contribution is calculated by dividing the mean absolute SHAP value of each feature by the sum of mean absolute SHAP values for all features, and then multiplying by 100 as shown in Equation (4.2). Percentage contributions are rounded to one decimal place and may not sum to exactly 100% due to rounding.	137
4.13	SHAP value contributions of features in the normalized XGBoost model trained under the mixed hold-out configuration. All features presented are normalized relative to national values and expressed as percentages. The mean SHAP value represents the average absolute SHAP value for each feature across all observations. The percentage contribution is calculated by dividing the mean absolute SHAP value of each feature by the sum of mean absolute SHAP values for all features, and then multiplying by 100 as shown in Equation (4.2). Percentage contributions are rounded to one decimal place and may not sum to exactly 100% due to rounding.	138

4.14	Actual and predicted regional solar PV capacity in 2023 using the XGBoost models trained under the mixed hold-out configuration. Results are shown for both the absolute and normalized models. Percentages represent each region's share of its national total, while MW values denote capacity estimates in megawatts. "Scaled" values are adjusted to ensure that regional totals match national capacity. Blank cells indicate regions with unavailable data, whereas 0.0 denotes regions with no installed capacity within countries for which regional data exist.	140
4.15	Regional error metrics for the naïve carry-forward baseline and the scaled normalized model. Models are evaluated over the full 2017–2023 horizon and for 2023 only. Reported metrics are region-level R^2 , MAE, MSE, and RMSE, presented for both megawatts (MW) and percentage shares (%).	159
4.16	Country-level regional error metrics for 2023 comparing the naïve carry-forward baseline and the scaled normalized model across Germany (DE), Spain (ES), the United Kingdom (UK), Belgium (BE), France (FR), and Italy (IT). Metrics are region-level R^2 , MAE, MSE, RMSE, and MAPE (%). Belgium (BE) has missing naïve values because regional PV capacity data for 2010–2016 were unavailable.	160
5.1	Regional error metrics for the XGBoost and MLR models, reported separately for training and test sets in both megawatts (MW) and percentages (%). For the Scaled XGBoost model, regional predictions are adjusted to ensure they sum to the national capacity. Metrics reported are the coefficient of determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).	173

5.2	National error metrics for the XGBoost unscaled and scaled models, evaluated over the entire dataset. These metrics compare the actual national capacity with the estimated national capacity, which is derived from summing the regional capacities. Scaling refers to adjusting regional capacities so that their sum matches the actual national capacity.	174
5.3	Normalized data analysis. Features considered for modelling the percentage of regional solar photovoltaic capacity. All features presented are normalized relative to national values and expressed as percentages. The data availability, coefficient of determination (R^2), Pearson correlation, Spearman correlation, correlation average are shown. Relevant literature that explores similar features is cited. Land cover features are preceded by their CORINE Land Cover (CLC) classification codes (e.g., 1.2.1 for industrial or commercial units). . . .	181

Glossary

Actual installed solar PV capacity The total direct current (DC) capacity of solar photovoltaic (PV) systems that have been physically installed and connected to the grid. This reflects the true system capacity, regardless of whether it is reported or accurately measured. DC capacity is the sum of module capacity (not inverter capacity).

Reported solar PV capacity The PV capacity figures officially reported by governments, regulators, or other institutions. Therefore, national capacity estimates are often compiled from multiple sources that may miss particular system types. These may be incomplete due to under-reporting (e.g., small or off-grid systems not registered) or inaccurate due to reporting errors, such as duplicated systems or outdated data.

Modelled solar PV capacity Estimates of installed solar PV capacity derived from predictive models. These aim to approximate actual capacity, especially in cases where reported or measured data are missing or unreliable. Modelled capacity provides a consistent basis for analysis and comparison.

Forecasting PV capacity The process of estimating future solar PV capacity, either in the short or long term. Forecasts may be based on trends, policy scenarios, or predictive models that incorporate geographic, economic, and technological factors.

Benchmarking PV capacity The comparison of reported vs modelled solar PV deployment levels across countries or regions. Benchmarking helps evaluate performance relative to geographic, socioeconomic, and environmental conditions to identify under- or over-performing areas.

Disaggregation of PV capacity The process of breaking down national-level solar PV capacity figures into estimates at finer geographic scales, such as regions or localities. This enables more detailed analysis and localized planning.

Geographic factors Spatially varying characteristics that influence solar PV deployment, including social (e.g., population density), economic (e.g., income levels), climatic (e.g., solar radiation), and land use (e.g., availability of artificial surfaces or agricultural land) conditions.

Centralized PV capacity Centralized PV capacity refers to large-scale solar power installations, such as utility-scale solar farms, that are typically connected to the high-voltage transmission or primary distribution grid. These systems are usually ground-mounted, centrally monitored, and fully visible to the system operator, allowing for direct integration into grid management and planning.

Distributed PV capacity Distributed PV capacity refers to smaller, often rooftop-mounted solar systems installed on residential, commercial, or public buildings and connected to the low- or medium-voltage distribution network. These systems are generally located close to the point of consumption and may not be fully visible to the system operator, especially in markets without comprehensive monitoring or registration systems, making their impact on grid planning and operation more difficult to assess.

List of Abbreviations

CART	Classification and Regression Trees
CfD	Contract for Difference
DNO	Distribution Network Operator
EU	European Union
F1 score	Harmonic Mean of Precision and Recall
FCN	Fully Convolutional Network
FiT	Feed-in Tariff
FPN	Feature Pyramid Network
GDHI	Gross Disposable Household Income
GDP	Gross Domestic Product
GDPR	General Data Protection Regulation
GERD	Gross Expenditure on Research and Development
GHI	Global Horizontal Irradiance
GLMs	General Linear Models
GVA	Gross Value Added
IoU	Intersection over Union
IRENA	International Renewable Energy Agency
ITL	International Territorial Level
LSOA	Lower Layer Super Output Area
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error

MCS	Microgeneration Certification Scheme
MSE	Mean Squared Error
MW	Megawatt
NUTS	Nomenclature of Territorial Units for Statistics
OBIA	Object-Based Image Analysis
Ofgem	Office of Gas and Electricity Markets
PCA	Principal Component Analysis
PV	Photovoltaic
REPD	Renewable Energy Planning Database
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
ROCS	Renewable Obligation Certificates
SHAP	Shapley Additive Explanations
SPVDI	Solar PV Deployment Index
SVM	Support Vector Machine
TSO	Transmission System Operator
U-Net	U-shaped Convolutional Neural Network
XGBoost	Extreme Gradient Boosting
R^2	Coefficient of Determination

Chapter 1

Introduction

1.1 The Evolution of Solar Photovoltaics: From Scientific Discovery to Global Deployment

The photovoltaic effect, the physical principle underlying solar photovoltaic (PV) technology, was first observed in 1839 by French physicist Alexandre-Edmond Becquerel. He discovered that certain materials produced an electric current when exposed to light, marking the beginning of a field that would later revolutionize energy systems [1, 2]. In the late 19th century, further progress was made by Willoughby Smith and William Grylls Adams, who demonstrated the photoelectric properties of selenium, confirming that light could generate electricity without the need for moving parts [3, 4].

In 1883, Charles Fritts constructed the first true solar cell by coating selenium with a thin layer of gold. Although its efficiency was below 1%, it represented the first practical application of the photovoltaic effect [5]. This was followed by the development of a copper-copper oxide solar cell in 1904 by Wilhelm Hallwachs [6].

A major theoretical breakthrough came in 1905, when Albert Einstein published his explanation of the photoelectric effect, demonstrating how light could eject electrons from certain materials [7]. This work laid the theoretical foundation for modern solar energy conversion and earned him the Nobel Prize in Physics in 1921 [8].

In the early 20th century, the search for alternative materials and methods of solar energy conversion continued, although widespread application remained limited due to high costs and low efficiencies. The decisive breakthrough for practical solar PV came in 1954, when Daryl Chapin, Calvin Fuller, and Gerald Pearson at Bell Labs developed the first silicon-based solar cell, achieving an efficiency of approximately 6% [9]. This milestone is widely regarded as the beginning of the modern photovoltaic era.

Solar PV's reliability and potential were soon demonstrated in space applications. In 1958, the Vanguard I satellite was launched with solar cells that powered its radio transmitter, proving solar PV's effectiveness in harsh and remote environments [10]. This success led to widespread adoption of PV technology in space exploration.

The oil crises of the 1970s revived global interest in alternative energy sources, prompting governments—particularly in the United States—to invest in solar research and development [11]. These investments facilitated gradual improvements in solar cell efficiency and reductions in manufacturing costs. By the 1980s, innovations in manufacturing processes helped raise its efficiency to approximately 10% [12, 13].

Between 1980 and 2010, technological innovation in solar PV accelerated significantly. Patent data from this period show an average annual growth rate of 10.5%, with a marked concentration in five countries: Japan, the United States, China, Germany, and South Korea. Together, these nations accounted for over 92% of all solar PV-related patents, indicating a high level of geographic concentration in technological innovation [14].

As a result of these advancements and strong policy support, global installed PV capacity grew rapidly—from just over 1 GW in 2000 to nearly 40 GW by 2010 as shown in Figure 1.1 [15, 16]. Much of this growth was driven by feed-in tariff (FiT) policies in Europe, especially in Germany, as well as incentive programs in Japan and the United States.

During the same period, the cost of solar PV modules declined significantly—from over \$20 per watt in the early 1980s to under \$3 per watt by 2010 as shown in Figure 1.2. This dramatic cost reduction was enabled by improvements in manufacturing, driven by government subsidies, and the economies of scale, with the rise of China as a dominant

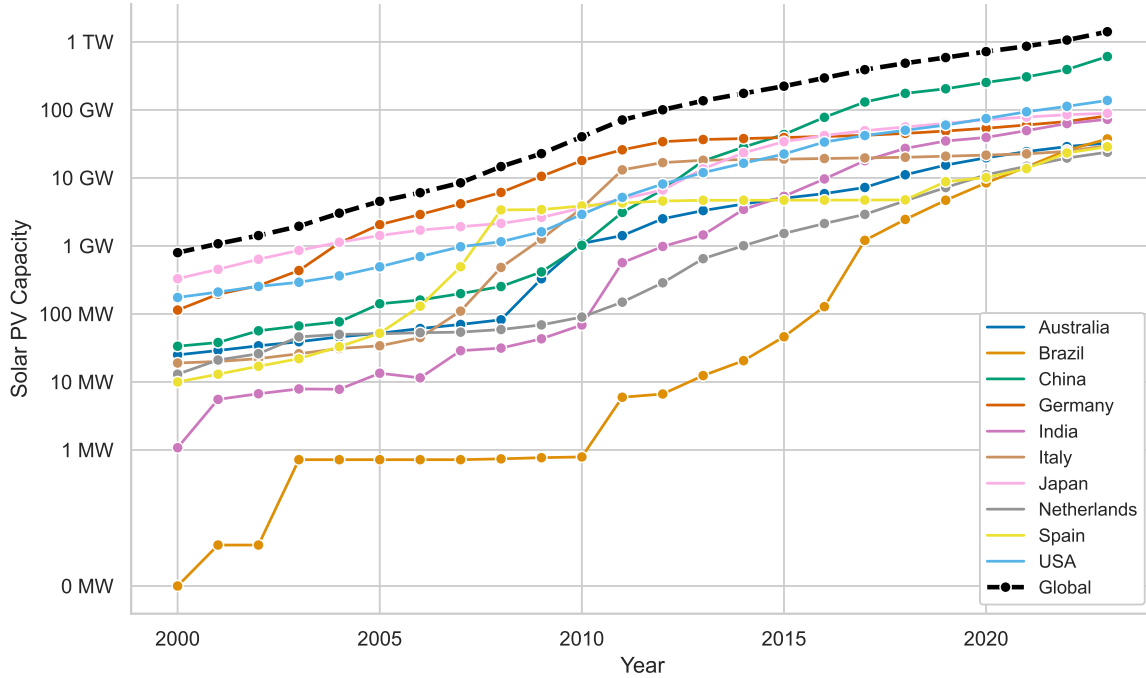


Figure 1.1: Historical deployment of solar photovoltaic capacity for the top 10 countries by installed capacity: China, USA, Japan, Germany, India, Brazil, Australia, Italy, Spain, and the Netherlands. Global cumulative capacity is shown as a black dashed line. Data sourced from IRENA [17]. In the early 2000s, Japan, the United States, and Germany led global solar PV deployment, with installed capacities significantly higher than China's. Japan's capacity in 2000 was nearly ten times greater than China's, while the United States and Germany had fivefold and threefold higher capacities, respectively. However, by 2015, this pattern had reversed: China overtook all three countries to become the world's leading market for cumulative installed solar PV capacity.

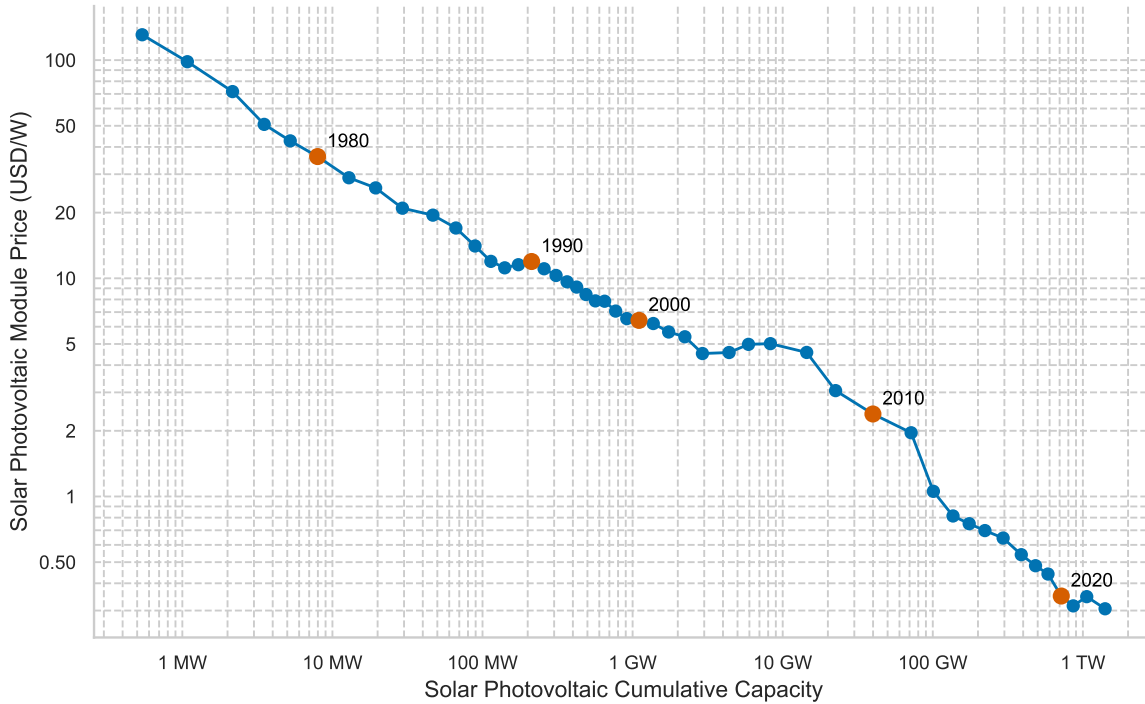


Figure 1.2: Historical evolution of solar photovoltaic module prices (USD/W) against cumulative global installed PV capacity from 1975 to 2023, demonstrating the consistent decrease in module prices with increased deployment. Data sourced from IRENA [19].

player in the global solar supply chain [18].

Since 2010, solar PV has undergone unprecedented cost reductions and deployment growth, solidifying its position as one of the most economically viable and scalable sources of electricity. Global cumulative installed PV capacity increased from 40 GW in 2010 to 1,412 GW by the end of 2023, reflecting an exponential growth trajectory [20]. As shown in Figure 1.1, global PV capacity has doubled approximately every three years. This rapid expansion has been underpinned by a dramatic decline in technology costs, particularly the price of solar panels.

According to the International Renewable Energy Agency (IRENA), the global weighted average levelized cost of electricity (LCOE) for utility-scale solar PV projects fell by 90% between 2010 and 2023—from USD 0.460 per kilowatt-hour (kWh) to just USD 0.044/kWh—making solar PV one of the most cost-effective sources of electricity worldwide [20]. This reduction has been driven primarily by a 96% drop in PV module prices, alongside significant

decreases in balance-of-plant costs, operation and maintenance (O&M) expenses, and the cost of capital. Further contributing factors include advances in module efficiency, economies of scale, and greater vertical integration across the global supply chain [20].

As a result, by 2023, the global weighted average LCOE for utility-scale solar PV was 56% lower than the average cost of electricity from fossil fuels [20]. Additionally, the total installed costs of utility-scale solar PV fell by 86% from 2010 to 2023, from USD 5,310/kW to USD 758/kW, reflecting advancements in both technology and installation processes. These significant cost reductions, combined with increased efficiency and global market growth, have established solar PV as one of the most affordable and scalable energy sources available today [20].

A central factor behind these dramatic cost reductions is the phenomenon described by Swanson’s Law, which posits that the price of solar PV modules declines by approximately 20% with each doubling of cumulative installed capacity [21]. This learning effect is illustrated by the curve presented in Figure 1.2, which captures how sustained technological innovation, increasing manufacturing scale, and efficiency improvements have driven long-term reductions in PV module costs.

In the early 2000s, solar PV technologies remained prohibitively expensive and required substantial government subsidies to stimulate market uptake. Japan, the United States, and Germany played a crucial role in this early phase by introducing policy incentives that supported deployment despite high upfront costs. These early investments catalysed global growth by driving cost-reducing innovations throughout the PV supply chain. As illustrated in Figure 1.1, Japan’s capacity in 2000 was nearly ten times greater than China’s, while the United States and Germany had fivefold and threefold higher capacities, respectively.

Between 2010 and 2015, PV module prices declined to a critical threshold that enabled solar PV to become economically competitive even in the absence of extensive subsidies. This turning point stimulated exponential global growth in installed capacity, particularly in China, where large-scale investments and aggressive industrial policies accelerated deployment at an unprecedented pace. As shown in Figure 1.1, China quickly overtook early market

leaders such as Germany and the United States, becoming the dominant force in the global solar PV landscape.

Several countries have been at the forefront of the global solar revolution, driving the rapid increase in solar PV capacity. China remains the undisputed leader in global solar capacity, with an installed capacity of 609,351 MW by 2023 [17], contributing more than 43% of the total global installed PV capacity. This dominance is largely attributed to proactive government policies [22, 23], including substantial investments, direct subsidies, feed-in tariffs, and tax incentives across the entire PV supply chain - from polysilicon production to wafers, cells, and modules. These measures have enabled Chinese companies to scale production efficiently, driving down costs and securing more than 80% of the global market share [22, 23].

In the United States, solar PV capacity has also experienced remarkable growth, reaching 137,725 MW by the end of 2023 [17]. This expansion has been driven by a combination of federal and state-level policies aimed at incentivizing renewable energy deployment. Chief among the federal mechanisms is the Investment Tax Credit (ITC), which has been instrumental in promoting investment in solar projects. Since its introduction in 2006, the U.S. solar market has expanded by more than 10,000%, underscoring the ITC's pivotal role in shaping market dynamics [24]. In addition to the ITC, the Modified Accelerated Cost Recovery System (MACRS) allows businesses to recover 10–25% of their solar system cost through tax deductions over a five-year period, further incentivizing capital investment in solar PV infrastructure [24].

State-level policies have also been crucial in supporting deployment. As of 2020, 30 U.S. states and the District of Columbia had adopted renewable or clean energy mandates, though only 12 states and D.C. had committed to achieving 100% clean energy targets by 2050 or earlier [25]. These mandates, in combination with state subsidies, net metering schemes, and renewable portfolio standards, create a diverse and multi-layered policy landscape. Federal initiatives such as the SunShot Initiative, launched by the U.S. Department of Energy in 2011, have further accelerated cost reductions and innovation, aiming to make solar power cost-

competitive with conventional energy sources [24]. Together, these policies have positioned the United States as a global leader in solar PV deployment.

Germany has long been regarded as a pioneer in solar PV deployment, with a total installed capacity of 82,000 MW as of 2023 [17]. The country’s leadership in solar energy can be traced back to a series of progressive government initiatives beginning in the early 1990s. The launch of the “1000 Rooftops Program” in 1990 marked one of the first large-scale public investments in solar PV, aiming to demonstrate the viability of rooftop solar systems [26]. Its success led to the “100,000 Rooftops Program” in 1999, which further expanded market awareness and infrastructure, laying the groundwork for widespread residential adoption [27, 28].

However, it was the introduction of the Renewable Energy Sources Act (Erneuerbare-Energien-Gesetz, EEG) in 2000 that became the defining moment for Germany’s solar PV sector. The EEG established one of the world’s earliest and most influential feed-in tariff (FiT) systems, guaranteeing long-term, above-market prices for electricity generated from solar PV and other renewable sources. This stable and investor-friendly framework catalysed rapid deployment, transforming Germany into a global leader in solar adoption throughout the 2000s and early 2010s [29]. Germany’s experience not only demonstrated the feasibility of large-scale solar integration but also served as a model for similar FiT-based policies worldwide [30].

The Netherlands has emerged as a significant player in solar PV deployment, with a total installed capacity of 24,000 MW by the end of 2023 [17]. This growth has been driven by a succession of feed-in premium and subsidy schemes aimed at stimulating both residential and utility-scale PV adoption. One of the earliest incentives was the Energy Contribution Regulation (EPR), introduced in 2001, which offered EUR 3.50 per watt-peak (Wp) and a 25% system cost discount for installations on existing homes. The abrupt termination of the EPR in 2003 triggered a rush of applications from investors seeking to secure the subsidy before its expiration [31, 32]. In the same year, the Environmental Quality of Electricity Production (MEP) scheme was introduced [31], providing a fixed premium per kilowatt-hour

(kWh) to renewable electricity producers feeding into the grid. PV received the highest level of support—EUR 0.097/kWh on top of wholesale electricity prices—before the scheme was discontinued in 2006 [33]. To maintain momentum, the government launched the Stimulation of Sustainable Energy Production (SDE) in 2008, initially targeting small-scale PV installations (0.6–3.5 kWp) with a feed-in premium of EUR 0.33/kWh for 15 years [34]. The scheme was expanded in 2009 to support systems up to 100 kWp, with tiered subsidy rates depending on system size [35]. By 2010, the total SDE budget for PV had reached EUR 94 million, supporting 25 MW of capacity [36]. However, the program was discontinued in 2011 to make way for the SDE+ scheme, which focused exclusively on systems larger than 15 kWp and adopted a more cost-competitive, technology-neutral allocation approach [36]. More recently, the Sustainable Energy Production and Climate Transition Incentive Scheme (SDE++) has supported large-scale solar deployment, while the annual net metering policy has played a critical role in stimulating residential PV adoption by allowing households to offset electricity bills with self-generated solar power [37, 38]. These policy instruments—together with decreasing technology costs and public enthusiasm—have collectively enabled the Netherlands to become one of Europe’s fastest-growing solar PV markets.

Together, these countries contributed to a global solar energy revolution, where policies, technological innovations, and economic factors drove substantial increases in installed capacity. As solar technology continues to evolve and costs continue to fall, it is expected that even more nations will make significant strides in increasing their solar PV installations, helping to mitigate climate change and promote a cleaner, more sustainable energy future.

1.2 Significance and Challenges of Tracking Solar PV Capacity

1.2.1 Tracking National PV Capacity

Several key organizations and institutions provide national-level PV capacity data, each employing distinct methodologies, classification systems, and update frequencies. These data

sources vary in their degree of openness, detail, and reliability, which can affect their suitability for different analytical purposes.

The International Renewable Energy Agency (IRENA) is one of the most widely used sources for global solar PV capacity data. IRENA publishes annual statistics on installed and cumulative renewable energy capacity, disaggregated by technology type, for more than 180 countries through its Renewable Capacity Statistics reports [39]. These statistics primarily reflect capacity that is installed and connected at the end of each calendar year, though in some cases they may also include completed projects that are not yet grid-connected. IRENA gathers data from a mix of sources, including national statistical offices, energy ministries, regulators, industry associations, and its own questionnaire-based surveys. Figures in IRENA’s datasets are annotated to indicate their provenance: “o” for official statistics, “u” for unofficial sources such as industry reports and news media, and “e” for figures estimated by IRENA from various sources [40]. While this multi-source approach enables broad coverage and comparability, it also introduces variability in data quality, depending on the accuracy and completeness of national submissions.

The International Energy Agency (IEA) also publishes solar PV capacity data through its “Renewables 2024: Analysis and Forecast to 2030” report [41], and the World Energy Outlook. The IEA dataset includes both historical trends and forward-looking deployment scenarios, covering over 40 countries and regional aggregates such as the European Union and G20 [42]. One of the key features of IEA’s data is the segmentation of PV systems into detailed subcategories. These include residential PV systems (PVRES, <10 kW), commercial and industrial systems (PVCOMM, 10 kW–1 MW), utility-scale systems (PVUT, >1 MW), and off-grid systems (PVOFF), which include solar home systems, mini-grids, and remote commercial setups. Distributed PV systems (PVDIST) encompass all installations under 1 MW, including residential, commercial and industrial systems, and off-grid application types. The IEA also tracks emerging categories such as PV systems dedicated to hydrogen production (PVH2), reflecting the growing integration of solar energy into green hydrogen strategies. Importantly, IEA distinguishes between grid-connected and off-grid systems in

its methodology: grid-connected systems are counted at the time of connection, while off-grid systems are included at installation. This consistent classification supports international comparability across countries and system types [42]. BloombergNEF (BNEF) is a commercial provider of solar market data, offering high-frequency updates on national PV capacity and project pipelines, including insights into future developments. BNEF’s data is widely used by financial analysts, policymakers, and researchers.

Taken together, these data sources offer varying levels of detail, access, and reliability. While organizations like IRENA support open access and broad global comparability, their data is limited by the quality of national reporting. The IEA and BNEF provide more granular and segmented data, including forecasts and market breakdowns, but their commercial nature restricts accessibility. National statistical agencies remain essential for fine-grained capacity tracking, although their methodologies and data structures differ.

Mao et al. [43] compared methods for estimating PV installation and mentioned the advantages and disadvantages of each method. Several methods exist for estimating PV installations, each with distinct advantages and limitations. Official registers typically provide comprehensive and reliable information for centralized PV systems—large-scale installations connected to the high-voltage transmission or primary distribution grid—including detailed data on area, capacity, and tilt angle. However, they often lack complete records for distributed PV systems, which are smaller, typically rooftop installations connected to the low- or medium-voltage distribution network and often not visible to the system operator. In addition, updates regarding decommissioned installations may be delayed. Crowdsourced field surveys allow for flexible, small-area data collection and can capture a wide range of PV system characteristics. Yet, the quality of volunteer-submitted data can vary significantly, and maintaining or updating the dataset is labor-intensive and costly. Behind-the-meter analysis offers the advantage of directly predicting the actual power output of PV systems, which is valuable for grid management. Nonetheless, it suffers from high prediction errors for distributed systems, limited access to meter-level data, and complications from battery storage, making it unsuitable for estimating installation potential. Lastly, satellite and aerial

imagery analysis enables broad, relatively low-cost coverage with high accuracy in identifying location, area, and system orientation. It offers fast processing and easy updates, and it can also support assessments of untapped PV potential when combined with building data. However, this method requires significant computational resources and does not directly measure system capacity; rather, it captures panel surface area, from which capacity must be inferred using empirical relationships or estimation models. Additionally, global access to high-resolution imagery remains limited.

1.2.2 Tracking Regional and Subregional PV Capacity

Tracking regional and subregional solar PV capacity presents greater challenges than monitoring national-level deployment. Data availability, resolution, and methodological consistency vary significantly across countries and data providers. While some nations have developed solutions to monitor solar capacity at fine geographic scales, others rely on national aggregates with limited spatial granularity. Nonetheless, several countries in Europe—including the United Kingdom, Italy, Spain, Belgium, Germany, and France—provide regional PV capacity data through a combination of government agencies, transmission system operators (TSOs), and research institutions.

In the United Kingdom, high-resolution solar PV capacity data is made available through Sheffield Solar, a research group based in the Department of Physics and Astronomy at the University of Sheffield [44]. The group publishes estimates of installed solar capacity at the Lower Layer Super Output Area (LSOA) level, updated quarterly with monthly temporal resolution. These estimates are compiled from a combination of sources, including the Microgeneration Certification Scheme (MCS) for domestic and small commercial systems [45], the Feed-in Tariff (FiT) register for systems that previously qualified for tariff payments [46], and the Renewable Energy Planning Database (REPD) for utility-scale projects [47]. Additional data is sourced from Solar Media Ltd’s commercial dataset [48]. This integrative approach ensures broad coverage of the UK’s solar fleet and enables robust spatial and temporal analyses of PV deployment. However, there are still known gaps in the 50 kW

to 1 MW range due to incomplete reports in the source data. Despite the extensive data coverage, capacity estimates remain subject to several sources of uncertainty that affect both domestic, commercial, and utility-scale PV systems [49]. One key issue is the presence of unreported systems—installations that are operational but missing from official registers and site lists. Conversely, some systems included in the datasets may have been decommissioned yet still appear as active entries, leading to potential overestimations of installed capacity. Transcription errors also contribute to uncertainty, as inaccuracies in reported capacity may result from rounding or data entry mistakes. Additionally, revisions to system size following installation introduce ambiguity, since such changes—such as upgrades or partial removals—are not always captured. The operational status of systems adds further complexity: some may be temporarily offline due to component failures, maintenance, or grid frequency issues, reducing their contribution to generation despite being listed as installed. Finally, network outages—whether planned or unplanned—can result in short-term disconnections that are not reflected in static capacity records. These factors underscore the importance of interpreting PV capacity data with caution, particularly when conducting spatial or temporal analyses.

In Italy, regional and provincial solar PV capacity data is published by Terna, the national transmission grid operator [50]. These statistics are available from the year 2000 onwards and are updated annually, with a temporal resolution of one year. The dataset includes geographic detail down to the province level, making it useful for regional planning and historical analysis.

Spain’s TSO, Red Eléctrica de España (REE), provides PV capacity data disaggregated by region and available at both monthly and annual resolutions [51]. The dataset includes capacity figures for multiple renewable technologies, including solar PV, and is updated monthly, making it one of the more temporally detailed sources among European TSOs.

In Belgium, Elia, the national transmission system operator, publishes regional solar PV capacity data at the provincial level, corresponding to NUTS 2 regions [52]. While Elia provides highly granular generation and forecast data—updated every 15 minutes—the

frequency with which capacity data is revised is not explicitly documented on its Open Data Portal. This creates some uncertainty about the timeliness of installed capacity figures, although the data remains useful for spatial analyses.

Germany offers one of the most detailed public datasets through its Marktstammdatenregister (MaStR), maintained by the Federal Network Agency (Bundesnetzagentur, BNetzA) [53]. The MaStR tracks all power and gas units across the country, including solar PV systems, and provides geographic data at multiple levels, including postal codes and, in some cases, exact coordinates. However, full spatial precision is generally limited to large-scale systems, while residential systems often lack coordinate-level information. The register is updated daily, providing near-real-time access to newly registered systems and allowing continuous tracking of capacity changes. The reliability and completeness of this dataset were evaluated by Kotthoff et al. [54], who conducted a comprehensive analysis of data quality in the MaStR. Their study applied a series of validation tests covering structural consistency, system sizing, and geolocation accuracy. Basic tests ensured that critical fields were populated, unit IDs were unique, and net power did not exceed gross or inverter power. In addition, they verified the validity of municipality IDs and postal codes using regular expression formats, identifying only minor structural issues (e.g., 30 PV systems missing location codes). System size validation involved checking correlations between installed capacity, inverter size, module count, and land area for ground-mounted systems. While most records passed these checks, some inconsistencies were noted: 0.6% of systems (23,251 units) showed extreme mismatches between module and inverter power, 1.12% had anomalous module counts, and 21.3% of ground-mounted PV systems exhibited unrealistic land-to-capacity ratios. Additional scrutiny of balcony PV systems revealed classification and capacity errors affecting nearly 10,000 units. Geolocation accuracy was also assessed, though limited to systems above 30 kWp, for which coordinate data are publicly available. Within this subset—representing just 5% of all PV installations—2.9 GW of capacity was found to be geospatially misaligned with its reported municipality. Technology-specific checks further flagged 1,169 PV systems with implausible installation years (e.g., prior to 1980), which con-

tradict regulatory expectations. Overall, while MaStR is an exceptionally rich and regularly updated resource for tracking PV deployment in Germany, the analysis by Kotthoff et al. [54] highlights targeted issues—particularly in system sizing and geolocation—that should be considered when using the data for detailed spatial or capacity analyses.

In France, the Open Data Réseaux Énergies (ODRÉ) platform serves as the central repository for electricity production and storage data, sourced from a range of national energy operators [55]. Solar PV capacity is reported at the IRIS level—the highest statistical resolution available for installations under 36 kW. IRIS (Ilots Regroupés pour l’Information Statistique) represents small-scale geographic units similar to census tracts [56]. While the spatial resolution is high, the data is only updated annually, limiting its use for near-term monitoring.

In the United States, the Energy Information Administration (EIA) provides detailed data on solar PV capacity through two complementary datasets: EIA-860 and EIA-861M [57, 58]. Form EIA-860 collects generator-level information for existing and planned electricity generators with a total nameplate capacity of 1 MW or greater. It includes technical details such as generator size, fuel type, geographic location, and technology type (e.g., crystalline silicon). In contrast, EIA-861M focuses on distributed PV systems smaller than 1 MW, providing estimates of both installed capacity and generation from net-metered and non-net-metered installations. These smaller systems are not covered by EIA-860. Their inclusion via EIA-861M ensures a more comprehensive accounting of solar generation across the U.S. The EIA estimates the EIA-861M capacity using standardized formulas and adjustment procedures detailed in its technical documentation [59].

In academic research, Stowell et al. [60] presents a high-coverage, open-access geographic dataset of more than 260,000 solar PV installations across the UK, developed through a large-scale crowdsourcing campaign. The dataset focuses especially on small-scale domestic systems, which had previously been poorly documented in official sources. As of September 2020, the dataset explicitly maps 10.66 GW of installed capacity, and—after accounting for missing capacity not directly observed—estimates the total UK capacity at 13.93 GW. This

corresponds to approximately 86% of the national installed capacity at that time. Compared to traditional administrative sources, the dataset provides improved spatial precision and metadata, enhancing its value for energy planning, forecasting, and grid management. Although the dataset is UK-specific, the methodology is globally applicable and designed to support continual updating as solar deployment evolves.

These examples demonstrate the diversity in approaches to regional PV capacity tracking across Europe and the US. They also highlight the importance of integrating multiple data sources—ranging from official registries and TSO statistics to crowdsourced and commercial datasets—to achieve high-resolution, reliable, and up-to-date information. Although significant progress has been made in certain countries, the availability and quality of subnational solar PV data remain uneven, posing challenges for comparative research, infrastructure planning, and policy evaluation at regional scales.

1.2.3 Significance of Tracking Solar PV Capacity

Accurate and timely tracking of solar PV capacity is essential for the effective operation of electricity grids, reliable energy forecasting, informed policy design, and long-term infrastructure planning. As solar PV becomes an increasingly prominent share of electricity generation across many countries, the availability of high-quality, spatially and temporally resolved data on installed capacity has become critical. This section outlines key applications of capacity data, progressing from real-time operational requirements to long-term strategic planning.

Real-Time Grid Operation and Balancing

The growing share of solar PV in electricity systems presents new challenges for real-time grid operation. Unlike conventional power plants, solar PV is weather-dependent, distributed, and often embedded in local distribution networks, making it less visible to transmission system operators. Without accurate and geographically resolved capacity data, grid operators struggle to estimate the power outturn of embedded generation, which appears as a reduction in demand, thereby introducing significant uncertainty into power flow forecasts [61]. This

complicates the task of maintaining grid stability and balancing [62, 49].

Solar PV monitoring services estimate generation at national or regional scales by multiplying yield by installed capacity [49]. Yield refers to the normalized output of PV systems relative to the total nominal capacity of the fleet, typically expressed in megawatts (MW) per megawatt-peak (MWp). A basic approach to determining yield involves collecting generation data from a sample of PV systems - referred to as reference systems - and statistically scaling these outputs to estimate broader regional or national yields [63, 64, 65, 66]. More sophisticated approaches integrate meteorological data with statistical or physical models to derive yield estimates, which are subsequently upscaled in a similar manner [64, 67, 68, 69, 70, 71]. There are also alternative methods that bypass upscaling entirely [61, 72, 73], though these are less widely used due to constraints such as the need for historical data from every PV site or the requirement of detailed information on the orientation of PV modules.

Regardless of the methodology employed, accurate data on installed capacity is a fundamental requirement. However, many nations still lack reliable and comprehensive records of regional PV installations, posing challenges for effective grid operation and policy development. While countries such as the United Kingdom, France, Germany, Spain, Italy, and Belgium do publish regional PV capacity data [44, 50, 51, 52, 53, 55], a notable share of the total capacity often remains unattributed to specific geographic areas. Furthermore, even where such data is available, inaccuracies in capacity records can significantly affect the accuracy of estimated solar PV generation at the regional level [49].

Grid Connection and Infrastructure Planning

The ability to forecast where solar PV capacity will be deployed is critical for planning new infrastructure and processing grid connection requests. In many countries, the rapid pace of solar development has outstripped the expansion of electricity infrastructure, resulting in widespread grid congestion and connection delays. For example, in Great Britain, a total of 732 GW of projects were awaiting connection to the transmission network as of October 2024, with renewable energy projects comprising approximately 363 GW of the queue [74].

This backlog is partly driven by speculative behaviour, whereby developers submit multiple applications for similar projects across different locations to increase their chances of securing a connection. Given the uncertainty around where and when approvals will be granted, this strategy has become common practice. However, it artificially inflates the size of the connection queue with projects that are unlikely to materialize, creating inefficiencies for system operators. Grid operators must then allocate time and resources to assess a large volume of non-viable applications, slowing the approval process for feasible projects [75]. These issues are further compounded by limited physical network infrastructure—such as substations, transformers, and transmission lines—which constrains the grid’s capacity to accommodate new connections [76, 77]. These challenges reflect the growing strain on Great Britain’s electricity system as it undergoes rapid decarbonisation, a trend mirrored in many other countries worldwide [78].

The Netherlands is facing similar issues, with significant grid connection delays that are stalling the deployment of solar PV. This problem has been exacerbated by the rapid development of utility-scale solar projects in the north-eastern provinces, where the availability of inexpensive land has attracted concentrated investment. The resulting demand for grid connections has exceeded the capacity of the existing infrastructure, leading to widespread congestion. By late 2023, much of the Dutch electricity grid was unable to accommodate new large-scale energy users, effectively halting further PV integration in several regions [37]. This situation was underscored by Minister of Climate and Energy Rob Jetten, who stated in a letter to the Dutch Parliament that “the electricity grid in all provinces is largely full, probably full or almost full” [79].

Grid connection constraints are not unique to the United Kingdom or the Netherlands. Similar challenges have been reported in a wide range of countries, including Austria [80], Bulgaria [80], Croatia [81], Chile [82], Finland [81], France [81], Germany [81], Greece [81, 80], Hungary [80], Ireland [81], Italy [81], Poland [81], Spain [81, 83], Sweden [81], Turkey [81], and the United States [84].

Without accurate and geographically disaggregated PV capacity data, operators cannot

prioritise infrastructure upgrades or identify areas where future constraints are likely to occur. This increases costs, delays decarbonisation, and reduces investor confidence in renewable energy markets.

Evaluating Renewable Energy Policies

Tracking capacity deployment over time and space is essential for evaluating the effectiveness of renewable energy policies such as feed-in tariffs, tax incentives, net metering, and auctions. Policy implementation without a clear understanding of the complex interactions between geographic factors and PV deployment can lead to unintended consequences—such as over-deployment in areas with limited grid capacity, or failure to meet deployment targets in areas with high potential. Spain’s experience in 2008 offers a cautionary example: the introduction of generous feed-in tariffs triggered a rapid expansion in solar PV installations, placing excessive strain on the electricity grid. In response, the Spanish government introduced abrupt policy changes to curb escalating costs, which ultimately caused a market collapse in the years that followed [85, 86].

A similar outcome was observed in Japan following the launch of its national Feed-in Tariff (FiT) in July 2012. The policy led to a dramatic increase in solar PV capacity approvals, reaching nearly 70 GW by late 2014. However, the pace of deployment quickly overwhelmed the grid, particularly in regions such as Kyushu with high solar penetration. As a result, Kyushu Electric Power Company suspended new grid connection contracts in September 2014, and six other utilities soon followed. These disruptions prompted Japan’s Ministry of Economy, Trade and Industry (METI) to implement corrective measures, including curtailment policies and evaluations of grid hosting capacity [87].

Greece implemented a generous Feed-in Tariff (FiT) scheme, aligned with European directives, alongside simplified licensing procedures to encourage PV deployment [88]. This policy led to a rapid surge in PV installations between 2012 and 2013, during which installed capacity nearly tripled—from 650 MWp to almost 2.4 GWp in just 14 months [89]. However, the speed and scale of this expansion outpaced planning efforts, exposing weaknesses in grid

infrastructure and the FiT financing mechanism. These issues created a significant deficit for the market operator, prompting the government to impose additional taxation and ultimately reduce FiT rates, including for existing contracts under the so-called “new deal” [89]. As a result, PV deployment stagnated for approximately six years following March 2013 [39].

In contrast, India introduced a wide array of supportive policies—including feed-in tariffs, renewable purchase obligations, and tax incentives—but still fell short of its 2022 target of 100 GW of installed solar capacity, achieving only 54 GW [90, 91]. This gap highlights the importance not only of supportive policy design but also of effective implementation, infrastructure readiness, and continuous capacity monitoring.

Comprehensive capacity data is essential for developing predictive models that can benchmark expected deployment levels based on a country’s geographic characteristics. Such models can aid in the early identification of over- or under-deployment by comparing actual progress against projections derived from countries further along the deployment curve. This benchmarking capability supports more adaptive and geographically informed policymaking, reducing the risk of unintended outcomes.

Benchmarking Deployment and Identifying Gaps

Beyond individual policy evaluation, capacity data enables benchmarking—comparing actual deployment against what would be expected based on a region’s solar potential, infrastructure readiness, and socio-economic context. Currently, few tools offer this level of analysis.

Benchmarking supports more objective assessments of deployment performance. It helps identify underperforming regions, diagnose policy gaps, and target interventions where the solar potential remains underutilized. Setting realistic capacity targets also requires historical data and contextual geographic factors.

Strategic Planning and Land-Use Management

Regional capacity data also informs long-term planning and land-use management. The rapid expansion of utility-scale PV installations increasingly raises land-use concerns, particularly

when agricultural land is repurposed for solar energy production. This can result in direct competition between food and energy systems, potentially undermining both public support and long-term sustainability goals [92, 93, 94, 95, 96, 97]. In the absence of clear and proactive policy frameworks, such conflicts can escalate—slowing deployment and eroding public trust. A notable example is Brandenburg, Germany, where weak land-use regulations have enabled land grabbing by solar developers, triggering disputes with local farmers and generating resistance to further utility-scale projects [96].

To address these challenges, forecasting models that incorporate regional PV capacity data and geographic factors can be used to identify areas with high potential for land-use conflict. These models can inform policy decisions by highlighting regions where future deployment is likely to create tension or require additional planning safeguards. They can also support the strategic rollout of alternative deployment approaches, such as agrivoltaics, which allow for the simultaneous use of land for both food and energy production.

By combining spatial capacity forecasting with land-use and environmental data, policymakers can design more nuanced and geographically tailored solar strategies that align energy objectives with food production, biodiversity protection, and local interests.

1.2.4 Challenges of Tracking Solar PV Capacity

Accurately tracking solar PV capacity is essential for effective planning, forecasting, and policy evaluation in the context of the energy transition. However, a number of methodological and systemic challenges limit the reliability and consistency of PV capacity data. Most countries rely on administrative sources—such as planning permission applications, grid connection records, or registrations for subsidy schemes—to monitor solar installations. Yet these data sources are often incomplete, particularly when it comes to small-scale and residential systems. In the United Kingdom, for example, small PV systems can legally be connected to the grid without formal registration, meaning they are frequently omitted from official records [60]. This under-reporting poses a significant challenge given that small-scale systems can collectively contribute a substantial share of total generation capacity.

The quality and completeness of PV capacity data also vary widely across countries. The International Renewable Energy Agency (IRENA) compiles global capacity statistics using a combination of official government submissions, industry association reports, third-party publications, and its own estimates. Data entries are tagged to indicate their source—whether official, unofficial, or estimated—but the overall reliability of the dataset depends heavily on the practices of national authorities and the availability of comprehensive reporting [39, 40]. Moreover, differences in national reporting standards, changes to policy incentives (such as feed-in tariffs), and inconsistent registration procedures contribute to temporal and cross-country inconsistencies. The relative immaturity of the renewable energy sector further compounds these issues [60], as many countries are still in the process of developing robust data collection and verification frameworks.

In addition to administrative and statistical challenges, many public data sources suffer from spatial imprecision, missing metadata, and duplication errors. For example, the open geographic dataset developed by Stowell et al. [60], which provides data on over 260,000 solar PV installations across the UK, illustrates several of these issues. Within this dataset, duplication often arises due to multiple entries representing geographically separated components of the same PV system. In some cases, these entries are manually tagged as related, but more commonly, they are not explicitly linked, making it difficult to identify them as parts of a single installation. Similar issues are present in the UK’s Renewable Energy Planning Database (REPD), where it is common for a PV farm to appear in multiple entries—one corresponding to the original planning proposal and another to a subsequent extension—without clear indication that these records refer to the same site [60]. These inconsistencies introduce noise and redundancy into the data, complicating efforts to generate accurate and comprehensive capacity estimates.

In response to the limitations of traditional data sources, researchers have increasingly turned to satellite and aerial imagery combined with machine learning techniques to detect solar PV installations. Notable examples include models such as DeepSolar, which use computer vision to identify solar panels from overhead imagery [98]. These approaches offer the

potential to map PV deployment at scale and with greater geographic specificity. However, they also present their own challenges. First, the PV capacity estimated by computer vision models is typically calculated using an equation where capacity is proportional to the surface area of the detected solar array [99]. However, the proportionality constant used in this calculation is not universal and varies across studies. This variation arises because the constant depends on several array-specific factors, including the manufacturer, age and maintenance history, and the type of photovoltaic technology employed (e.g., thin-film, monocrystalline, or polycrystalline). Consequently, different studies may adopt different values or assumptions for this constant, resulting in inconsistencies in capacity estimates even when based on similar surface area measurements.

Furthermore, the accuracy of computer vision models is strongly influenced by the quality of the ground truth data used for training and validation. These datasets often contain errors or inconsistencies, including duplicate entries, mislabelled installations, or incomplete metadata. Evaluation metrics used to validate these models can also be misleading. As Hu et al. [99] point out, conventional approaches for assessing the accuracy of automated solar PV detection using satellite imagery may produce overly optimistic results due to common flaws in the validation process. These include distribution shift, where the statistical characteristics of the data differ between the model’s training or evaluation setting and its real-world deployment, leading to discrepancies in performance. Additionally, the ground truth data may be unreliable, introducing further uncertainty. The scale at which the analysis is conducted also affects accuracy, as performance can vary significantly across different levels of assessment—whether at the pixel, individual installation, or regional level—with evaluation metrics and results differing across these scales.

Finally, the practical application of satellite-based detection methods faces several resource constraints. Open-access satellite imagery, such as that provided by Sentinel [100], offers broad geographic coverage but typically lacks the resolution required to detect small-scale residential systems. Higher-resolution imagery is better suited for this task but is often costly or unavailable. Additionally, the computational resources required to run large-scale

machine learning models over extensive image datasets can be substantial, limiting the accessibility of these methods for many research institutions and policymakers [60].

1.3 Capacity Definitions and Scope of Use

Understanding and clearly defining different types of solar PV capacity is essential for interpreting data, evaluating deployment trends, and developing robust models. In this thesis, I distinguish between three key types of PV capacity: actual installed capacity, reported capacity, and modelled capacity. Actual installed PV capacity refers to the true DC capacity of PV systems that have been physically installed and connected to the grid. However, this value is rarely directly observable. Instead, most available data correspond to reported capacity figures published by governments or institutions, which may be incomplete or inaccurate due to under-reporting, duplication, or administrative inconsistencies. Reported capacity can also vary in terms of the unit used: some data are reported in alternating current (AC) capacity, which refers to the output rating of the inverter, while others are in direct current (DC) capacity, which refers to the rated output of the PV modules themselves. In many cases, the unit is not specified, leading to uncertainty about the true capacity. To overcome these limitations, this thesis uses modelled capacity estimates generated from predictive models as a consistent and scalable proxy for actual capacity. Since the actual installed capacity is unknowable in practice, comparisons are made between modelled and reported capacities to assess deployment patterns, validate model performance, and identify potential discrepancies or gaps in the reported data.

1.4 Research Focus and Overview

This thesis focuses on the development of analytical tools and models to enhance the spatial and temporal understanding of solar PV deployment. Specifically, it aims to estimate solar PV capacity across different geographic scales and time periods, and to apply these models to three key use cases: *disaggregation*, *benchmarking*, and *forecasting*.

- **Disaggregation:** The process of allocating national or otherwise unlocated PV capacity to specific subnational regions. This is essential in countries where official statistics lack spatial resolution or omit detailed regional breakdowns. Regional PV capacity data are often limited due to data privacy regulations—such as the General Data Protection Regulation (GDPR) in the UK—as well as restrictions on commercially held data. These limitations hinder direct access to granular capacity records, making disaggregation necessary to improve the spatial accuracy of generation estimates and to enable regional monitoring and planning. This use case is addressed through the European and Great Britain models presented in Chapters 4 and 5.
- **Benchmarking:** The comparison of reported PV deployment levels to modelled values derived from geographic factors—spatially varying characteristics that influence solar PV deployment, including social (e.g., population density), economic (e.g., income levels), climatic (e.g., solar radiation), and land use (e.g., availability of artificial surfaces or agricultural land) conditions. Benchmarking allows for the identification of regions that are overperforming or underperforming relative to their potential, helping to evaluate the effectiveness of past or ongoing policy interventions. It also informs the setting of realistic capacity targets and the design of targeted incentives. This use case is demonstrated across all three models developed in this thesis—the Global, European, and Great Britain models presented in Chapters 3 to 5.
- **Forecasting:** Estimating future PV deployment patterns to inform long-term policy planning, infrastructure development, and land-use management. The models developed in this thesis can be used to forecast the spatial distribution of PV capacity across regions, offering critical insight into where deployment is likely to expand. Such forecasts are essential for strategic planning, including grid reinforcement, land allocation, and the timely formulation of regulatory frameworks. In addition, forecasting helps anticipate and address challenges related to land competition, permitting processes, and public acceptance by enabling targeted policy interventions. Accordingly, the models serve not only as tools for explaining historical deployment but also as instruments for

guiding future energy transitions. This use case is addressed in the Global, European, and Great Britain models presented in Chapters 3 to 5.

To support these use cases, this thesis investigates the relationship between installed solar PV capacity and geographic factors through a series of modelling efforts. Each model is developed for a distinct spatial scale and geographical context, contributing to a broader aim of improving the accuracy and spatial resolution of PV capacity data.

Development of a Global Model for Estimating Solar PV Capacity at Country Level

Presented in Chapter 3, this model estimates national solar PV capacity worldwide based on geographical factors. It includes the creation of a Solar PV Deployment Index that compares countries' deployment levels relative to other countries with similar geographic factors. The model supports global benchmarking, the setting of realistic and geographically informed deployment targets, and the evaluation of policy effectiveness by comparing observed deployment to expected values derived from each country's geographical context.

Development of a European Model for Estimating Solar PV Capacity in European Regions

Presented in Chapter 4, this model focuses on NUTS 2 regions across Europe to support benchmarking and disaggregation at the subnational level. It provides estimates of regional solar PV capacity, including the allocation of unlocated capacity to specific geographic areas. As part of this, a Regional Solar PV Deployment Index is developed to compare deployment levels between regions with similar geographical characteristics—either across Europe or within the same country.

Development of a Great Britain Model for Estimating Solar PV Capacity in Great Britain Regions

Presented in Chapter 5, this model applies the methodology to NUTS 3 regions in Great Britain, addressing disaggregation and forecasting in a policy-relevant national context. It includes the development of a Subregional Solar PV Deployment Index to compare deployment levels across geographically similar regions within Great Britain. The model supports grid operators in generation monitoring and strategic network planning by identifying regions where capacity is likely to concentrate. Ultimately, the model contributes to resolving grid connection delays, guiding network investment, and addressing potential land-use conflicts.

Analysis of the Relationship Between Geographical Factors and Solar PV Capacity

Presented in Chapter 6, this component identifies the key drivers of solar PV deployment at multiple spatial resolutions by examining geographical factors—a term used here to encompass climatic, land-use, economic, and sociological variables. The aim is to analyse the relationship between these factors and installed solar PV capacity across different spatial scales—national, regional, and subregional—in three geographic contexts: globally, across Europe, and within Great Britain. These contributions aim to improve solar PV data availability and support the strategic planning, evaluation, and management of the global energy transition.

Chapter 2

Literature Review

This chapter provides an overview of the main factors shaping the deployment of solar PV systems, the modelling techniques used to estimate installed capacity, and the policy tools implemented worldwide to encourage solar adoption. It concludes by outlining the key knowledge gaps that this thesis seeks to address.

2.1 Factors Associated With Solar PV Deployment

The adoption and diffusion of solar PV technologies are shaped by a wide range of factors that operate across multiple geographic scales. These include climatic conditions, land-use patterns, socioeconomic variables, and institutional or policy frameworks, influencing deployment at the subregional, regional, and national levels. This section categorises these drivers into three spatially defined groups—household and subregional, regional, and national—and reviews key empirical studies within each.

2.1.1 Household and Subregional Factors

Jan, Ullah, and Ashfaq [101] identified key factors influencing household solar PV adoption in Northwest Pakistan, which include income, energy consumption costs, education level, information about the solar PV market, and source of awareness about solar PV systems. These factors accounted for 38% of the variation in adoption. This study draws on a survey con-

ducted with 100 randomly selected households across three villages in Khyber Pakhtunkhwa, Pakistan. Within the study area, 46% of households used solar PV systems. A binary logistic regression model was employed to analyse the factors influencing the social acceptance of solar PV technology.

Letchford, Kiran Lakkaraju, and Yevgeniy Vorobeychik [102] conducted sensitivity analysis using multiple methods to identify important predictors of household solar PV adoption in the San Diego region in the USA. Key factors included property size, owner occupancy, national unemployment rate, income, electricity costs, and peer effects, which explain 33% of the variation in adoption. In rural India, Aklin, Cheng, and Urpelainen [103] investigated solar PV adoption at the household level using logistic regression, and found that wealthier households with access to banking services were more likely to adopt solar power.

Robinson and Rai [104] used a GIS-integrated agent-based model to study household solar PV adoption in Austin, Texas, USA. They found that financial factors alone provided an adequate model of PV adoption, but incorporating agents' attitudes towards solar PV and social interactions significantly improved model performance. Additionally, using household-level insolation data enhanced the model's accuracy.

Fuentes, Khalilpour, and Voinov [105] used regression analysis to explore the relationship between solar PV adoption and various factors across postal areas in Australia. The findings indicated that gender, the proportion of certain age groups, land area, and dwellings with a vehicle did not correlate with solar PV adoption. In contrast, there was a positive correlation between solar PV adoption and the share of the married population, unemployment rate, population with non-school qualifications, the number of rented residential units, and the number of bedrooms. Peer effects were also observed. Negative correlations were found with the share of the population aged 25 to 40, weekly household income, population density, dwelling density, and the number of occupants per dwelling.

Alderete Peralta, Balta-Ozkan, and Longhurst [106] used an integrated neural networks and agent-based modelling approach to characterize the spatio-temporal adoption patterns of domestic solar PVs at the postcode level in Birmingham, UK, identifying income, electricity

usage, and average household size as the best predictors of solar PV adoption.

Collier et al. [107] investigated the determinants of domestic solar PV uptake at the Lower Layer Super Output Area (LSOA) level in England and Wales. To account for spatial and structural variation, the authors used a series of regression-based methods, including Ordinary Least Squares (OLS) regression, fixed effects regression with local authority dummy variables, multi-level modelling, and a spatial error model. Residential solar uptake was found to be higher among older age groups and in detached houses, while higher housing density negatively impacted uptake. Spatial effects on solar adoption were evident at both the LSOA and Local Authority levels.

Graziano and Gillingham [108] analysed the impact of multiple factors on PV adoption within census block groups in Connecticut, USA, using a linear fixed effects regression model. Neighbour influence, the built environment (housing density and the proportion of renters), and policy showed a stronger relationship with PV adoption compared to social, economic, and political factors.

Laura Williams and Mita Kerai [109] examined the factors influencing PV deployment at the LSOA level under a feed-in tariff scheme in England, UK, using a descriptive statistical approach. The findings highlighted that electricity and gas consumption, gas coverage, age demographics, the index of multiple deprivation and its domains, dwelling stock by tenure and type, urban or rural classification, council tax band, and fuel poverty are all critical in explaining solar PV deployment.

Yu et al. [98] used correlation analysis to identify key social and economic factors associated with solar deployment density at the census tract level in the USA. These included solar radiation, population density, annual household income, Gini index, and education level.

2.1.2 Regional Factors

Westacott and Candelise [110] used a GIS framework to investigate the relationship between solar PV deployment and multiple factors across regions in the UK, finding that PV deployment is strongly correlated with the level of policy support and with rural areas that have

high levels of irradiation.

Thormeyer, Sasse, and Trutnevyte [111] investigated the spatial diffusion of solar PV projects across 2,222 municipalities in Switzerland using two methods: spatial analysis of hot and cold spots and stepwise regression. Their findings revealed that the urban–rural divide and exploitable PV potential are the primary drivers of solar PV diffusion. Municipalities with a stronger focus on agriculture and forestry typically have more PV projects, whereas urban municipalities tend to have fewer. Other factors, such as voting behaviour and electricity prices, also influence adoption, but to a lesser extent.

Rigo et al. [112] analysed PV diffusion across 5,570 municipalities in Brazil using regression-based machine learning algorithms, including Light Gradient Boosting Machine and Random Forest algorithms. They identified electricity tariffs as the primary predictor of adoption, followed by solar irradiation and municipal Gross Domestic Product (GDP). Other relevant predictors included the number of companies, minimum wage, the education component of the Human Development Index (HDI), demographic density, and the vehicle fleet.

McEachern and Hanson [113] used multivariate linear regression to study solar PV adoption across 120 villages in Sri Lanka, showing that adoption is driven by expectations regarding future grid connection by the government and by tolerance for non-conformist behaviour. Similarly, Aklin, Cheng, and Urpelainen [103] used logistic regression with fixed effects to investigate the factors influencing solar adoption at the village level in rural India. They found that remote, large, and poor villages with high levels of solar radiation are more likely to adopt solar technology as a substitute for grid electricity.

Mayer et al. [114] analysed socioeconomic factors correlating with PV system adoption in 53 counties in North Rhine-Westphalia, Germany, and found a strong positive correlation between agricultural gross value added and PV adoption ($r = +0.75$). In contrast, unemployment rate and population density showed moderate negative correlations with PV adoption ($r = -0.61$ and $r = -0.64$, respectively).

Balta-Ozkan, Yildirim, and Connor [115] applied spatial econometric methods to study PV uptake at the NUTS 3 level in the UK. The study found positive correlations between

PV deployment and income per capita, education level, electricity sales, irradiation, and the share of detached houses. Conversely, a higher share of owned houses, population density, and the average number of households negatively correlated with solar PV deployment.

2.1.3 National Factors

Liu et al. [116] investigated the correlation between social and economic factors and the installed capacity of solar PV in China using Grey Relation Analysis (GRA). The analysis demonstrated that GDP, final consumer expenditure, industrial added value, solar energy generation, and solar energy consumption are all strongly correlated with PV capacity.

Celik and Özgür [117] examined the relationship between solar PV capacity, GDP per capita, and solar radiation in Turkey and five other European Union countries using descriptive and comparative analysis. The study found no direct relationship between solar PV capacity and either GDP or solar radiation.

Bunea et al. [118] applied innovation diffusion models to study the adoption of solar PVs in the UK, and found that public incentives play a critical role in driving adoption.

Kruitwagen et al. [119] explored the relationship between land cover and solar PV deployment for systems larger than 10 kW on a global scale. The findings indicated that most solar PV systems are located on cropland, followed by arid lands and grasslands.

In summary, these studies demonstrate that geographical factors—including socioeconomic, climatic, and land-use characteristics—play a significant role in shaping the adoption and deployment of solar PV systems. However, the findings also reveal considerable variability across studies, reflecting differences in methodological approaches, spatial scales, and geographic contexts. The relationship between solar PV deployment and its drivers is complex and often context-specific, influenced by local policy environments, market maturity, and demographic patterns. As such, developing generalisable models requires careful consideration of the spatial and temporal characteristics of each study area, as well as the interactions

between multiple factors. This complexity underscores the need for multi-scalar and data-driven approaches to better understand and predict solar PV adoption patterns across diverse settings.

2.2 Methods for Modelling Solar PV Capacity

2.2.1 Statistical Models

Several academic studies have proposed data-driven models for estimating solar PV capacity, though most are designed for specific countries and rely on non-generalisable datasets.

Yu et al. [98] developed a two-stage machine learning model to estimate solar PV deployment density in the United States using more than 90 socioeconomic and environmental features. The first stage uses a random forest classifier to predict the presence of PV in a census tract, and the second stage uses a regressor to estimate density. Census tracts are small, relatively permanent statistical subdivisions of a county, typically designed to contain between 1,200 and 8,000 people, with an optimum population size of around 4,000. The model achieved a cross-validation R^2 of 0.72. While the model is highly accurate, its reliance on extensive US-specific data makes it difficult to apply in other contexts. In addition, it only takes into account residential PV systems.

Liu et al. [116] developed a bidirectional long short-term memory (BiLSTM) neural network to forecast China's annual installed solar PV capacity up to 2035. The model was trained on historical data from 1996 to 2019. To extend predictions to 2035, the authors used a support vector regression model to estimate future values of the input variables which were then used as inputs to the BiLSTM model. The BiLSTM model achieved a mean absolute percentage error (MAPE) of 6%, a mean absolute error (MAE) of 6.6 GW, and a root mean squared error (RMSE) of 7.7 GW. To interpret the influence of input features, the authors performed a mean impact value analysis. The results showed that solar power generation and consumption were the most influential factors, contributing 26% and 27% respectively, followed by gross domestic product (17%), final consumer expenditure (15%),

and industrial added value (14%). While the model achieved high predictive accuracy using a relatively small set of input variables, its reliance on solar generation and consumption data limits its applicability in countries where such data are not consistently available.

Bunea et al. [118] modelled the diffusion of national solar PV capacity in the UK across residential, commercial, and utility-scale sectors between 2010 and 2021 using a Generalized Bass Model (GBM). While GBMs are effective for capturing adoption trends over time, their reliance on time-series data makes them less useful in contexts with limited or incomplete historical records. The study does not report standard error metrics such as R^2 , MAE, RMSE, or MAPE. Instead, model performance is evaluated visually using plots that compare observed and fitted values, without providing a quantitative assessment of predictive accuracy.

Li et al. [120] used logistic growth models to fit residential PV system counts across 2,670 Australian postcodes from 2005 to 2021. Principal component analysis (PCA) and cluster analysis were then applied to identify three distinct diffusion patterns. This study estimated the number of systems rather than their capacity.

Thormeyer, Sasse, and Trutnevyte [111] conducted a stepwise regression analysis on the number of feed-in tariff PV systems per 1,000 inhabitants across 2,222 Swiss municipalities in 2016. The model incorporated variables such as exploitable solar PV potential, electricity demand, socio-demographic factors, and other municipal characteristics. These factors account for 43% of the variation in the number of PV projects per 1,000 inhabitants. Further analyses were performed by disaggregating the data into German-speaking and French-speaking municipalities, and by including cantons—26 larger administrative units with lower spatial resolution than municipalities—as dummy variables. The respective models accounted for 45%, 49%, and 47% of the variation in the number of PV projects per 1,000 inhabitants. Consequently, splitting the regions by language or incorporating cantons as dummy variables enhances the model’s predictive capacity.

Davidson et al. [121] employed stepwise regression to model the logarithm of the number of cumulative residential PV systems per census block group and ZIP code in California, USA, from 2007 to 2013. Two models were developed for each spatial resolution: a best-fit

model that included all variables using stepwise regression, and a parsimonious model that used a branch-and-bound algorithm to select a subset of one to eight variables. The best-fit model at the block group level had a lower predictive performance, with an adjusted R^2 of 0.48 and an MSE of 0.48 (in $\log(\text{installations})^2$), compared to the ZIP code level model, which achieved an adjusted R^2 of 0.58 and an MSE of 0.38 (in $\log(\text{installations})^2$). The parsimonious model at the block group level reached an adjusted R^2 of 0.46, while the ZIP code level model achieved 0.49. The variables considered included population demographics, housing characteristics, foreclosure rates, solar irradiance, vehicle ownership preferences, and other relevant factors. Additionally, the study assessed the model's applicability across California's three Investor Owned Utility regions by training the model on one region and testing it on the other two. The results indicated that historical PV diffusion trends and population statistics were sufficiently consistent across regions, enabling models trained in one area to generalize well to other regions.

Müller and Trutnevyte [122] employed linear and spatial regression models to estimate installed solar PV capacity across 143 districts in Switzerland. They developed a multiple linear regression (MLR) model, a simultaneous autoregressive (SAR) model, and a spatial error model (SEM), using a comprehensive dataset comprising 68,341 PV installations recorded between 2010 and 2017. These models incorporated techno-economic and socio-demographic predictor variables to generate spatial projections of PV capacity and were evaluated through both in-sample and out-of-sample accuracy testing. The results for in-sample testing in 2017 showed high explanatory power, with R^2 values ranging from 0.88 to 0.89 and root mean squared logarithmic error (RMSLE) values around 0.3. When tested in-sample across all years from 2010 to 2017, RMSLE values ranged from approximately 0.3 to 0.8. For 1-year-ahead out-of-sample projections, RMSLE increased to a range of 0.4 to 1.2 across the models. Incorporating a time-lagged response variable (y_{t-1}) as a predictor improved out-of-sample projections, yielding RMSLE values ranging from approximately 0.1 to 1.2 for 1-year-ahead forecasts. For longer-term projections (1- to 5-years-ahead), based on model fitting using 2012 data and including the lagged response variable, RMSLE values ranged from 0.3 to 0.7.

Spatial regression models underscore the importance of accounting for spatial autocorrelation in the data, revealing spatial spillover effects among neighbouring districts. In-sample accuracy tests demonstrate that spatial regression models exhibit marginally superior performance, supporting the inclusion of spatial terms in the formulation of spatial PV forecasts. However, for out-of-sample projections spanning one to five years, the traditional multiple linear regression model performs comparably to the spatial models, suggesting that simpler approaches may be sufficient in some forecasting contexts, offering a potential reduction in analytical complexity.

These studies use a variety of methods to model solar PV deployment, ranging from diffusion of innovation models to machine learning approaches. However, these models often differ in their target variables—some estimate installed capacity, others model the number of installations, or even the logarithm of installations—making direct comparison of results challenging. In addition, model performance is reported using different error metrics, and in some cases, no quantitative error metrics are reported at all. Many of these models are developed for specific locations and rely on proprietary or country-specific datasets, which are not available for most other regions. This limits the transferability of the methods, particularly in data-scarce contexts. Furthermore, some studies focus on specific market segments (e.g., residential, commercial, or utility-scale systems), while others model aggregate national capacity, further complicating cross-study comparisons. To overcome these limitations, this thesis aims to develop a generic and transferable modelling framework based on open-access, globally available datasets. Such a framework would support comparative assessments and enable consistent application across regions with varying levels of data availability.

2.2.2 Vision-Based Models

Computer vision is increasingly used to detect solar PV installations in satellite and aerial imagery, offering a scalable and efficient alternative to manual surveys. According to Mao et al. [43], PV detection tasks can be grouped into three main types: image classification, object detection, and semantic segmentation. Image classification simply determines whether

or not a PV system is present in an image, classifying it as either positive or negative. Object detection goes a step further by identifying where PV systems are located within the image and estimating how many there are. Semantic segmentation offers the highest level of detail by analysing every pixel in the image to determine whether it belongs to a PV system, which allows for accurate estimation of the system's area. To carry out these tasks, researchers typically use either conventional machine learning techniques or deep learning methods. Conventional approaches are further divided into two types: pixel-based image analysis (PBIA), which looks at each pixel individually, and object-based image analysis (OBIA), which groups pixels into meaningful objects before classifying them. Deep learning methods, particularly convolutional neural networks (CNNs), have become popular due to their ability to automatically learn features and handle complex image patterns with high accuracy.

To evaluate the performance of these computer vision models, a range of error metrics are employed depending on the nature of the task. In classification-based tasks, such as image classification or object-level detection, precision, recall, and the F1 score are commonly used. Precision measures the proportion of correctly identified instances among all predicted positives, reflecting how accurate the model's detections are. Recall assesses the proportion of actual positives that were correctly identified, indicating the model's ability to detect relevant instances. These two metrics often trade off against each other, especially in imbalanced datasets where one class dominates. The F1 score, defined as the harmonic mean of precision and recall, provides a balanced summary metric that penalizes extreme values in either. A high F1 score indicates that the model achieved a strong balance between detecting the majority of PV installations and minimizing both false positives and false negatives. For tasks involving spatial localization—such as object detection and semantic segmentation—the Jaccard Index, also known as Intersection over Union (IoU), is widely used. IoU quantifies how well the predicted region aligns with the ground truth by dividing the area of overlap (intersection) by the area covered by both predicted and true regions (union). The resulting value ranges from 0 to 1, where a score of 1 indicates perfect spatial alignment and 0 indicates

no overlap. This metric is particularly relevant for PV mapping from imagery, as it allows researchers to assess not only whether a PV system is detected, but also how accurately its shape and location are captured. The Dice coefficient is another commonly used spatial metric that quantifies the similarity between predicted and ground truth regions; it is calculated as twice the area of intersection divided by the sum of the predicted and ground truth areas.

Pixel-Based Image Analysis (PBIA)

Several studies have applied pixel-based image analysis (PBIA) methods to detect rooftop PV systems using high-resolution satellite imagery. Ji et al. [123] used 1.2 m resolution imagery from Oldenburg, Germany, and applied a statistical classification approach, achieving overall accuracy ranging from 92.8% to 99.3%. Karoui et al. [124] worked with 0.84–1.6 m resolution images over a section of Toulouse, France, using multi-part nonnegative matrix factorization, resulting in a Normalized Mean Square Error (NMSE) of 23.73% for PV detection. Malof et al. [125] employed random forest classification on 0.3 m imagery from Fresno, USA, reporting a precision of 60% and recall of 70%.

Object-Based Image Analysis (OBIA)

Several studies have utilized object-based image analysis (OBIA) methods to detect PV systems, focusing on both rooftop and large-scale centralized installations. Malof et al. [126] used high-resolution (0.3 m) imagery from 100 building images in Lemoore, CA, USA to detect rooftop PV systems, achieving a recall of 94% using a Support Vector Machine (SVM) classifier. For centralized PV systems, Zhang et al. [127] employed random forest classification on 30 m resolution imagery in Ningxia Province, China, identifying PV areas larger than 0.21 km² with a precision of 98.53% and recall of 92.19%. Similarly, Plakman, Rosier, and Vliet [128] used 10–20 m imagery in the Netherlands and applied semantic segmentation via random forest to detect centralized PV systems larger than 1034 m², achieving a recall of 85.86%, precision of 92.39%, and a Jaccard Index (IoU) of 80.19%. In Denmark, Vasku [129] evaluated multiple classifiers—including Classification and Regression Trees (CART), Ran-

dom Forest, Support Vector Machine (SVM), and Naïve Bayes—on 10 m resolution imagery to detect large centralized PV fields ($>12\text{ km}^2$). The CART model achieved a precision of 70.73% and recall of 52.94%. Both Random Forest and SVM produced the same precision of 71.60%, while recall was 54.90% for Random Forest and 52.94% for SVM. The Naïve Bayes classifier performed worse, with a precision of 63.64% and recall of 37.25%. Finally, Wang et al. [130] used ultra-high-resolution (0.05 m) drone imagery and applied a region–line primitive association and template matching approach for semantic segmentation of centralized PV systems, reporting both precision and recall exceeding 99%.

Deep Learning

Semantic segmentation is a key approach in detecting solar PV installations, as it enables pixel-level classification of imagery to accurately outline the boundaries of PV systems. This detailed extraction allows for estimating both the area and, indirectly, the capacity of PV installations by counting the number of pixels identified as PV. Semantic segmentation models for PV detection using deep learning can be categorized according to their training methods and network structures. Based on training strategies, models fall into three main types: fully supervised, weakly supervised, and unsupervised learning. Fully supervised models require large amounts of labelled training data and include various architectures such as models based on candidate regions, fully convolutional networks (FCNs), and recurrent neural networks (RNNs). FCN-based models themselves can be further divided into subtypes, including those using symmetric network structures, convolutional extensions, and residual networks with fusion features. In contrast, weakly supervised and unsupervised models aim to reduce reliance on detailed annotations, making them promising for scaling PV detection across large areas with limited labelled data [43].

The object detection approach based on candidate regions, initially developed for general detection tasks, has evolved to support semantic and even instance segmentation. Beyond these models, Mask-RCNN, which was originally designed specifically for instance segmentation, has also been applied to segment solar PV installations.

Several studies have applied candidate region-based methods, particularly Mask R-CNN, to segment solar PV installations across various contexts. Moradi Sizkouhi et al. [131] applied Mask R-CNN with a VGG16 backbone across 12 countries, focusing on centralized PV systems, and achieved an impressive accuracy of 96.99%. In computer vision, a *backbone* refers to a pre-trained neural network architecture primarily used for feature extraction from images [132]. Schulz, Boughattas, and Wendel [133] implemented Mask R-CNN with a ResNet101 + Feature Pyramid Network (FPN) backbone in Bad Brambach and Chemnitz, Germany, targeting all types of PV systems. Their model achieved strong results for rooftop PV, with a precision of 81%, recall of 73%, F1 score of 97%, and IoU of 77%. For free-field PV, the model reached a precision of 97%, recall of 72%, F1 score of 83%, and IoU of 86%. Liang et al. [134] used high-resolution imagery (0.15 m) in the United States to segment various scales of PV using a series of refined Mask R-CNN models with a ResNet50 + FPN backbone. The basic Mask R-CNN achieved a precision of 95.5%, recall of 92.1%, and IoU of 86.5%. When enhanced with an overlap-title mechanism, precision and recall improved to 95.7% and 93.5%, respectively, with an IoU of 87.4%. Adding a right-angle polygon fit algorithm further boosted performance, reaching a precision of 96.2%, recall of 95.5%, and IoU of 88.8%.

A wide range of semantic segmentation studies have applied Fully Convolutional Networks (FCNs) and FCN-based architectures to detect solar PV systems at varying resolutions, locations, and PV scales. Ishii et al. [135] used 30 m resolution images in Japan to detect centralized PV systems (>5 MW) using an FCN model, achieving an IoU greater than 50%, precision above 60%, and recall exceeding 75%. Edun et al. [136] focused on both centralized and distributed PV arrays across 387 locations in the U.S., using images with a resolution of 0.596 m and an encoder-decoder architecture with VGG16, obtaining 98.66% accuracy and a Dice coefficient of 81.74%. In the city of Fresno, Camilo et al. [137] tested rooftop PV detection using both CNN and SegNet (both with VGG backbones); when recall was fixed at 80%, CNN yielded about 50% precision, while SegNet significantly improved that to about 90% precision.

Pérez-González, Jaramillo-Duque, and Cano-Quintero [138] applied both Fully Convolutional Networks (FCN) and U-Net architectures across 12 countries to detect centralized PV systems. Using FCN, they achieved a recall of 94.16%, an IoU of 87.47%, and a Dice coefficient of 89.61%; with U-Net, performance improved slightly to a recall of 95.44%, IoU of 90.42%, and Dice of 91.42%. The study used the Amir dataset [139], which was also used by Moradi Sizkouhi et al. [131].

Kruitwagen et al. [119] applied a combined U-Net + ResNet50/RNN model to 1.5 m and 10 m resolution images globally to detect centralized PV systems (>10 kW). U-Net was used for semantic segmentation, while ResNet and RNN served as image classifiers to filter false positives. For installations with an area larger than $10,000 \text{ m}^2$, the model achieved 90% recall, 90% IoU, and 98.6% precision.

Wu and Biljecki [140] built a model to detect rooftop solar PV across 17 global cities using 0.5 m resolution images and a U-Net architecture in which the encoder layer was replaced by a pre-trained ResNet50 model. The model achieved a count recall of 91.9% and an area recall of 96.25%.

Castello et al. [141] developed a model to detect rooftop PV systems from 0.25 m resolution images in Switzerland, using a U-Net architecture. The model achieved an IoU of 64%, an accuracy of 94%, and an F1 score of 80%.

Zech and Ranalli [142] developed a model to detect rooftop solar PV systems from images with a resolution of 0.2 m in Oldenburg, Germany. A U-Net architecture was used with various backbones, including ResNet18, ResNet34, ResNet50, and ResNet101. The best results were achieved using U-Net with a ResNet50 backbone, yielding an IoU of 69%, precision of 84%, recall of 79%, and an F1 score of 81%.

Jie et al. [143] developed a model to detect rooftop PV systems using high-resolution aerial imagery in two regions: Fresno, Stockton, and Modesto in California (0.3 m resolution), and Songjiang and Pudong New districts in Shanghai, China (0.1 m resolution). The model combines multi-layer features with a gated fusion module and an edge detection network. In California, the model achieved an IoU of 73.60%, precision of 86.17%, recall of 83.45%, and

an F1 score of 84.79%. In China, where the imagery resolution is higher, the model achieved an IoU of 88.74%, precision of 93.88%, recall of 94.19%, and an F1 score of 94.03%. These results highlight the significant impact of image resolution on model accuracy. However, high-resolution imagery is not universally available, and when available, it is often not updated frequently enough.

Estimation of PV Capacity

Computer vision models usually compute PV capacity using an equation in which the capacity is proportional to the identified solar array’s surface area [99]. Yet, the proportionality constant utilized in this computation is not standardized and differs among studies. This difference stems from the fact that the constant is influenced by several factors specific to the array, such as manufacturer, age and maintenance record, and the photovoltaic technology type used (such as thin-film, monocrystalline, or polycrystalline). Therefore, studies might apply varying values or assumptions for this constant, leading to discrepancies in capacity estimates, even when surface area measurements are similar.

While computer vision models are effective for detecting existing solar PV installations in aerial or satellite imagery, they are primarily designed for identification, not forecasting. Although the surface area data generated by these models can support capacity estimation, several limitations hinder their reliability. Typically, capacity is estimated using a fixed ratio between the detected surface area and installed capacity [99]. However, this ratio is not universal and varies across studies due to differences in PV technology, manufacturer specifications, system age, and maintenance practices. As a result, even with accurate area detection, capacity estimates can vary significantly depending on the assumptions used.

Beyond these methodological challenges, practical limitations also constrain the utility of computer vision approaches for this study. These models require high-resolution imagery, which is often costly, inconsistently available, and not captured on a regular basis. Their accuracy depends heavily on the quality of training data and labelling, and they require substantial computational resources for implementation at scale. Given these constraints,

we chose not to adopt a computer vision-based approach in this research. Instead, we focus on modelling techniques that use geographic, environmental, and socio-economic variables, which are more widely available and better suited to estimating and forecasting PV capacity over space and time.

While both computer vision and statistical models are employed to estimate solar PV deployment, they differ in the types of accuracy they offer and the contexts in which they perform best. Statistical models report a wide range of accuracy depending on the input data and modelling approach. For example, spatial regression models have achieved in-sample R^2 values up to 0.89 and RMSLE values as low as 0.3 for estimating installed capacity across Swiss districts [122]. Machine learning models such as DeepSolar reported cross-validation R^2 of 0.72 at the census tract level in the United States [98], while BiLSTM models in China reached a mean absolute percentage error (MAPE) of 6% [116]. These models benefit from interpretable results and can be applied to forecasting, but their accuracy depends heavily on the availability and quality of structured socioeconomic and energy data.

By contrast, computer vision models excel at identifying visible PV systems in satellite or aerial imagery and are typically evaluated using image-based metrics such as precision, recall, F1 score, and IoU. For example, U-Net models have achieved IoU values ranging from 64% to over 90%, with recall values between 69% and 96%, depending on image resolution and geographic context [142, 138, 140]. Mask R-CNN models, especially those incorporating advanced post-processing, have reported IoU values up to 88.8% and recall exceeding 95% [134]. However, performance tends to decline when image resolution is lower or when training labels are less detailed. For instance, some pixel-based and object-based image analysis models report precision or recall as low as 50–70% [125, 129]. While effective at detecting the presence and geometry of PV systems, these models estimate capacity indirectly via surface area, which introduces uncertainty due to variation in technology type, panel tilt, system age, and maintenance.

In summary, statistical models offer stronger performance for direct capacity estimation and forecasting—especially when structured data is available—while computer vision models

provide detailed spatial coverage and detection of visible systems, with accuracy sensitive to image resolution and system characteristics. Their complementary strengths suggest potential for integrated approaches that combine spatial detection with statistical inference.

2.3 Solar PV Policies

The expansion of solar PV technology has been driven not only by technological advances and cost reductions but also by a range of supportive policy instruments. At both national and regional levels, governments have played a pivotal role in shaping solar PV deployment through targeted regulatory and financial interventions. This section examines the key policy mechanisms that have influenced the growth of solar PV, focusing on feed-in tariffs, net metering, renewable portfolio standards, and certificate-based schemes. These policies have collectively underpinned the emergence of solar energy as a mainstream component of the global electricity mix.

2.3.1 Feed-in Policies

Feed-in policies, encompassing feed-in tariffs (FiTs) and feed-in premiums (FiPs), have played a pivotal role in the development of solar PV markets. These mechanisms are designed to incentivise investment in renewable electricity generation by ensuring stable and predictable revenue streams for producers.

FiTs guarantee a fixed payment rate for each unit of electricity generated from renewable sources, such as solar PV, over a specified contractual period. This financial certainty mitigates investment risks and has historically spurred rapid deployment of PV systems. FiTs have been the primary policy tool supporting renewable electricity across OECD countries [144], and remain the only mechanism proven to successfully drive the development of large-scale gigawatt-level renewable energy markets [145].

FiPs serve as an alternative to feed-in tariffs by offering a supplementary payment in addition to the electricity market price, rather than guaranteeing a fixed rate [146]. This approach preserves market responsiveness while still providing financial support to renewable

energy producers.

Recent empirical studies continue to demonstrate the impact of FiTs and FiPs on investment behaviour. Lin and Xie [147] examined the impact of FiT subsidies on investment in the renewable energy sector, using data from 39 Chinese renewable energy generation companies between 2014 and 2021. The findings reveal that FiT subsidies have a strong positive influence on corporate investment, with each unit increase in subsidies being associated with a 3.4-unit rise in investment. This relationship is mediated by improvements in cash flow, increased profitability, and reduced financing constraints. The study also highlights the role of external environmental factors: economic policy uncertainty diminishes the effectiveness of FiT subsidies, while stronger environmental regulation amplifies their positive impact. Additionally, the investment-enhancing effect of FiT subsidies is found to be more significant in large, state-owned renewable energy enterprises.

In the European context, Alolo, Azevedo, and El Kalak [148] investigated the impact of Feed-in-System (FIS) policies on renewable energy investments in the European Union (EU) between 1992 and 2015 and provided valuable insights into the role of policy design in influencing capacity development for wind and solar PV technologies. Drawing on panel data from 27 EU countries and employing a fixed-effects model, the authors introduce a novel subsidy performance indicator (PvRev) that distinguishes between FiTs and FiPs. This indicator incorporates critical contractual and market factors, including tariff price, contract duration, digression rates, electricity market prices, production costs, and interest rates. The study finds that a 5% increase in PvRev is associated with a 28.9% increase in solar PV capacity. These results suggest that the effectiveness of FIS policies is not determined merely by their existence, but rather by the specific features of the contracts and the prevailing market conditions. Poorly designed subsidy mechanisms are shown to have limited impact on investment, highlighting the necessity for carefully calibrated policy instruments that reflect local economic and energy market dynamics. Additionally, the study underscores the importance of revenue certainty in influencing investment decisions, noting that uncertainties arising from political, economic, and technological factors can significantly

affect the attractiveness of renewable energy projects.

Dijkgraaf, Dorp, and Maasland [149] evaluated the effectiveness of FiT policies in promoting PV solar energy development across 30 OECD countries between 1990 and 2011, using panel data analysis. Unlike previous studies, which often focused on the average impact of FiTs, this research placed specific emphasis on FiT design features—such as tariff level, contract duration, the presence of capacity caps, and overall policy consistency—to assess their influence on per capita annual additions of PV capacity. The findings confirmed that FiTs, on average, contributed positively to PV capacity development, resulting in a 108% increase in per capita PV capacity relative to the sample average when a FiT was in place. However, the effectiveness of FiTs varied substantially depending on their design. A well-designed FiT, combining a high tariff, long contract duration, and consistent application, was found to be up to seven times more effective than the average FiT. In contrast, the study found that effectiveness declined sharply when there was high variability—measured by the standard deviation—in tariff levels and contract durations. In extreme cases, this inconsistency resulted in a negative effect equivalent to approximately 588% below the sample average. The presence of a capacity cap also significantly reduced effectiveness, lowering the impact of the FiT by an average of 200% relative to the sample average. These results emphasized that policy stability and clarity were critical to maximizing the impact of FiT schemes. Particularly when tariffs were low, consistent implementation became essential to maintain investor confidence and policy credibility. The study concluded that the effectiveness of FiTs was highly dependent on their structure, and that many previous studies likely underestimated the potential impact of FiTs by not accounting for these design and consistency factors.

The Italian Conto Energia feed-in tariff programme played a central role in promoting PV deployment in Italy, resulting in an increase of approximately 17.6 GW in domestic PV capacity between 2006 and 2018. While the programme was instrumental in accelerating the growth of solar energy, its economic efficiency has been subject to critical scrutiny. Throughout its implementation period, the capital-grant equivalent values of the feed-in tariffs consistently exceeded the prevailing market prices of PV systems, indicating an overcompensation

relative to actual investment costs. The total net subsidisation cost associated with Conto Energia is estimated to have exceeded EUR 60 billion, raising concerns about the fiscal burden imposed by the policy. Consequently, although effective in terms of capacity expansion, the programme is considered not to have been a cost-effective instrument for promoting PV deployment [150].

In practice, FiPs have gained popularity as they integrate renewable producers more directly into the electricity market, promoting competition and efficiency. However, their effectiveness depends on the structure of the premium and the level of exposure to market risks. For instance, Marques, Fuinhas, and Macedo [146] investigated the effects of FiTs, FiPs, and capacity payments on electricity generation by source in Spain, using monthly data from January 2010 to February 2017 and applying an Autoregressive Distributed Lag (ARDL) model. The findings show that FiTs did not have a statistically significant effect on solar PV electricity generation, thereby calling into question the effectiveness of these policy instruments within the Spanish context. This limited impact may be explained by the structural characteristics of FiTs themselves: as price-based mechanisms, FiTs guarantee producers a fixed payment regardless of actual output, weakening incentives to maximise generation or invest in efficiency improvements. In addition, the challenge of setting optimal tariff levels—due to information asymmetries between policymakers and producers [151]—can result in prices that are either too low to stimulate investment or too high to be cost-effective. As a result, while FiTs can drive capacity installation, they may not translate into proportional increases in electricity generation.

In the United Kingdom, the introduction of the FiT scheme in April 2010 played a pivotal role in accelerating solar PV deployment across all market segments [110]. A strong correlation was observed between the level of policy support and the rate of PV deployment. Annual installations rose sharply from just 30 MWp in 2010 to around 1 GWp in 2011, largely driven by growth in the domestic and ground-mounted segments, while non-domestic installations contributed to a lesser extent. However, policy revisions—such as the Fast-Track Review in August 2011 and Comprehensive Review Phase 1 in February 2012—reduced FiT rates across

all system sizes, leading to a significant drop in deployment in 2012. This decline continued into 2013 for the domestic and non-domestic sectors, following further tariff cuts introduced under Comprehensive Review Phase 2A. Despite these reductions, the ground-mount segment experienced a significant increase in capacity during 2013, reflecting a strategic shift toward larger-scale installations and a growing reliance on the Renewables Obligation (RO) scheme [110]. These trends underscore the importance of market-pull policies in stimulating deployment.

As the cost of solar technology has declined, many governments have moved away from FiTs in favour of market-based mechanisms [152] to control expenditure and enhance price competitiveness. Instruments such as renewable energy auctions [153] have become increasingly widespread. The International Renewable Energy Agency (IRENA) highlights a global trend toward auctions, motivated by policymakers' aims to procure renewable electricity at the lowest possible cost while also advancing objectives related to energy security and socio-economic development [153]. This shift has been particularly evident in European countries such as Germany, Italy, and Switzerland, where the advent of grid parity was accompanied by a reduction in FiTs. However, empirical evidence indicates that the withdrawal of FiTs without the introduction of stable alternative incentives resulted in a decline in investment in new capacity, raising concerns over the feasibility of achieving long-term low-carbon energy targets [152].

While these studies consistently show that FiTs play a significant role in driving early-stage solar PV deployment, it is important to recognise that their impact may change over time as markets mature. Most empirical evaluations focus on the early years of policy implementation, when solar markets were still emerging and cost barriers were high. In this context, the presence or removal of FiTs had an outsized influence on deployment levels. However, once markets are more established—due to falling technology costs, increased investor confidence, and accumulated industry experience—the removal of FiTs does not necessarily lead to sustained declines in deployment. For example, in the United Kingdom, cumulative solar PV capacity grew steadily to 13.35 GW by 2019—the year the FiT scheme was closed

to new applicants. Although growth slowed slightly in 2020, reaching 13.55 GW, capacity increased by more than 2.1 GW over the following three years, reaching 15.66 GW by 2023. This trajectory demonstrates that, although FiTs were instrumental in catalysing early deployment, the UK market continued to expand in their absence. This suggests that while FiTs are essential for market formation, their influence diminishes as solar markets become more competitive and self-sustaining.

2.3.2 Net Metering

Net metering policies enable PV system owners to export surplus electricity to the grid and receive credit on their electricity bills. Under net metering, the exported electricity is credited at the full retail rate, offering a strong economic incentive for investment in distributed PV systems. By contrast, net billing provides compensation at a lower rate—often close to the wholesale electricity price—meaning that system owners receive less value for electricity fed into the grid [154]. As a result, while both policies promote distributed generation, net metering is generally more favourable to consumers and has been more effective in driving residential solar adoption.

Numerous studies have evaluated the effectiveness of net metering in promoting solar PV deployment, particularly in the United States. Doris and Krasko [155] conducted a regression analysis to examine the relationship between various policy measures and newly installed PV capacity across all sectors in 2010. The variables analysed included interconnection standards, net metering standards, renewable portfolio standards (RPS), and solar carve-outs. Their findings indicated that all four policy measures were significantly linked to PV deployment, with each one-point improvement in a state’s net metering score corresponding to a 4.6% increase in PV installations.

A follow-up study by Steward et al. [156] supported these findings, showing that U.S. states implementing both net metering and RPS experienced more substantial growth in solar PV installations. However, not all findings were uniformly positive. Yin and Powers [157] used regression analysis covering the years 1990 to 2006 to assess whether a state’s renewable

portfolio standard (RPS), net metering standard, and interconnection standard influence the annual share of non-hydro renewable energy capacity—such as wind, solar, geothermal, and various forms of biomass. RPS requirements had a statistically significant effect on non-hydro renewable energy capacity, while net metering and interconnection standards were not statistically significant predictors in their model. This suggests that the influence of net metering may be more pronounced in certain regulatory or economic contexts, and that other complementary policies may be necessary to maximize its impact.

More recent studies have reinforced the importance of net metering, particularly for non-utility and residential solar installations. Gilbert Michaud [158] studied the effectiveness of state-level policies in encouraging non-utility solar PV installations in the US using hierarchical regression analysis and cross-sectional data from 2012 to 2013. By comparing policy impacts to other influencing factors—such as electricity prices, market deregulation, per capita income, and solar resource availability—the study found that net energy metering was the most influential policy in driving non-utility PV capacity growth.

In a panel analysis of rooftop solar adoption across 27 U.S. states from 2008 to 2018, Ros and Sai [159] found that net metering significantly increased consumer demand for rooftop PV, with adoption rising by at least twofold under supportive compensation schemes. Kim et al. [160] also identified a statistically significant positive effect of net metering on solar deployment in Colorado through a predictive model incorporating a range of social and environmental variables. However, the model showed that other factors—particularly socioeconomic characteristics and the average time required to obtain a permit (measured in business days from initial submission to local approval)—played a more dominant role in explaining deployment outcomes, suggesting that while net metering supports adoption, its overall predictive importance was relatively limited.

2.3.3 Renewable Portfolio Standards and Certificate-Based Mechanisms

Utility quota obligations, commonly known as Renewable Portfolio Standards (RPS), are regulatory mandates requiring utilities to procure a specified percentage of their electricity

from renewable sources. These standards are designed to stimulate investment in renewable energy technologies, lower greenhouse gas emissions, and diversify the energy mix to enhance long-term energy security. Compliance can be achieved either through direct generation or through the acquisition of Renewable Energy Certificates (RECs), which represent proof that one megawatt-hour (MWh) of electricity was generated from a renewable source.

Several U.S. states have enacted RPS policies as a market-pull mechanism to accelerate the transition toward renewable energy. While early studies on their effectiveness used cross-sectional methods or overlooked the diversity in policy design, recent work by Yin and Powers [157] introduces a refined measure of RPS stringency that accounts for design features such as percentage requirements, eligible technologies, and compliance flexibility. Their panel analysis from 1990 to 2006 finds that RPS policies have had a statistically significant and positive impact on in-state renewable energy capacity. Specifically, a one-percentage-point increase in the mandated renewable share led, on average, to a 0.56 percentage point rise in non-hydro renewable capacity. However, the study also notes that allowing unrestricted trading of RECs—referred to as “free trade”—can significantly weaken the effectiveness of an RPS by decoupling renewable development from the state implementing the policy. These findings suggest that not only the presence but also the design of RPS policies is critical to their success.

RECs, and their UK equivalent Renewable Obligation Certificates (ROCs), are tradable instruments awarded for every MWh of renewable electricity generated. These certificates create a market-based mechanism for compliance, allowing utilities and obligated parties to either generate renewable electricity themselves or purchase certificates from producers to meet regulatory targets. ROCs specifically operate under the UK’s Renewable Obligation (RO) scheme, wherein electricity suppliers must meet annual renewable energy targets. Both systems are intended to enhance flexibility and cost-effectiveness in meeting renewable obligations, and they play a key role in incentivizing the expansion of renewable generation capacity. In the United Kingdom, the ROC scheme was closed to new generating capacity in 2017 and subsequently replaced by Contracts for Difference (CfDs) as the primary sup-

port mechanism for large-scale renewable energy projects [161]. However, the ROC scheme continues to operate for existing accredited generators, who will receive support for up to 20 years from their accreditation date, with the final certificates expiring in 2037 [162]. CfDs guarantee renewable electricity generators a fixed 'strike price' for the electricity they produce, compensating them when market prices fall below this level and requiring repayment when prices exceed it. This mechanism reduces revenue uncertainty and investment risk, thereby encouraging continued investment in low-carbon technologies [163].

However, the effectiveness of these certificate-based systems varies depending on the market segment and complementary policy environment. As observed by Westacott and Candelise [110], different solar PV market segments have responded unequally to market-pull mechanisms such as FiTs and the RO. While the non-domestic rooftop segment saw limited uptake, the ground mounted solar sector expanded rapidly—transitioning from FiTs to ROCs as economies of scale made larger projects more financially viable.

In summary, RPS policies and certificate-based compliance mechanisms such as RECs and ROCs have proven effective tools in expanding renewable energy capacity, particularly when well designed. However, their impact is sensitive to policy design, market conditions, and the specific structure of certificate trading.

2.3.4 Other Policies

In addition to regulatory measures such as feed-in tariffs, net metering, and renewable portfolio standards, governments employ a range of financial incentives to support the development and deployment of solar energy projects. These include tax reductions, investment or production tax credits, public investment programmes, and direct financial incentives such as rebates and grants. Among the most common tools are tax incentives designed to lower the upfront cost burden for businesses and homeowners investing in solar technologies. Investment tax credits (ITCs) reduce the capital cost of solar projects by providing credits based on the total investment, while production tax credits (PTCs) are based on the amount of energy generated. These mechanisms reduce financial risk, enhance return on investment,

and improve the economic viability of solar energy systems.

Beyond tax-based instruments, governments also support solar energy through public investment schemes, including low interest loans, grants, capital subsidies, and rebates. These programmes are particularly valuable in addressing high initial capital costs, which often present a significant barrier to entry. Public investment loans typically offer favourable repayment terms, while grants and capital subsidies provide direct financial support for equipment procurement and installation. Empirical evidence shows that such instruments are highly effective: Crago and Chernyakhovskiy [164] found that each additional \$1 per watt in rebate funding is associated with a nearly 50% increase in annual residential PV installations in the U.S., underscoring the powerful role of rebates in stimulating solar adoption.

Cash incentives have also been identified as a particularly effective mechanism in the commercial solar market. Using panel data across 27 programmes in 16 U.S. states, Shrimali and Jenner [165] showed that cash incentives significantly increased commercial PV adoption, while interconnection standards were more influential in promoting residential deployment. The study also highlighted property tax incentives as a promising policy tool for supporting the uptake of commercial-scale systems.

In addition, competitive auctions or tenders are increasingly used as market-based mechanisms to drive down costs and promote efficient deployment of large-scale solar projects. In these schemes, developers bid to supply electricity at the lowest possible price, with successful bidders awarded contracts to construct and operate solar facilities. Auctions promote transparency and cost-effectiveness, encouraging technological innovation and market competitiveness.

As noted by Jacobs and Sovacool [145], while public spending on research and development (R&D) can help establish early markets and demonstration projects, tax and investment incentives remain essential to attracting private capital and creating commercially viable niches for solar PV. Collectively, these financial and market-based mechanisms play a critical role in accelerating the energy transition by lowering investment barriers, reducing levelized costs, and de-risking early-stage projects, particularly in markets where solar technologies

are still emerging.

2.3.5 Future Policies

As the deployment of solar PV systems continues to expand, future policy frameworks must evolve to address emerging technical and regulatory challenges, particularly those related to grid access, system integration, and infrastructure adequacy. One pressing issue is the growing length of grid connection queues, where a significant number of proposed PV projects await approval and interconnection. The increasing demand for grid access has resulted in delays that hinder the timely deployment of renewable energy infrastructure. These constraints are especially pronounced in regions with outdated or insufficient grid capacity, where utilities may lack the technical and administrative means to process and integrate a large volume of new connections efficiently.

Another challenge is the intermittent nature of solar power generation, which complicates grid balancing and reliability. As solar PV penetration increases, variability in electricity supply—caused by factors such as cloud cover and daily solar cycles—can create supply-demand imbalances. Without adequate grid flexibility or energy storage solutions, high levels of solar generation during peak sunlight hours may overwhelm local distribution networks, while periods of low generation may strain overall system adequacy. These dynamics introduce new complexities in maintaining grid stability, especially in energy systems that lack responsive demand side measures or dispatchable backup capacity.

In response to these challenges, future policies must prioritize streamlining interconnection procedures and enhancing grid flexibility. This includes investment in smart grid technologies, which enable dynamic monitoring and control of electricity flows, and facilitate the integration of distributed energy resources. Policies should also promote the deployment of energy storage systems, which can store surplus electricity generated during periods of high solar output and release it when generation drops. Furthermore, a forward looking regulatory environment should incentivise the development of flexible and modernized grid infrastructure, capable of supporting high levels of distributed generation and ensuring equitable access for all energy

producers.

By proactively addressing these technical and institutional challenges, policymakers can support the continued scaling of solar PV in a manner that is both reliable and resilient. Ensuring a smooth energy transition will require not only increasing solar capacity but also transforming the underlying systems and rules that govern electricity generation, distribution, and consumption.

2.4 Knowledge Gap

While significant progress has been made in tracking and modelling solar PV capacity, important gaps remain in the current literature and methodological landscape that limit the ability to support coordinated and data-driven energy planning. First, there is no global model that estimates national-level solar PV capacity based on geographical determinants. Such a model is useful not only for understanding how geographic factors influence deployment patterns across countries, but also for benchmarking national progress and informing national PV deployment targets. Existing global assessments typically focus on technical potential or large-scale utility installations, without systematically modelling actual installed capacity using comparable, location-specific predictors.

Second, within Europe, there is a lack of models that estimate regional PV capacity across the full set of European regions using consistent geographic inputs. A model of this kind would support the disaggregation of national capacity figures into regional estimates, which is critical for real-time generation monitoring and the development of grid management tools. It would also enhance the ability to forecast regional deployment trends for long-term infrastructure and grid expansion planning. Additionally, such models could help inform land-use policy, particularly in resolving conflicts between solar PV development and agricultural activity. Although regional disparities in solar deployment are well recognized, there remains a need for a consistent and spatially explicit framework to explain and predict these differences.

Third, there is no comprehensive national model that estimates PV capacity—across residential, commercial, and utility scale systems—at fine geographical scales for an entire

country. High-resolution capacity modelling is crucial for identifying under-deploying areas, informing grid infrastructure investment, and improving the accuracy of generation monitoring systems. While large-scale installations can sometimes be detected using satellite or computer vision methods, these approaches are computationally intensive, often omit small-scale systems, and provide only static snapshots of deployment. Their reliance on high-resolution imagery, which is not consistently updated, makes them poorly suited for real-time monitoring or forecasting of rapidly evolving PV deployment patterns.

Moreover, while geographical factors have been used to model specific PV market segments, there is currently no unified modelling framework that integrates both small- and large-scale systems using a consistent set of predictors. This fragmentation limits cross-scale comparability and undermines the development of cohesive energy strategies that span multiple administrative levels.

Another important limitation is that most existing studies are confined to countries or regions where detailed subnational capacity data are already available. Consequently, large parts of the world lack regionally disaggregated estimates, which are vital for informed energy planning, grid integration, and performance benchmarking. There is a need for scalable, generalisable models that can be applied even in data-scarce contexts, using open and globally available datasets.

To address these challenges, this study develops a consistent and scalable modelling framework to estimate solar PV capacity at multiple spatial levels: globally at the national level, regionally across NUTS 2 regions in Europe, and subregionally within NUTS 3 units in Great Britain. A key strength of the approach lies in its exclusive reliance on open-access, frequently updated, and globally available data sources, which enhances the transferability, reproducibility, and scalability of the methodology. This framework is designed to support energy planning and policy development in both data-rich and data-scarce environments, offering a practical solution for regionally resolved capacity estimation across diverse geographical contexts.

Chapter 3

Global Model

3.1 Abstract

Setting solar photovoltaic capacity targets and implementing supportive policies is a widespread strategy among nations aiming to achieve decarbonisation goals. However, policy implementation without a thorough understanding of the intricate relationship between social, economic, and land-use factors and solar photovoltaic deployment can lead to unintended consequences, including over- or underdeployment and failure to reach targets. To address this challenge, an investigation was conducted into the relationship between 36 factors and solar photovoltaic deployment across 143 countries from 2001 to 2020 using correlation analysis and principal component analysis. From these factors, five key variables were identified that collectively explain 79% of the year-to-year variation in photovoltaic capacity. Using these variables, a neural network model was constructed, enabling the estimation of capacity additions by country with an error of less than 10%. Additionally, a solar photovoltaic deployment index was developed, serving as a benchmark for comparing a country's actual historical photovoltaic deployment to similar nations. Furthermore, the model's utility in evaluating the impact of solar photovoltaic policies was explored. Through three distinct use cases—forecasting solar photovoltaic capacity additions, developing a solar photovoltaic deployment index, and assessing the impact of solar photovoltaic policies—the model emerges

as a potentially powerful tool for governments and policy makers to assess solar photovoltaic deployment effectively and formulate strategies to promote sustainable solar energy growth.

3.2 Introduction

Solar photovoltaic (PV) electricity is now, on average, cheaper than fossil fuel electricity [166], and one of the cheapest sources of power production [167]. Consequently, solar PV contributes substantially to the decarbonisation strategies of many countries. For example, China aims to increase the capacity of solar and wind to over 1200 GW by 2030 [168]. Japan targets 108 GW of installed solar capacity by 2030, equivalent to 15% of its total power generation [169]. EU countries Italy, Germany, and Spain are aiming for 52 GW [170], 215 GW [171], and 37 GW [172] of solar capacity by 2030. The United Kingdom's target is 7% of electricity from solar PV by 2030 [173], and South Africa's target is 8 GW of solar PV, which would account for 11% of total installed capacity [174].

These capacity targets are often backed up by a range of policies to support investment in solar power production. Historically and presently, these include feed-in tariffs, where system owners are paid to export power to the grid (e.g., China, Japan, and Vietnam [173]); net metering, which compensates system owners for surplus electricity fed into the grid [175] (e.g., Botswana, Zimbabwe, Saudi Arabia, and Belgium [173, 166]); utility quota obligations or renewable portfolio standards, which require a minimum percentage of generation to be provided by renewable energy of which a portion is solar PV [166] (e.g., Australia, Sweden, and the United Kingdom [173]); and tradeable renewable energy certificates (RECs), which are awarded per MWh and can be bought or sold separately from the electricity [144]. Other policy options include a reduction in tax associated with energy (e.g., Finland, France, Germany, Italy, Spain, and Japan [173]), direct investment or production tax credits to encourage businesses to develop and operate solar energy projects (e.g., Germany, Greece, Italy, Spain, and the United States [173]), and public investment loans, grants, capital subsidies or rebates, and auctions or tenders.

Implementing these kinds of policies without prior understanding of the complex interac-

tions of geographic factors (e.g., sociological, economic, land-use, climatic, and technological) involved in PV deployment can lead to unintended consequences such as overdeployment or failure to meet the targets. Spain's experience in 2008 exemplifies this, where rapid growth in solar PV deployment, driven by new feed-in tariffs, strained the electricity grid and necessitated sudden policy adjustments to curb costs [85]. This eventually led to market collapse in the following years. On the other hand, India, despite implementing various supportive measures such as feed-in tariffs, obligation certificates, and tax reductions, fell short of its ambitious 100 GW target by 2022, achieving only 54 GW [90, 91]. Furthermore, evaluating a country's progress in solar PV deployment based on its self-defined targets may not accurately reflect the actual deployment compared to realistic expectations, as it fails to account for the complex interplay between PV deployment and geographical factors.

The development of models that can explain the importance and interactions of the different adoption factors could be very valuable in the refinement of policies to support solar PV and its integration into the wider electricity system within a country. Such models could be used to forecast solar PV capacity in order to support network planning, and they could help with the early identification of over- or underdeployment by providing a benchmark in terms of what might be expected in terms of a comparison with countries further ahead in the deployment curve.

The aim of this study is to investigate the factors influencing country-level historical solar PV deployment, culminating in the construction of a comprehensive global model capable of estimating total PV capacity additions for any country. This model will serve multiple purposes: Firstly, as a forecasting tool for PV capacity, facilitating effective planning and generation monitoring. Secondly, as a benchmarking mechanism, allowing for comparisons with similar countries. Lastly, as an evaluative instrument for policy assessment, by comparing estimated outcomes against actual developments.

3.3 Literature Review

To construct a comprehensive global model, it is imperative to examine the factors influencing deployment and integrate them into our framework. To the best of our knowledge, there have not been any studies that investigate the relationship between geographic factors and solar PV deployment on a global scale. However, there are studies that investigate the relationship between these factors on household, subregional (i.e., census areas of approximately 1000–8000 people depending on country [109, 176]), regional (i.e., state/county/village), and country level.

On a household level, Jan, Ullah, and Ashfaq [101] identified key factors for explaining solar PV adoption in northwest Pakistan, which are income, cost of energy consumption, education level, information about the solar PV market, and source of awareness about solar PV systems. They explain 38% of the variation in adoption. Letchford, Kiran Lakkaraju, and Yevgeniy Vorobeychik [102] performed a sensitivity analysis using multiple methods to determine which features were important predictors of solar PV adoption in the San Diego region, US. Property size, whether the owner lived on the property, national unemployment rate, income, cost of electricity, and peer effect are all key factors that explain 33% of the variation in adoption. Aklin, Cheng, and Urpelainen [103] investigated solar PV adoption at the household level in rural India and found that households that are wealthy and have access to banking are more likely to use solar power.

Graziano and Gillingham [108] studied the influence of multiple factors on PV adoption in Connecticut, US, and showed that the influence of neighbours, the built environment (housing density and share of renters), and policy have a strong relationship with PV adoption compared to social, economic, and political factors. Alderete Peralta, Balta-Ozkan, and Longhurst [106] characterised the spatio-temporal adoption patterns of domestic solar PVs in Birmingham, UK, and found that income, electricity usage, and average household size are the best predictors of solar PV adoption.

Yu et al. [98] identified key social and economic factors correlating with solar deployment density in the US, which are solar radiation, population density, annual household income,

Gini index, and education level. Laura Williams and Mita Kerai [109] analysed factors that play a role in the deployment of PVs under a feed-in tariff scheme in England, UK, and found that electricity consumption, gas consumption, gas coverage, age of population, index of multiple deprivation and its various domains, dwelling stock by tenure and type, urban or rural classification, council tax band, and fuel poverty are all key in explaining solar PV deployment.

McEachern and Hanson [113] studied the adoption of solar PVs across 120 villages in Sri Lanka and found that solar PV adoption is driven by expectations of whether the government will connect the villages to the electricity grid, as well as tolerance for nonconformist behaviour. Aklin, Cheng, and Urpelainen [103] investigated factors that determine solar adoption at the village level in rural India and showed that remote, large, and poor villages with high levels of solar radiation adopt solar technology as a replacement for grid electricity. Mayer et al. [114] analysed the socioeconomic factors correlating with PV system adoption in 53 counties in the state of North Rhine Westphalia, Germany, and found that gross value added by agriculture was highly correlated with PV adoption with a Pearson correlation coefficient of +0.75, while unemployment rate and population density were moderately correlated with PV adoption at -0.61 and -0.64 , respectively. Liu et al. [116] investigated the correlation between social and economic factors and the installed capacity of solar PV in China and showed that GDP, final consumer expenditure, industrial added value, and solar energy generation and consumption are strongly correlated with PV capacity.

When considering quantitative models that can forecast PV capacity additions, there are very few models. The World Energy Model (WEM) initially forecasts total capacity additions, irrespective of technology, driven by demand. Subsequently, the share of solar PV capacity additions is determined according to the regional value-adjusted levelised cost of electricity [177]. However, this method of estimating the required capacity and subsequently deriving the share of solar PV capacity introduces compounding errors. Historically, the International Energy Agency, the US Energy Information Agency, Bloomberg New Energy Finance, Photon, and Greenpeace all underestimated PV capacity additions [178].

The International Renewable Energy Agency (IRENA) projects global solar PV capacity additions based on current and planned policies and targets of countries, as well as the trajectory of the global energy system aimed at limiting the rise in global temperatures to well below 2 degrees Celsius above preindustrial levels [179]. However, these projections are susceptible to errors due to several factors. Firstly, the capacity targets set by countries are not always met, leading to discrepancies between projected and actual outcomes. Secondly, the emphasis on global temperature objectives does not fully account for the diverse geographic factors within each country, which can significantly influence the deployment of solar PV capacity.

In academia, Yu et al. [98] developed a machine learning model that uses socioeconomic and environmental factors to accurately predict solar PV deployment density in US sub-regions. The model is a two-stage model that uses a random forest classifier to determine whether any solar PV systems exist in a census area and a random forest regressor to estimate the solar deployment density. The model achieved a cross-validation R^2 of 0.72, but it uses a large number of US-specific input features (>90), which makes it difficult to replicate in another country. In addition, it only takes into account residential PV systems.

Liu et al. [116] built a bidirectional long short-term memory neural network model to forecast China's solar PV installed capacity and achieved a mean absolute percentage error (MAPE) of 6%. A mean impact value analysis was performed to determine the contribution of each factor in the model. Solar power generation, solar power consumption, gross domestic product, final consumer expenditure, and industrial added value contributed 26%, 27%, 17%, 15%, and 14%, respectively. This model uses a small number of input features, but it relies on solar generation and consumption data, which are not available for most countries.

Remote sensing-based methods have emerged as a promising solution for acquiring information on PV installations. These techniques use overhead imagery and deep neural networks to detect and map solar PV capacity using computer vision. For instance, Ravishankar et al. [180] devised a deep learning framework to estimate the global capacity of solar farms from high-resolution satellite imagery, achieving an average error rate of 4.5% when validated

against publicly available data; while this method effectively detects large-scale solar installations, it can be computationally expensive. Detecting small-scale solar PV installations is more complicated as it necessitates high-resolution imagery to maintain model performance [99, 181]. Recent advancements, such as the development of electric-dipole gated phototransistors, offer high-performance imaging capabilities with reduced power consumption, promising improved machine vision imaging models in the future [182].

These studies investigated factors associated with PV adoption, which were used as a guide for selecting factors in the present study. The focus extends to total national capacity additions, regardless of type (residential, commercial, utility scale). Existing global models estimate capacity based on national targets, leading to inaccuracies. Furthermore, national or subregional capacity models often rely on data unavailable in many countries, and while remote sensing-based vision models offer a promising solution for acquiring information on PV installations, their reliance on high-resolution imagery can be impractical in regions where such data are scarce or expensive to obtain.

To overcome these difficulties and develop a common framework for analysis of national PV capacity across many countries, an attempt is made to build a generic model that relies on open source global databases. Fortunately, global databases of historical data are available. Weather data are available as a record of meteorological variables and include irradiance [183], the key determinant of solar PV production. The International Energy Agency (IEA) and Euro-Mediterranean Center on Climate Change (CMCC) provide records of averaged weather parameters specific to each country, such as temperature, daylight hours, snowfall, cloud coverage, and precipitation [184]. The World Bank documents country-level demographics such as population, average age, level of education, employment, and national economics such as gross domestic product and gross national income [185]. The World Bank also documents land use such as urban, rural, and agricultural. The International Renewable Energy Agency (IRENA) reports country-level solar PV capacity additions on an annual basis [186]. The Energy Information Administration (EIA) tracks annual electricity consumption and generation by source such as nuclear, fossil fuels, and renewables [187], and the Renewable

Energy Policy Network for the 21st Century (REN21) documents national and subnational solar PV policies [166].

3.4 Methodology

3.4.1 Determining Key Features Associated with Global Solar PV Capacity Additions

To build a global model, potential geographic features with global coverage needed to be determined. Table 3.1 shows the investigation conducted into the relationship between 36 climatic, social, economic, and land-use features and solar PV capacity additions. The Pearson's correlation coefficient was calculated for each factor in relation to solar PV capacity additions. The coefficient of determination (R^2) was determined by fitting linear regression models between each feature and solar PV capacity additions. To evaluate similarity among features, a variable clustering analysis was performed in JMP Pro using the VARCLUS algorithm developed by SAS [188, 189]. The algorithm starts with all variables in a single cluster and iteratively partitions them based on shared variance. Within each cluster, principal components are computed, and if the second eigenvalue exceeds 1, the cluster with the largest second eigenvalue is selected for splitting. Splitting is guided by an orthoblique rotation, where variables are assigned to one of two new clusters depending on whether they show a stronger squared correlation with the first or second rotated component. After each split, variables are re-evaluated and reassigned to the cluster where they have the highest squared correlation with the first principal component. This process continues until no further improvements can be made. For each final cluster, the proportion of variance explained by the first principal component, as well as by individual variables, was calculated. The resulting structure provided interpretable groupings of similar features to inform further analysis. The data in Table 3.1 cover 143 countries around the world, span the years 2001 to 2020, and have a temporal resolution of one year.

Table 3.1: Features considered for modelling solar photovoltaic capacity additions. The definition, category, and availability of the data are shown. The correlation and coefficient of determination (R^2) with solar photovoltaic capacity additions are calculated. Principal component analysis (PCA) is performed, and similar features are clustered together. Finally, the literature that used the same or similar features is linked.

Feature	Definition	Category	Availability	Pearson Correlation/ R^2	PCA Cluster/ R^2	Literature	Source
Temperature ($^{\circ}\text{C}$)	Annual mean temperature at 2 m.	Climate	Global / complete	-0.1/0.01	2/0.85	[98]	[184]
Average theoretical potential GHI ($\text{kWh}/\text{m}^2/\text{day}$)	Annual mean theoretical global horizontal irradiance.	Climate	Global / complete	-0.08/0.007	2/0.77	[98, 103]	[183]
Snowfall (mm/h)	Annual mean snowfall.	Climate	Global / complete	0.04/0.002	2/0.61		[184]
Daylight hours (minutes/day)	Annual mean daylight hours.	Climate	Global / complete	0.04/0.002	2/0.82		[184]
Precipitation (mm/h)	Annual mean precipitation.	Climate	Global / complete	-0.02/0.0006	5/0.87		[184]
Cloud coverage (%)	Annual mean cloud coverage.	Climate	Global / complete	0.007/0.00006	5/0.87		[184]

Feature	Definition	Category	Availability	Pearson Correlation/ R^2	PCA Cluster/ R^2	Literature	Source
Tertiary education	The number of population enrolled in tertiary education. This is calculated by multiplying the population by the tertiary gross enrollment ratio.	Social	Global / incomplete	0.8/0.6	4/0.87	[101, 106, 98] ¹	[190]
Labour force	The number of people aged 15 or older who supply labour for the production of goods and services.	Social	Global / complete	0.5/0.3	4/0.96	[98] ²	[191, 192]
Population	Count of people in a country.	Social	Global / complete	0.5/0.2	4/0.96	[106, 98, 103] ³ , [116]	[193]
Primary education	The number of population enrolled in primary education. This is calculated by multiplying the population by the primary gross enrollment ratio.	Social	Global / incomplete	0.5/0.2	4/0.96	[101, 106, 98] ¹	[190]

Feature	Definition	Category	Availability	Pearson Correlation/ R^2	PCA Cluster/ R^2	Literature	Source
Total unemployment	The total number of unemployed labour force that is without work but available and seeking employment. Calculated by multiplying the labour force with unemployment percentage.	Social	Global / complete	0.5/0.2	4/0.95	[102, 108, 114] ⁴	[192]
Secondary education	The number of population enrolled in secondary education. This is calculated by multiplying the population by the secondary gross enrollment ratio.	Social	Global / incomplete	0.4/0.1	4/0.99	[101, 106, 98] ¹	[190]
KOFGI	The KOF Globalisation Index measures the economic, social, and political dimensions of globalisation.	Social and economic	Global / complete	0.1 /0.01	6/0.85		[194, 195]
Duration of compulsory education (years)	The number of years that children are legally obliged to attend school.	Social	Global / complete	0.03/0.001	7/0.58	[98] ⁵	[190]

Feature	Definition	Category	Availability	Pearson Correlation/ R^2	PCA Cluster/ R^2	Literature	Source
Last year's PV cumulative capacity (MW)	The solar photovoltaic cumulative capacity of the previous year.	Economic	Global / complete	0.8/0.7	1/0.43		[186]
Electricity net generation (billion kWh)	Annual total electricity generation.	Economic	Global / complete	0.8/0.6	1/0.96	[116]	[187]
Electricity net consumption (billion kWh)	Annual total electricity consumption.	Economic	Global / complete	0.7/0.6	1/0.94	[116, 102, 106, 109]	[187]
Fossil fuels electricity net generation (billion kWh)	Annual fossil fuel electricity generated by a country.	Economic	Global / complete	0.7/0.6	1/0.90		[187]
Agriculture, forestry, and fishing, value added (current USD)	Value added by agriculture, forestry, and fishing sectors.	Economic	Global / complete	0.7/0.5	4/0.88	[114]	[196, 197]
Manufacturing, value added (current USD)	Value added by the manufacturing sector.	Economic	Global / complete	0.7/0.5	1/0.95		[198, 197]

Feature	Definition	Category	Availability	Pearson Correlation/ R^2	PCA Cluster/ R^2	Literature	Source
Industry (including construction), value added (current USD)	Value added by the industrial sector.	Economic	Global / complete	0.7/0.5	1/0.97	[116]	[199, 197]
GDP (current USD)	Gross domestic product (GDP) measures the gross value added by production of goods and services in a country on a yearly basis.	Economic	Global / complete	0.6/0.3	1/0.93	[116], [102] ⁶ , [114] ⁷	[200, 201]
GNI (current USD)	Gross national income (GNI) is defined as GDP plus net income from abroad.	Economic	Global / complete	0.5/0.3	1/0.92		[202, 203]
Nuclear electricity net generation (billion kWh)	Annual nuclear electricity generated by a country.	Economic	Global / complete	0.3/0.1	1/0.31		[187]
Ease of doing business rank	Ease of doing business ranks economies of countries from best to worst based on how the regulatory environment is conducive to business operation.	Economic	Global / incomplete ⁸	-0.2/0.05	2/0.57		[204]

Feature	Definition	Category	Availability	Pearson Correlation/ R^2	PCA Cluster/ R^2	Literature	Source
Research and development expenditure (% of GDP)	Percentage of GDP spent on research and development.	Economic	Global / incomplete	0.2/0.03	6/0.57		[190]
Public investments in solar energy (2019 million USD)	Annual public investment in solar energy.	Economic	Global / incomplete	0.1 /0.01	6/0.23		[205]
Solar PV module cost (2019 USD per W)	Global average price of solar photovoltaic modules.	Economic	Global / complete	-0.09/0.008	7/0.58		[186, 206]
Access to electricity (% of population)	Percentage of population with access to electricity.	Economic	Global / complete	0.07/0.005	6/0.56	[101, 98, 109, 113, 103] ⁹	[207]
Investment in energy with private participation (current USD)	Investment in energy generation, transmission, and distribution projects with private participation.	Economic	Global / incomplete	0.03/0.001	3/0.22		[208]
Gini index (World Bank estimate)	Gini index is a measure of income inequality within a country.	Economic	Global / incomplete	-0.02/0.0004	2/0.57	[98]	[209]

Feature	Definition	Category	Availability	Pearson Correlation/ R^2	PCA Cluster/ R^2	Literature	Source
Urban land area (km ²)	Urban land area which is based on population counts, settlement points, and presence of nighttime lights.	Land use	Global / complete	0.4/0.1	1/0.76	[109] ¹⁰	[210]
Agricultural land (km ²)	Total area of land used for agriculture within a country.	Land use	Global / complete	0.3/0.1	3/0.70		[211]
Land area (km ²)	Total land area of a country excluding water bodies.	Land use	Global / complete	0.2/0.06	3/0.97	[106, 98, 103] ¹¹	[212]
Rural land area (km ²)	Rural land area which is based on population counts, settlement points, and presence of nighttime lights.	Land use	Global / complete	0.2/0.06	3/0.97	[109] ¹⁰	[210]
Forest area (km ²)	Total forest land area within a country.	Land use	Global / complete	0.1/0.02	3/0.84		[213]

¹ Similar but not identical education measures. ² Used employment rate rather than number of employed people. ³ Used population density rather than population count. ⁴ Used unemployment rate instead of number of unemployed people. ⁵ Used the number of years a person spent pursuing education. ⁶ Used change in GDP. ⁷ Used GDP per person. ⁸ Only available for 2019. ⁹ These studies use features that directly or indirectly measure access to electricity. ¹⁰ Used urban and rural classification instead of land area. ¹¹ Used population density rather than land area.

PCA results in Table 3.1 show that the first cluster's members are mostly economic features. It explained 22.4% of the variation in the feature data and was highly correlated with PV capacity additions. When fitting the cluster features in a linear model, it explained 76.9% of the variation in PV capacity, as shown in Appendix 3.7.2. The most important features in the first cluster were last year's PV cumulative capacity, electricity net generation, and consumption.

The second cluster comprised mostly climate features and explained 11.6% of the variation in the feature data. When the cluster features were fitted into a linear model, they accounted for 21.1% of the variation in PV capacity, as shown in Appendix 3.7.2. However, the number of data points for this cluster was less than 1% of the dataset, a limitation attributed to the methodology of fitting multiple features simultaneously. Specifically, when fitting more than one feature, rows with missing data points for any feature were excluded from the analysis. As shown in Table 3.1, features from this cluster were not correlated with PV capacity additions and did not explain the variation in PV capacity.

The third cluster consisted mainly of land area features and explained 10.3% of the variation in the feature data. When the cluster features were fitted into a linear model, they accounted for 22.2% of the variation in PV capacity, as shown in Appendix 3.7.2. The most significant feature in this cluster was agricultural land area.

The fourth cluster consisted of social features and explained 18.2% of the variation in the feature data. When the cluster features were fitted into a linear model, they accounted for 46.3% of the variation in PV capacity, as shown in Appendix 3.7.2. The most significant feature in this cluster was tertiary education. The remaining clusters explained less than 10% of the variation in the feature data.

Economic factors played the largest role in explaining the variation in solar PV deployment, followed by social factors. Land use played an important role but was less significant compared to economic and social factors. When it came to explaining the variation in the feature data, economic and social factors contributed equally, while land-use factors contributed about half as much as social or economic factors.

The Pearson's correlation coefficient and the coefficient of determination (R^2) in Table 3.1 show that climate features did not play a significant role in the additions of solar PV capacity on a global scale. This was not the case on smaller scales such as regional, subregional, and household levels, where solar irradiance was a key factor [98, 103].

Tertiary education was highly correlated with solar PV capacity additions (0.8) and explained 60% of the variation on a global scale, which was also the case on subregional [98] and household levels [101]. It was the most significant social feature on a global scale. This may have been because it acted as an indicator of the economy and population count in addition to education. Population count, primary and secondary education, labour force, and total unemployment were moderately correlated with PV capacity additions and explained between 10% and 30% of the variation.

The previous year's cumulative PV capacity, economic value added by agriculture, forestry, fishing, manufacturing, industry, electricity net consumption and generation, and fossil fuel electricity net generation were highly correlated with added PV capacity and explained a high percentage of the variation (between 50% and 80%). They played the largest role in explaining solar PV deployment on a global scale. Gross domestic product (GDP) and gross national income (GNI) were moderately correlated with added PV capacity. This was not the case on the country or regional level, where GDP was highly correlated with solar PV deployment on a country level [116] and weakly correlated on a regional level [114].

Solar PV module price was very weakly correlated with PV capacity additions. This may have been because some countries implemented policies that supported the adoption of solar PVs early on. The Gini index, which is a measure of income inequality, was very weakly correlated with solar PV capacity on a global scale, although it was strongly correlated at the subregional level [98]. Access to electricity was very weakly correlated with PV capacity on a global scale but was significant on smaller scales such as regional and subregional [109, 113, 103]. Using available but incomplete data, investment in energy and public investment in solar energy were weakly correlated with PV capacity additions.

Urban land area was moderately correlated with added PV capacity and explained 10% of

the variation, while rural land area was weakly correlated and explained 6% of the variation. This suggests that urban land area was a more important factor on a global scale compared to rural land area, but this was not the case on a subregional level where rural areas were associated with more PV installations [109]. This was explained by the high correlation between urban land area and GDP (0.90). Agricultural land area had a higher correlation and explained more of the variation in PV additions compared to total land area. This was probably because agricultural land was suitable for large-scale solar farms. Forest land area was not associated with PV installations on a global scale.

3.4.2 Feature Screening and Selection

The aim was to develop a globally applicable model, which required reducing the number of features due to variations in data availability across different countries. Moreover, some features displayed high correlation, which, when added to the model, increased complexity without substantially enhancing performance or explanatory power. To reduce the number of features, variables were grouped into clusters using the VARCLUS algorithm [188, 189]. Clusters explaining less than 5% of the variability in PV capacity or containing fewer than 500 data points were excluded (Appendix 3.7.2). The retained clusters were clusters one, three, and four, which explained 22.4%, 10.3%, and 18.2% of feature variability, respectively. Features were drawn from these clusters in proportion to their explanatory contribution: 40% from clusters one and four each, and 20% from cluster three.

Within each cluster, features were prioritised using a composite score based on three criteria: the Pearson correlation coefficient (Corr_i) between feature i and PV capacity additions; the number of valid observations for feature i (N_i); and the number of studies referencing feature i (L_i). Each criterion was normalised within the feature's cluster to account for differences in scale. The normalised correlation was computed as shown in Equation (3.1), the normalised data availability as shown in Equation (3.2), and the normalised literature

frequency as shown in Equation (3.3):

$$\text{NormCorr}_i = \frac{|\text{Corr}_i|}{\max_{k \in C_i} |\text{Corr}_k|} \quad (3.1)$$

$$\text{NormAvail}_i = \frac{N_i}{\max_{k \in C_i} N_k} \quad (3.2)$$

$$\text{NormLit}_i = \frac{L_i}{\max_{k \in C_i} L_k} \quad (3.3)$$

Here, C_i denotes the set of features in the same cluster as feature i .

The composite score was then calculated as shown in Equation (3.4) giving greatest weight to correlation while still accounting for coverage and prior usage.

$$\text{Score}_i = 0.6 \cdot \text{NormCorr}_i + 0.25 \cdot \text{NormAvail}_i + 0.15 \cdot \text{NormLit}_i, \quad (3.4)$$

Features in the same cluster were ranked from highest to lowest composite score. Ranking was restricted to features whose (Cluster, Category) pair was in Equation (3.5):

$$A = \{(4, \text{social}), (1, \text{economic}), (3, \text{land use})\}. \quad (3.5)$$

The optimal number of features was determined by incrementally fitting linear models and evaluating R^2 and RMSE. Features were introduced according to the cluster weights (40% from clusters 1 and 4 each, and 20% from cluster 3), using a repeating five-step cycle: The repeating cycle of clusters was:

$$1, 4, 3, 1, 4, 1, 4, 3, \dots$$

Within each cluster position, features were added in descending order of their composite score, ensuring that the highest-scoring features entered first.

Alternative feature selection approaches, including stepwise regression with both AICc and BIC criteria, were also applied. Detailed results of the stepwise regression analysis are

provided in Appendix 3.7.5.

3.4.3 Model Building

The models considered were a multiple linear regression (MLR) model, a second-order polynomial regression (PR) model, a neural network model (NN), and finally, a combined model, which is a neural network model that took the second-order polynomial features as input. For all models, the dataset was partitioned into training (80%) and test (20%) sets using the `train_test_split` function from Scikit-learn [214]. To ensure comparable distributions of the target variable (annual PV capacity additions) in both sets, stratified sampling was applied. The target variable was first discretised into six bin labels using `numpy.digitize` with bin edges defined by `numpy.linspace` between 0 MW and 1,843 MW. The six labels arose from `digitize` behaviour, which assigns separate labels for values below the first bin edge and above the last bin edge. These labels were then used as the stratification variable in the split procedure. A fixed random seed (`random_state=1`) was set to ensure reproducibility.

For the linear models, Scikit-learn, a Python package encompassing various advanced machine learning algorithms for medium-scale supervised and unsupervised tasks, was used [214]. Two model types were considered: multiple linear regression (MLR) and second-order polynomial regression (PR). The MLR model can be written as:

$$y = \beta_0 + \sum_{i=1}^p \beta_i x_i + \varepsilon \quad (3.6)$$

where y is the target variable, β_0 is the intercept, β_i are the model coefficients, x_i are the predictor variables for $i = 1, \dots, p$, and ε is the error term capturing the discrepancy between the observed and predicted values.

The PR model extended this formulation by including all quadratic and pairwise interaction terms:

$$y = \beta_0 + \sum_{i=1}^p \beta_i x_i + \sum_{i=1}^p \beta_{ii} x_i^2 + \sum_{i < j}^p \beta_{ij} x_i x_j + \varepsilon \quad (3.7)$$

where y is the target variable, β_0 is the intercept, β_i are the coefficients for the linear terms,

β_{ii} are the coefficients for the squared terms, β_{ij} are the coefficients for the interaction terms between predictors x_i and x_j , and ε is the error term. The parameters of both MLR and PR models were estimated using Ordinary Least Squares (OLS).

In the case of neural network models, JMP Pro 17 software was employed [215]. Multilayer perceptron (MLP) neural networks, which are fully connected feed-forward artificial neural networks, were used [216]. The neural network model contained a single hidden layer with five nodes and a hyperbolic tangent (Tanh) activation function. For the combined model, two hidden layers with five nodes each were used. These architectures were selected empirically by testing a range of configurations with one or two hidden layers and up to five nodes per layer, and then choosing the structures that provided the best balance between model performance and complexity.

Polynomial features were passed to the neural network models. This was motivated by the fact that interaction terms involving the previous year's PV capacity help distinguish between countries with no installed capacity and those with capacity. For countries with zero previous capacity, these interaction terms reduce all related polynomial inputs to zero, effectively encoding the absence of capacity in a way that improves the model's ability to separate such cases from those with existing capacity.

The JMP Neural platform applies a penalty on model parameters to prevent overfitting, with the penalty values tuned using a validation set. Model parameters were optimised using K -fold cross-validation ($K = 5$), which divided the data into five subsets (folds). Each fold served as the validation set once, while the remaining folds formed the training set, resulting in five fitted models. For linear models, the reported validation R^2 corresponds to the average across all folds, whereas for neural networks, it corresponds to the fold achieving the best validation performance.

Since this is the first attempt at creating a global model, there are no similar models available for benchmarking. The challenge is further complicated by instances where countries experienced years with either no or minimal capacity additions, resulting in actual added capacity being zero or near-zero in these cases. Mean absolute error (MAE), mean squared

error (MSE), and root mean squared error (RMSE) were used to compare the models. Additionally, new error metrics such as the global error, country error, and yearly error were defined. The global error, calculated according to Equation (3.8), serves to evaluate the overall performance of the model and facilitate comparison between the different models. Country error, determined using Equation (3.9), allows for comparison of errors between countries. The yearly error assesses the error for each year, based on Equation (3.10).

$$\text{Global Error} = \frac{\text{MAE}_{\text{Global}}}{\text{mean cumulative capacity}_{\text{Global}}} \quad (3.8)$$

$$\text{Country Error} = \frac{\text{MAE}_{\text{per country}}}{\text{mean cumulative capacity}_{\text{per country}}} \quad (3.9)$$

$$\text{Yearly Error} = \frac{\text{MAE}_{\text{per year}}}{\text{mean cumulative capacity}_{\text{per year}}} \quad (3.10)$$

3.4.4 Model Application

After selecting the best model, it is used as a benchmark against which solar PV deployment in different countries is evaluated. A solar PV deployment index (SPVDI) is developed to assess solar PV deployment in a country relative to other countries with similar social, economic, and land-use factors. The SPVDI is calculated based on Equation (3.11), where t_{start} is the initial year and t_{end} is the final year for which the analysis is conducted. The index sums the difference between actual and predicted capacity for country i over the specified time range. The result is the quantity of capacity additions with a corresponding sign. A positive sign indicates that the country has more capacity than expected, while a negative sign indicates less capacity than expected. The SPVDI enables the comparison of countries' performance across multiple years and time ranges. Additionally, it serves as a tool to rank countries based on their performance in terms of PV deployment.

$$\text{SPVDI}_i = \sum_{j=t_{\text{start}}}^{t_{\text{end}}} (\text{Actual Capacity Additions}_{i,j} - \text{Predicted Capacity Additions}_{i,j}) \quad (3.11)$$

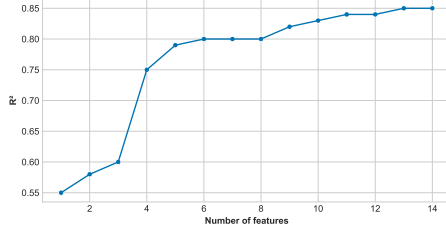
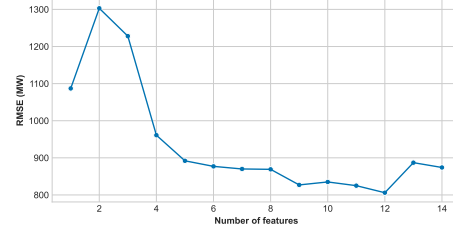
Another application is the use of the model to assess the effectiveness of implementing or removing a policy. This evaluation of policy interventions involves comparing actual capacity additions to the estimated capacity additions. When actual additions surpass expectations following policy implementation, it indicates success, signifying that the policy effectively increased capacity beyond initial projections. Conversely, if actual additions fall short of expectations, it suggests a policy failure, as it did not achieve the anticipated increase in capacity. To illustrate this use case, a specific set of countries is selected, and their policy interventions are evaluated.

3.5 Results and Discussion

Based on the analysis of 36 geographic factors in Table 3.1, economic factors emerge as the largest contributors to solar PV deployment, with the previous year's PV capacity playing the most important role. Social factors, especially education, also contribute significantly. Land-use factors have a smaller yet measurable effect, while climatic factors are not significant.

Five key features are selected—last year's cumulative PV capacity, population, agricultural land area, tertiary education, and electricity net consumption—which collectively account for 79% of the variation in PV capacity, to be fed into the models. Illustrated in Figure 3.1 is the relationship between the number of features, explained variation in PV capacity additions, and root mean squared error (RMSE) of the linear model used for feature selection. The analysis reveals marginal gains beyond these five features, with less than a 6% increase in explanatory power and no significant decrease in RMSE. Appendix 3.7.3 shows summary statistics for the selected features, and Appendix 3.7.4 shows the Pearson correlations between the selected features.

A comparison of the results obtained from the developed models is presented in Table 3.2.

(a) Number of features versus R^2 

(b) Number of features versus RMSE

Figure 3.1: Relationship between number of features and model performance metrics. (a) Number of features versus coefficient of determination, and (b) number of features versus root mean squared error after fitting a linear model. Features are numbered as follows: (1) electricity net consumption, (2) tertiary education, (3) agricultural land area, (4) previous year's PV cumulative capacity, (5) population, (6) electricity net generation, (7) total unemployment, (8) land area, (9) GDP, (10) primary education, (11) value added by industry (including construction), (12) labour force, (13) rural land area, (14) value added by manufacturing.

Detailed equations describing these models can be found in Appendices 3.7.7 and 3.7.8. The best results are achieved by the combined model, which has a global error value of 9.8%. The neural network model comes in second place with a global error of 10.3%. The polynomial PR model has a global error of 18.9% which is almost double the error of the more complex NN and combined models. The MLR model has the highest global error of 36.1%, and its validation R^2 score is -1.57 , which shows that its predictions are worse than a constant function that predicts the mean of PV capacity additions, deeming it unsuitable to model capacity additions. Considering that the error in measuring national PV capacity is at least 5% [49], the combined model's prediction error of 9.8% provides a reliable estimate of the actual capacity.

In Table 3.2, the polynomial regression (PR) model reduces the mean absolute error (MAE) relative to MLR, with values decreasing from 330 to 173. However, the root mean squared error (RMSE) increases from 675 to 1393. This increase is almost entirely attributable to a single outlier (China, 2018), which contributes 93.7% of the total squared error. When this case is excluded, PR's RMSE decreases substantially from 1393 to 349 and MAE from 173 to 103. Thus, although PR improves typical accuracy as reflected by MAE, its RMSE

Table 3.2: Comparison between the results of the different models. MLR is the Multiple Linear Regression model, PR is the Polynomial Regression model, NN is the Neural Network model, and CM is the Combined Model. The training and test sets used in each model were identical.

Model	Train R^2	Val R^2	Test R^2	Train MAE (MW)	Train RMSE (MW)	Train Global Error (%)	Test MAE (MW)	Test MSE (MW ²)	Test RMSE (MW)	Test Global Error (%)
MLR	0.72	-1.57	0.92	342	946	44.8	330	4.56×10^5	675	36.1
PR	0.96	0.63	0.68	96	343	12.5	173	1.94×10^6	1393	18.9
NN	0.92	0.95	0.98	91	445	11.9	95	1.49×10^5	386	10.3
CM	0.97	0.94	0.98	70	292	9.2	90	1.28×10^5	358	9.8

is dominated by the outlier, highlighting the disproportionate influence of squared errors on large deviations.

Appendix 3.7.6 shows the importance of each feature in the combined model, the correlation between the features and the PV capacity additions, and how much variation each feature explains in the PV capacity additions. The interaction between the previous year's PV cumulative capacity and the other features increased the correlation and explained more of the variation in the PV capacity additions compared to single features, which explains why these terms are the most important in the model. This also explains why the MLR model had the lowest training R^2 (0.72) compared to the other models, which had a much higher training R^2 (>0.96).

Figure 3.2 shows the linear and logarithmic actual versus predicted capacity additions for the combined model. The combined model is highly accurate for capacity additions greater than 1 GW (global error = 4.5%), has a medium accuracy for capacities between 1 MW and 1 GW (global error = 19.1%), and has a low accuracy for capacities below 1 MW (global error = 177%).

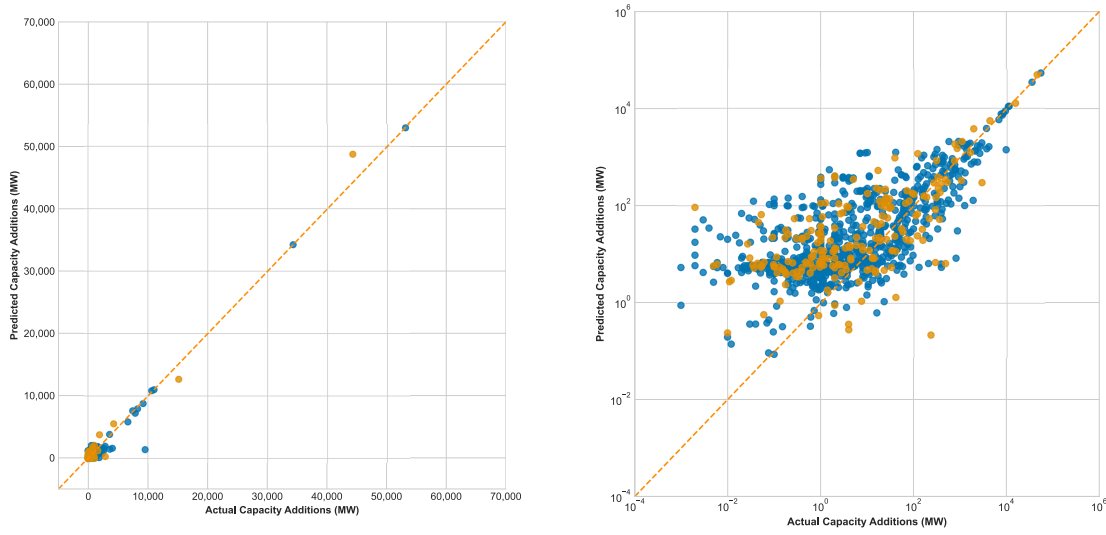


Figure 3.2: Linear and logarithmic actual vs. predicted solar photovoltaic capacity additions for the combined model. Blue points represent the training set and orange points represent the test set.

Figure 3.3 shows the average cumulative PV capacity per year and the yearly error for the combined model. The error is highest in earlier years when the average capacity was low, but it drops significantly once the capacity starts to increase. This suggests that the model works well once a capacity threshold has been surpassed. This threshold is probably related to the state of the solar PV market.

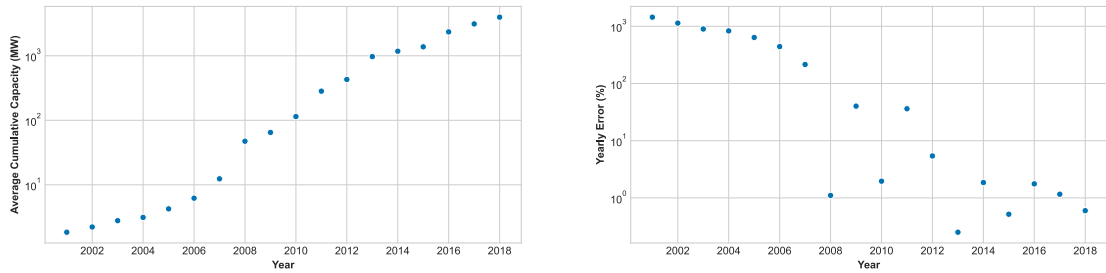


Figure 3.3: Average cumulative solar photovoltaic capacity and error per year for the combined model. The yearly error was calculated from the entire dataset by dividing the mean absolute error per year by the average cumulative capacity that year.

Figure 3.4 shows the country error for the combined model. The largest errors are for countries with low to no PV capacity additions, as shown in Figure 3.5. These countries increase the model's global error to about 10%, but as shown in Figure 3.4, countries with

high capacities have errors that are significantly smaller than 10%. For example, Germany, the United Kingdom, China, Australia, Greece, and Japan have an error of less than 1%. Portugal, India, Austria, Denmark, Bulgaria, Belgium, and Switzerland have an error of less than 5%. The model works well as a forecasting tool in countries with a mature solar PV market but performs less well in countries with emerging markets.

While the model may exhibit significant errors in certain countries, its appropriateness depends on the specific application at hand. For instance, when forecasting solar PV capacity for monitoring PV generation, an error of around 10% or less is typically acceptable. In the context of ranking countries' PV deployment based on geographic factors, the observed error describes deviations from expected capacity additions compared to similar nations, offering valuable comparative insight. Lastly, when evaluating policy effectiveness using the model, discrepancies between actual and projected capacities can serve as indicators of policy impact, given the model's inherent exclusion of policy inputs.

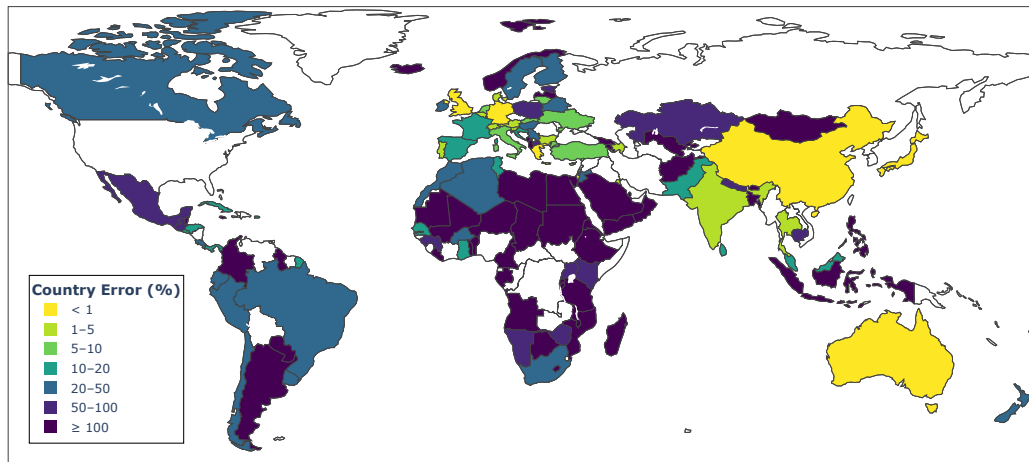


Figure 3.4: Country error of the combined model, determined by dividing the mean absolute error per country by the respective mean cumulative photovoltaic capacity. Countries without available data are represented in white.

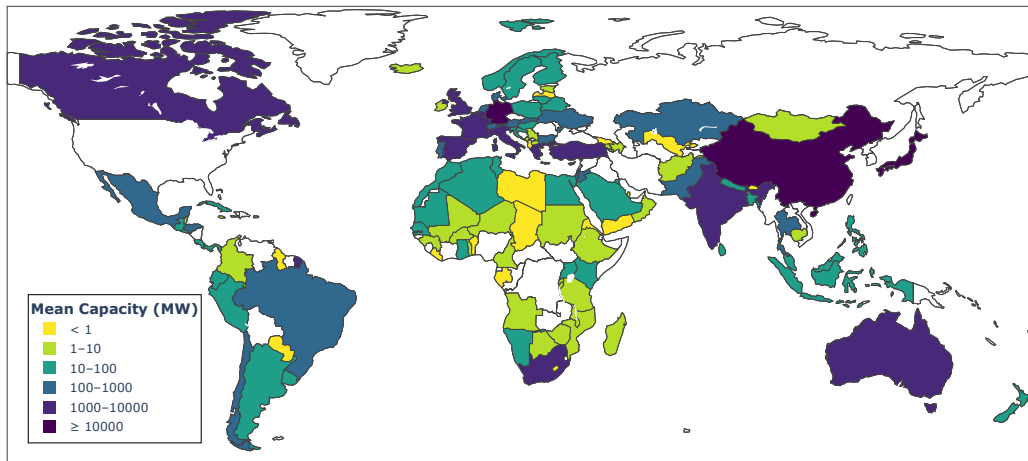


Figure 3.5: Mean cumulative photovoltaic capacity across various countries from 2001 to 2018. Countries without available data are represented in white.

Figure 3.6 shows the solar PV deployment index (SPVDI) for countries during the period from 2010 to 2018. Notably, Italy installed more capacity than expected from similar countries by 10 GW, the United Kingdom by 3.2 GW, Mexico by 2.4 GW, Chile by 1.3 GW, and Hungary by 0.5 GW. On the other hand, Spain installed less than expected by 9.1 GW, France by 5.1 GW, Canada by 2.8 GW, China by 1.7 GW, and India by 1.5 GW.

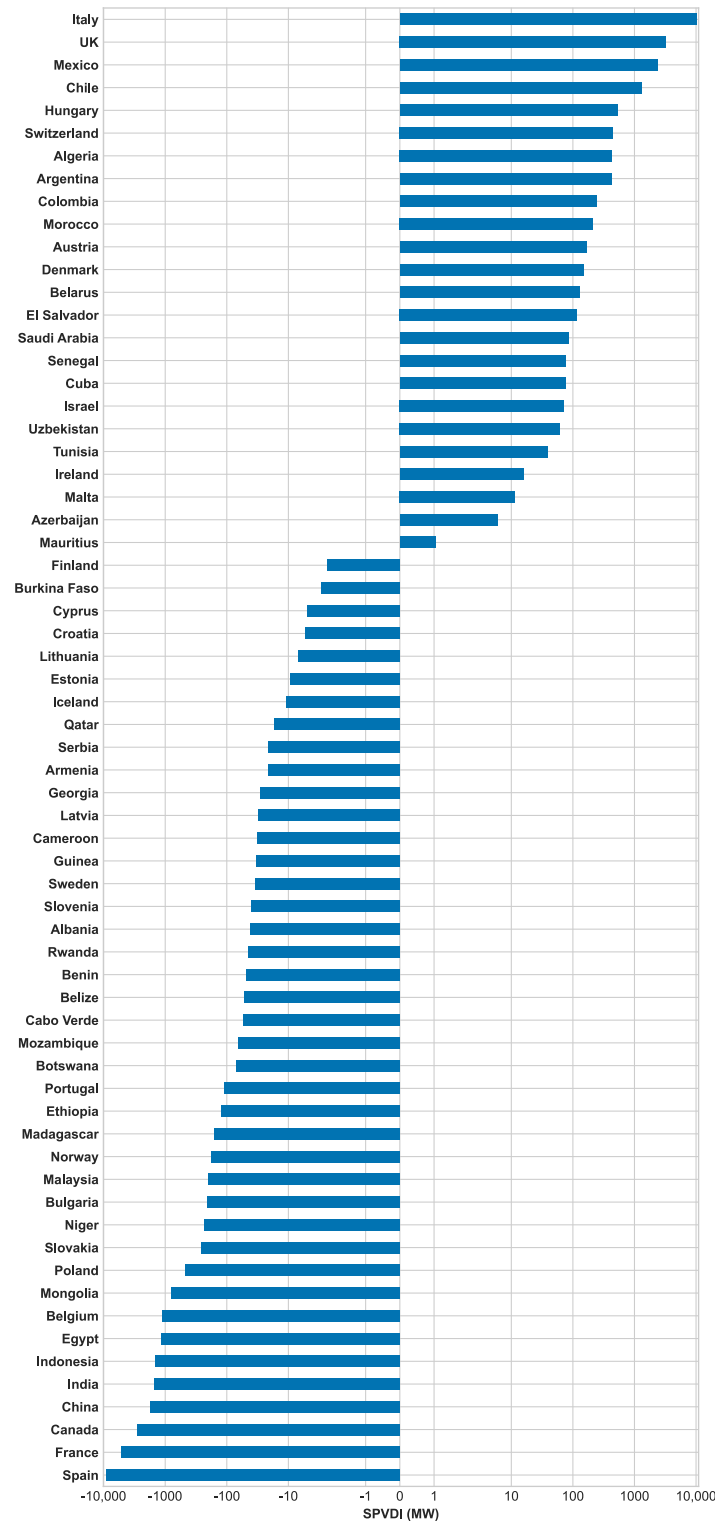


Figure 3.6: Solar PV deployment index (SPVDI) for different countries during the period from 2010 to 2018. Countries are ranked based on their solar photovoltaic deployment compared to other countries with similar social, economic, and land-use factors. A positive value indicates that a country has more capacity than expected, while a negative value means less capacity than expected from similar countries.

The top 10 countries in terms of PV deployment based on the SPVDI are Italy, the United Kingdom, Mexico, Chile, Hungary, Switzerland, Algeria, Argentina, Colombia, and Morocco. In contrast, the bottom 10 countries are Spain, France, Canada, China, India, Indonesia, Egypt, Belgium, Mongolia, and Poland. This does not necessarily imply that these countries are over- or underdeploying solar PV, as the index does not consider each country's individual targets. Nonetheless, the SPVDI offers valuable insights into a country's performance relative to others with similar geographic characteristics. Improving the index could involve integrating each nation's capacity targets; however, such data are often lacking for many countries and, when available, may not specifically pertain to solar PV but encompass renewable energy overall.

Figure 3.7 shows the actual versus estimated PV capacity additions for Italy, the United Kingdom, Mexico, and Spain from 2001 to 2018. Prior to 2008 in Italy, the actual capacity additions consistently fell short of expectations despite the implementation of the “Photovoltaic Roofs” program, which ran between 2001 and 2003 [217]. This initiative, offering up to 75% of installation costs for systems ranging from 1 kW to 20 kW connected to the distribution network [218], led to a deficit of 0.9 GW of installed capacity compared to what was expected. Feed-in tariff programs, “Conto Energia” (CE), were introduced in 2005, spanning five phases. The first CE, which ran between 2005 and 2006, achieved its 0.5 GW target [219]. Despite reaching the target, this phase led to a deficit of 0.7 GW compared to what was expected. The second CE was planned to last between 2007 and 2010, but it was extended to include PV systems installed before 31 December 2010 and operating before 30 June 2011, which led to a surge of investment to benefit from the feed-in tariffs [219]. The policy continued through the third CE, which entered into force in 2010 and was for PV systems commissioned between 1 January 2011 and 31 May 2011 [219]. The second and third CE programs were successful as they led to a surplus of 9.5 GW in installed capacity. The fourth CE program witnessed a significant reduction in tariffs on a monthly basis and expired in August 2012 [219]. The fourth CE was successful despite the significant reduction in rates as capacity additions were higher than expected by 10.3 GW. The fifth CE was introduced

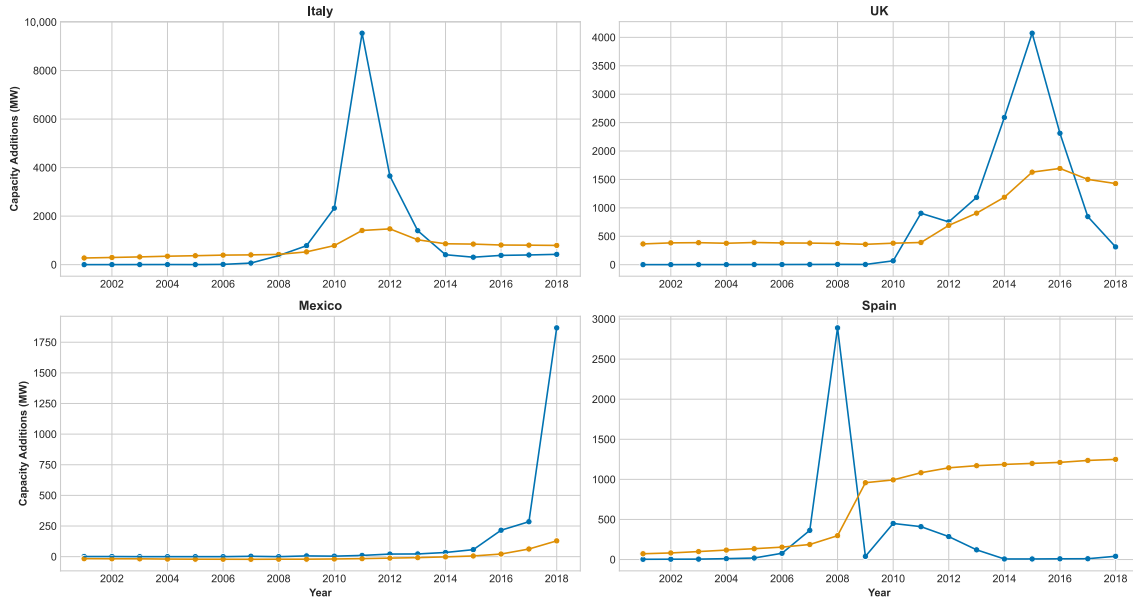


Figure 3.7: Actual versus estimated solar photovoltaic capacity additions for Italy, the United Kingdom, Mexico, and Spain for years between 2001 and 2018. Blue points represent actual capacity, while orange points represent estimated capacity.

in 2012 and ended in 2013, during which capacity additions were higher than expected by 2.6 GW. Overall, the CE programs led to 11.3 GW of capacity additions above what was expected. Following the conclusion of the CE scheme, a new tax credit system was implemented in 2013 [219]. However, capacity additions in the subsequent years dropped below expectations by 1.8 GW.

In the United Kingdom, there are two main policies when it comes to solar PV: renewable obligation certificates (ROCs) for systems above 50 kW of rated power and feed-in tariffs (FIT) for systems below 5 MW of rated power [220, 221]. ROCs were introduced in 2002 for England, Wales, and Scotland and in 2005 in Northern Ireland [162]. Despite the implementation of the scheme, capacity additions remained lower than expected until 2011, when the government declared that it would extend the scheme in England and Wales from 2027 to 2037 and would change it from a live-traded scheme to a fixed price certificate (FPC)-based scheme [162]. This increased capacity additions by 4.7 GW compared to the expected level until 2017 when the closure of the scheme [161] led to fewer additions than expected by 1.1 GW in the following year. The FIT scheme was launched in 2010 and ended in 2019. How-

ever, in 2016, a cap was applied to the number of new installations that could be accredited [220]. This led to fewer installations than expected by 1.8 GW in the next two years. Prior to this, the FIT scheme increased capacity additions by 5 GW compared to what was expected.

The actual additions are very close to the expected additions in Mexico up until the year 2015, after which capacity additions increased. This coincides with the introduction of long-term energy auctions. These auctions ran three times in 2015, 2016, and 2017. During these auctions, retailers would announce their requirement of capacity, consumption, or clean energy certificates, and generators would bid for them separately or in packages [222]. This led to more capacity than expected by 0.5 GW between 2015 and 2017. In 2018, major power consumers were required to buy 5% of electricity from power purchase agreements (PPAs) with clean power suppliers or through purchasing clean energy certificates [223], and the 15% customs duty on solar PV module imports which was introduced in 2015 was eliminated [224]. This led to more capacity additions by 1.7 GW compared to what was expected.

Spain introduced FITs in 1997. Generators could choose between fixed FITs that were adjusted annually or fixed premiums paid on top of the electricity market price. This was amended in 2004 so that FITs were set as a percentage of the electricity price and revised every 4 years [85], and were guaranteed to be paid for the lifetime of the solar power plants [225]. These policies had no significant impact on deployment, as capacity remained less than expected by 0.5 GW from 2001 to 2006. The FIT policy was revised again in 2007 so that FIT rates were fixed and updated every 4 years starting in 2010, or once 85% of the capacity target was reached, which ended up happening in the same year. This meant the government was going to lower the FIT rate, which led to a surge of investment to take advantage of the current FIT before the new tariff was implemented. During this period, capacity additions blew up and were higher than expected by 2.8 GW. The government responded by introducing policies that aimed to decrease deployment such as annual capacity quotas, the lifetime of FIT payments was reduced to 25 years for new plants, and FIT rates were reduced for small- and medium-sized solar PV. Further policies were introduced in 2010 such as limiting running hours eligible for FIT payments, reducing the FIT lifetime to 25 years for all existing plants,

and reducing FIT rates further. These measures were not enough, as the government had to implement additional measures in 2012 such as introducing a moratorium on support for new systems and revising FIT rates [85]. Finally, a sun tax was introduced in 2014 which aimed to cover the cost of balancing the grid [226]. All of these measures reduced additions to less than expected by 10.1 GW in the period between 2009 and 2018.

3.6 Conclusions

The previous year's cumulative PV capacity, population, agricultural land area, tertiary education, and electricity net consumption are identified as key features in explaining solar PV deployment. Using these features, the model achieves a global error of less than 10%. With country errors also below 10% in many cases, the model serves as a reliable forecasting tool across various nations.

Furthermore, the solar PV deployment index provides governments and policy makers with a benchmark for evaluating a country's performance relative to others with similar social, economic, and land-use characteristics. This index could aid in setting feasible solar PV targets. Additionally, the model offers a means to assess the efficacy of solar PV policies by comparing actual deployment against expected figures. Such analysis can inform policy refinement and enhance the likelihood of achieving national targets.

Future research should concentrate on enhancing model accuracy for countries with low capacity additions, extending its forecasting utility to these regions. Lastly, investigating the correlation between geographic factors and solar PV capacity types (residential, commercial, utility scale) presents an intriguing avenue for future exploration.

Data Availability

The dataset used for the analysis is publicly available and has been compiled from various open-source data sources. The compiled dataset, along with the code used for analysis, is accessible at: <https://doi.org/10.5281/zenodo.17143882>

3.7 Appendix

3.7.1 Data Availability for Correlation Analysis

Table 3.3: Number of data points used in the analysis presented in Table 3.1 to examine the correlation between each factor and solar PV capacity additions.

Feature	N
Last year's PV cumulative capacity (MW)	4320
Solar PV module cost (2019 USD per W)	4320
Land area (km ²)	3948
Cloud coverage (%)	3948
Daylight hours (min/day)	3948
Precipitation (mm/h)	3948
Snowfall (mm/h)	3948
Temperature (°C)	3948
Population	3939
Forest area (km ²)	3939
Average theoretical potential GHI (kWh/m ² /day)	3864
GDP (current USD)	3857
GNI (current USD)	3765
KOFGI	3718
Agriculture, forestry, and fishing, value added (current USD)	3686
Labour force	3675
Total unemployment	3675
Access to electricity (% of population)	3655
Industry (including construction), value added (current USD)	3646
Manufacturing, value added (current USD)	3556
Agricultural land (km ²)	3549
Nuclear electricity net generation (billion kWh)	3538
Duration of compulsory education (years)	3444
Electricity net generation (billion kWh)	3433
Fossil fuels electricity net generation (billion kWh)	3412
Electricity net consumption (billion kWh)	3400
Urban land area (km ²)	3381
Rural land area (km ²)	3381
Primary education	3028
Secondary education	2601
Tertiary education	2367
Research and development expenditure (% of GDP)	1603
Gini index (World Bank estimate)	1295
Public investments in solar energy (2019 million USD)	1001
Investment in energy with private participation (current USD)	754
Ease of doing business rank	178

3.7.2 Summary of PCA Cluster Characteristics

Table 3.4: Summary of the characteristics of principal component analysis (PCA) clusters, including the number of variables within each cluster, the coefficient of determination (R^2) obtained when fitting cluster variables into a linear model with PV capacity additions, and the corresponding number of data points used in the fitting process.

Cluster	Number of Features within Each Cluster	R^2	N
1	10	0.769	2257
2	6	0.211	21
3	5	0.222	570
4	7	0.463	1763
5	2	0.002	3760
6	4	0.163	327
7	2	0.010	3125

3.7.3 Descriptive Statistics of Model Features

Table 3.5: Descriptive statistics of the features used in the models. This table shows the count, mean, standard deviation, and distribution quartiles of the selected features.

	Added PV Capacity (MW)	Last Year's PV Cumulative Capacity (MW)	Population	Agricultural Land (km ²)	Tertiary Education	Electricity Net Consumption (Billion kWh)
count	1.84×10^3	1.84×10^3	1.84×10^3	1.84×10^3	1.84×10^3	1.84×10^3
mean	1.81×10^2	6.12×10^2	4.90×10^7	2.76×10^5	1.39×10^7	9.82×10^1
std	1.94×10^3	4.88×10^3	1.79×10^8	6.63×10^5	4.94×10^7	4.08×10^2
min	-5.60×10^1	0	4.39×10^4	6.60×10^0	5.94×10^2	3.00×10^{-2}
25%	0	0	3.03×10^6	1.51×10^4	6.00×10^5	2.85×10^0
50%	1.97×10^{-1}	1.13×10^0	9.79×10^6	5.05×10^4	2.40×10^6	1.29×10^1
75%	4.47×10^0	1.50×10^1	2.86×10^7	2.63×10^5	8.16×10^6	6.17×10^1
max	5.30×10^4	1.31×10^5	1.40×10^9	5.29×10^6	7.10×10^8	6.45×10^3

3.7.4 Pearson Correlation Matrix of Model Input Features

Table 3.6: Pearson correlation coefficients among selected variables used as inputs for the models.

	Added Capacity (MW)	Last Year's Cumulative Capacity (MW)	Population	Agricultural Land (km ²)	Tertiary Education	Electricity Net Consumption (Billion kWh)
Added Capacity (MW)	1.00	0.79	0.42	0.39	0.70	0.72
Last Year's Cumulative Capacity (MW)	0.79	1.00	0.30	0.27	0.54	0.58
Population	0.42	0.30	1.00	0.74	0.88	0.77
Agricultural Land (km ²)	0.39	0.27	0.75	1.00	0.73	0.75
Tertiary Education	0.70	0.54	0.88	0.73	1.00	0.92
Electricity Net Consumption (Billion kWh)	0.72	0.58	0.77	0.75	0.92	1.00

3.7.5 Feature Selection: Stepwise Regression

Feature selection was performed using the stepwise regression in JMP Student Edition 18 software. Prior to analysis, predictors with fewer than 2000 non-missing observations were excluded, leaving 32 candidate variables. Both forward selection and backward elimination were applied, with model fit evaluated using the corrected Akaike Information Criterion (AICc) and the Bayesian Information Criterion (BIC).

The corrected Akaike Information Criterion (AICc) was used instead of AIC because the ratio of sample size to parameters was relatively modest ($n/k \approx 33$), where AICc is recommended [227].

Backward elimination was emphasized because the sample size ($n = 1056$) was large relative to the number of predictors, allowing the full model to be estimated reliably. AICc was used to yield feature sets optimized for predictive performance, while BIC was used to

return more parsimonious subsets emphasizing interpretability. Forward selection was also tested to confirm whether results were robust to the choice of search direction.

The fit statistics for the models identified by each procedure are reported in Table 3.7. The AICc-based models retained 21–22 predictors, achieved $R^2 \approx 0.59$, and produced RMSE values of ≈ 412 MW. The BIC-based models retained fewer predictors (13–18), consistent with BIC’s stronger penalty on model complexity, but explained slightly less variance ($R^2 = 0.57$ – 0.59) with marginally higher RMSE.

The predictors retained in each model are listed in Table 3.8. Several variables were consistently selected across criteria and directions, including last year’s PV cumulative capacity, urban land area, forest area, GDP, GNI, primary and tertiary education, labor force, industry value added, and fossil fuel electricity generation. However, the exact subsets varied depending on the selection criterion and direction.

Overall, stepwise regression produced models with 13–22 predictors and $R^2 \approx 0.57$ – 0.59 . By comparison, the proposed feature selection method in Appendix 3.4.2 achieved higher explanatory power ($R^2 = 0.79$) using only five predictors.

Table 3.7: Fit statistics for stepwise regression models under different criteria and directions. SSE is the sum of squared errors, DFE is the error degrees of freedom, RMSE is the root mean square error, R^2 is the coefficient of determination, Adj. R^2 is the adjusted coefficient of determination, Cp is Mallows’ Cp statistic, p is the number of parameters (including the intercept), AICc is the corrected Akaike’s Information Criterion, and BIC is the Bayesian Information Criterion.

Criterion	Direction	SSE (MW ²)	DFE	RMSE (MW)	R^2	Adj. R^2	Cp	p	AICc	BIC
AICc	Backward	175574460	1035	411.87	0.5908	0.5829	17.55	21	15736.3	15844.5
AICc	Forward	175387153	1034	411.85	0.5912	0.5829	18.45	22	15737.3	15850.3
BIC	Backward	177159807	1038	413.13	0.5871	0.5803	20.86	18	15739.6	15833.1
BIC	Forward	184047714	1042	420.27	0.5711	0.5657	53.33	13	15771.6	15845.5

Table 3.8: Predictors selected by stepwise regression.

Criterion	Direction	Selected Features
AICc	Backward	Year; Last year's PV cumulative capacity (MW); Population; Urban land area (km ²); Agricultural land (km ²); Forest area (km ²); KOFGI; GDP (current US\$); GNI (current US\$); Primary education; Secondary education; Tertiary education; Labor force; Total unemployment; Agriculture, forestry, and fishing, value added (current US\$); Industry (including construction), value added (current US\$); Electricity net generation (billion kWh); Fossil fuels electricity net generation (billion kWh); Precipitation; Snowfall (mm/h)
AICc	Forward	Year; Last year's PV cumulative capacity (MW); Population; Urban land area (km ²); Agricultural land (km ²); Forest area (km ²); KOFGI; GDP (current US\$); GNI (current US\$); Primary education; Secondary education; Tertiary education; Labor force; Total unemployment; Agriculture, forestry, and fishing, value added (current US\$); Industry (including construction), value added (current US\$); Electricity net generation (billion kWh); Nuclear electricity net generation (billion kWh); Fossil fuels electricity net generation (billion kWh); Precipitation; Snowfall (mm/h)
BIC	Backward	Last year's PV cumulative capacity (MW); Population; Urban land area (km ²); Agricultural land (km ²); Forest area (km ²); KOFGI; GDP (current US\$); GNI (current US\$); Primary education; Secondary education; Tertiary education; Labor force; Agriculture, forestry, and fishing, value added (current US\$); Industry (including construction), value added (current US\$); Electricity net generation (billion kWh); Fossil fuels electricity net generation (billion kWh); Snowfall (mm/h)
BIC	Forward	Last year's PV cumulative capacity (MW); Urban land area (km ²); Forest area (km ²); GDP (current US\$); GNI (current US\$); Primary education; Tertiary education; Labor force; Total unemployment; Industry (including construction), value added (current US\$); Nuclear electricity net generation (billion kWh); Fossil fuels electricity net generation (billion kWh); Precipitation

3.7.6 Feature Importance, Correlation, and Variance Explained by Combined Model Features

Table 3.9: Feature importance, correlation, and variation explained by the combined model features. The main effect measures the contribution of the feature alone, while the total effect measures the contribution of the feature alone and in combination with other features.

Feature	Main Effect	Total Effect	Correlation with Capacity Additions	R^2
Last year's PV cumulative capacity (MW) * agricultural land (km ²)	0.13	0.72	0.95	0.90
Last year's PV cumulative capacity (MW) ²	0.04	0.42	0.81	0.65
Last year's PV cumulative capacity (MW) * population	0.06	0.26	0.97	0.93
Last year's PV cumulative capacity (MW) * electricity net consumption (billion kWh)	0.05	0.23	0.96	0.92
Agricultural land (km ²) * electricity net consumption (billion kWh)	0.10	0.21	0.65	0.42
Last year's PV cumulative capacity (MW) * tertiary education	0.04	0.20	0.96	0.93
Electricity net consumption (billion kWh) ²	0.03	0.18	0.77	0.60
Population * electricity net consumption (billion kWh)	0.05	0.18	0.67	0.45
Agricultural land (km ²) ²	0.09	0.09	0.36	0.19
Population * agricultural land (km ²)	0.06	0.08	0.46	0.22
Agricultural land (km ²) * tertiary education	0.04	0.06	0.71	0.51
Electricity net consumption (billion kWh)	0.05	0.05	0.67	0.45
Population ²	0.05	0.05	0.43	0.18
Population	0.05	0.05	0.40	0.16
Agricultural land (km ²)	0.04	0.04	0.36	0.13
Tertiary education ²	0.04	0.04	0.85	0.71
Tertiary education * electricity net consumption (billion kWh)	0.02	0.02	0.83	0.69
Population * tertiary education	0.02	0.02	0.69	0.47
Tertiary education	0.02	0.02	0.67	0.45
Last year's PV cumulative capacity (MW)	0.02	0.02	0.70	0.48

3.7.7 MLR Model

The multiple linear regression (MLR) model used to estimate solar PV capacity additions is summarised in Table 3.10. The predictors include previous year's cumulative PV capacity (PV), population (POP), agricultural land area (AL), tertiary education (ED), and electricity net consumption (EC).

Table 3.10: Parameter estimates for the MLR model.

Predictor	Estimate	Std. Error	t-ratio	p-value
Intercept	-124.87	27.17	-4.60	< .0001
PV	0.145	0.0075	19.36	< .0001
POP	-6.21×10^{-6}	3.63×10^{-7}	-17.08	< .0001
AL	-0.000279	0.000065	-4.27	< .0001
ED	3.86×10^{-5}	2.03×10^{-6}	19.02	< .0001
EC	0.659	0.179	3.69	0.0002

3.7.8 Polynomial Model

The polynomial model for estimating solar PV capacity additions is summarised in Table 3.11. The predictors include previous year's cumulative PV capacity (PV), population (POP), agricultural land area (AL), tertiary education (ED), and electricity net consumption (EC).

Table 3.11: Parameter estimates for the polynomial model.

Predictor	Estimate	Std. Error	t-ratio	p-value
Intercept	-4.465	12.09	-0.37	0.7120
PV	0.1318	0.0102	12.87	< .0001
POP	7.55×10^{-7}	4.67×10^{-7}	1.62	0.1062
AL	-0.000266	0.0000635	-4.19	< .0001
ED	-2.25×10^{-6}	2.16×10^{-6}	-1.04	0.2984
EC	1.251	0.214	5.85	< .0001
PV ²	-5.44×10^{-6}	3.39×10^{-7}	-16.07	< .0001
PV $\times$ POP	4.81×10^{-10}	9.03×10^{-11}	5.32	< .0001
PV $\times$ AL	-8.92×10^{-8}	1.21×10^{-8}	-7.40	< .0001
PV $\times$ ED	-1.10×10^{-9}	4.42×10^{-10}	-2.50	0.0125
PV $\times$ EC	0.000267	0.0000300	8.89	< .0001
POP ²	-1.10×10^{-15}	8.30×10^{-16}	-1.33	0.1838
POP $\times$ AL	-6.44×10^{-13}	2.39×10^{-13}	-2.70	0.0070
POP $\times$ ED	-2.30×10^{-14}	9.25×10^{-15}	-2.49	0.0131
POP $\times$ EC	9.05×10^{-9}	1.44×10^{-9}	6.29	< .0001
AL ²	3.14×10^{-10}	3.62×10^{-11}	8.68	< .0001
AL $\times$ ED	7.83×10^{-12}	1.99×10^{-12}	3.94	< .0001
AL $\times$ EC	-2.89×10^{-6}	3.78×10^{-7}	-7.63	< .0001
ED ²	6.12×10^{-14}	2.82×10^{-14}	2.17	0.0298
ED $\times$ EC	-2.22×10^{-8}	4.43×10^{-9}	-5.01	< .0001
EC ²	0.00169	0.000260	6.52	< .0001

Chapter 4

European Model

4.1 Abstract

The rapid expansion of solar photovoltaic (PV) technology has established it as a leading contributor to global renewable energy capacity. However, integrating solar PV into existing power grids presents significant challenges, primarily due to the variable nature of solar energy generation and the lack of accurate and complete data on installed PV capacity at the regional level. This study addresses this critical gap in capacity measurement by analysing the factors influencing regional solar PV deployment and developing models to estimate installed PV capacity across 333 regions of 36 European countries. We employed Pearson and Spearman correlation analyses to identify key geographic factors such as agricultural land area, solar irradiance and population, related to solar PV deployment. This informed the development of extreme gradient boosted parallel tree algorithm (XGBoost) models for estimating regional PV capacity. The models achieve a root mean squared capacity error (RMSE) of less than 272 MW, and explain more than 93% of the variation across 150 NUTS 2 EU regions. The models serve three primary purposes: disaggregating national PV capacity into regional figures, benchmarking inter- and intra-regional capacities, and forecasting future PV capacity distribution. The models presented in this study offer a comprehensive tool for policymakers and grid operators, enabling the design of more effective policy interventions

and enhanced solar PV monitoring services. This research contributes to more sustainable and efficient energy planning in Europe.

4.2 Introduction

Solar photovoltaic (PV) technology has emerged as one of the most cost-effective sources of electricity [41], leading to its widespread adoption globally. In 2023, solar PV continued to be the leading force behind global renewable capacity expansion, contributing to 65% of the growth [228]. Solar PV offers significant advantages in the energy transition: it is cost-effective, scalable, widely available, and has a low environmental footprint. However, it also presents operational challenges due to its strong seasonal dependence, day/night production variability, and the difficulty of short-term forecasting. These characteristics make it essential to incorporate PV into more granular and spatially resolved energy system models that can account for its variability and location-dependent performance.

As solar PV capacity continues to expand, managing and balancing the power grid becomes increasingly challenging due to the variable and intermittent nature of solar energy [62]. Historically, electricity system operators maintained grid stability by dispatching a limited number of large, controllable fossil fuel plants to meet predictable demand. However, the integration of weather-dependent renewable energy sources, which are often embedded within distribution networks and thus not directly visible to transmission system operators, complicates this traditional model [49]. Such embedded generation often manifests as a reduction in apparent demand, thereby introducing substantial uncertainties in power flow forecasts [61], which necessitates the development of solar PV monitoring services.

Solar PV monitoring services estimate national or regional generation as the product of yield and installed capacity [49]. Solar PV yield is defined as the output of PV systems normalized to the total nominal capacity of the entire fleet, and is measured in units of megawatts (MW) per megawatt peak (MWp). The simplest method for calculating yield involves using PV generation data from a sample of PV systems, typically referred to as reference systems, which are then statistically upscaled to estimate regional or national yield

[63, 64, 65, 66]. More advanced methods incorporate meteorological data into statistical or physical models to estimate yield, which is then similarly upscaled to provide regional or national estimates [64, 67, 68, 69, 70, 71]. Other methods exist for estimating regional or national PV generation that do not rely on upscaling [61, 72, 73]; however, these are less commonly used due to limitations such as the need for historical measurement data from every site, typically over a limited time period, to train the models, or the requirement of module orientation angles for a large number of PV systems. Regardless of the method, all approaches require knowledge of the installed capacity.

Despite the critical importance of regional capacity data, many European countries lack accurate records, hindering efficient grid management and policy-making. In 2023, 56% of European regions lacked regional PV capacity data, as shown in Appendix 4.6.5, highlighting the scale of this data availability problem. While countries such as the United Kingdom, Italy, Spain, Belgium, Germany, and France maintain data on regional solar PV capacity [44, 50, 51, 52, 53, 55], a portion of the national capacity in these countries often remains unallocated to specific geographic locations. Moreover, even where capacity data is available, inaccuracies can lead to significant errors in estimating regional PV output [49].

Regional PV capacity data is not only vital for real-time grid operations but also for long-term planning and policy-making. The expansion of large-scale PV projects can lead to conflicts over land use, particularly where agricultural land is repurposed for energy production, resulting in competition between farming activities and solar energy generation [92, 93, 94, 95, 96, 97]. The lack of comprehensive policies to address these conflicts can foster distrust and hinder the deployment of solar PVs. For example, in Brandenburg, Germany, the lack of a comprehensive legal framework to manage land use conflicts related to solar energy installations has led to land grabbing by solar investors, resulting in serious disputes with farmers and a lack of trust in large-scale PV projects [96].

The implementation of policies to support solar PV deployment targets without taking into account the complex relationship between geographic factors (e.g., social, economic, climatic, and land use) and PV deployment can result in unintended outcomes, such as

over or under deployment, and potentially hinder the achievement of targets. Moreover, assessing a region's solar PV progress based on self-defined targets may not accurately reflect its true deployment potential, as it overlooks the intricate interaction between deployment and geographical factors.

The development of models that estimate regional PV capacity based on geographical factors could significantly enhance solar PV monitoring services, help mitigate land use conflicts by anticipating potential hotspots, and enable more informed policy interventions. This study addresses a critical and often overlooked problem: the lack of accurate, complete, and spatially resolved data on installed solar PV capacity. We do not attempt to model peak generation or total energy production of the EU PV fleet. Instead, our focus is on the estimation of installed capacity at the regional level, a foundational input for both real-time generation monitoring and long-term planning models. Accurately quantifying where capacity is installed is essential for reliable yield modelling, power flow forecasting, and evidence-based policy design. The objective of this study is to investigate the geographic factors that influence regional solar PV deployment and to develop predictive models that estimate installed PV capacity across 36 European countries. The models are designed to serve three key purposes: first, as a disaggregation tool for breaking down national PV capacity to improve generation monitoring; second, as a benchmarking tool for comparing regions with similar characteristics; and third, as a forecasting tool to support effective long-term planning and policy-making for PV capacity.

4.3 Methodology

We previously reviewed and analysed the relationship between social, economic, climatic, and land use factors and national solar PV deployment, concluding that these variables can effectively model national PV capacity [229]. To identify factors associated with regional solar PV deployment, we adopt a similar methodology, analysing the relationships between these geographical factors and solar PV deployment at the regional level across 36 European countries. Given the potential non-linear relationships between these factors and PV capacity,

we analyse these relationships using both Pearson and Spearman correlation coefficients. Additionally, we calculate the coefficient of determination (R^2) by fitting linear regression models between each feature and solar PV capacity.

4.3.1 Data and Data Processing

Data for our analysis were obtained from various publicly accessible sources. Climate variables such as global horizontal irradiance (GHI), wind speed, total precipitation, sea-level pressure, and 2-meter air temperature were sourced from the Copernicus Climate Change Service [230]. Land use data came from the CORINE Land Cover 2018 dataset [231]. Socioeconomic indicators, including population, gross domestic product (GDP), employment, and age, were retrieved from Eurostat [232]. Solar PV capacity data were gathered from grid operators, government agencies, research institutions, and the International Renewable Energy Agency (IRENA) [233].

Our analysis is framed within the NUTS 2021 (Nomenclature of Territorial Units for Statistics) classification, a hierarchical system used to divide the European Union's territory for statistical purposes. NUTS regions are classified into three levels, each defined by specific population thresholds to ensure statistical consistency [234]. For this study, we focused on NUTS 2 regions, with populations ranging between 800,000 and 3 million, as data at this level are widely available across Europe. All data were processed according to the NUTS 2 regional classification and aggregated annually.

The CORINE land cover dataset provides a three-tier hierarchical classification system. Level 3, with its 44 thematic classes, offers detailed analysis but has a thematic accuracy of over 85% [235], potentially leading to some misclassification. Higher levels, such as Levels 1 and 2, have fewer categories, which generally increases thematic accuracy [236, 237, 238] and reduces misclassification risk. However, these broader classifications provide less detail on specific land use types, which is crucial for this study. To address this, we incorporate all three levels into the analysis: Level 3 provides detailed insights into specific land cover types, while Levels 1 and 2 offer a broader context to help mitigate misclassification risks.

For regional PV capacity data, we focused on countries where capacity information was available, such as the United Kingdom [44], Italy [50], Spain [51], Belgium [52], Germany [53], and France [55]. In these countries, capacity data were processed at both the NUTS 2 (regional) and NUTS 0 (national) levels, with regional capacity at the NUTS 2 level illustrated in Figure 4.1. For countries without regional data, national capacity figures were obtained from IRENA [233].

We constructed two primary datasets: one containing absolute values and another where values were normalized to either the national total or national average, depending on the variable. The absolute dataset is used to model capacity in megawatts (MW), whereas the normalized dataset is used to model the regional share of capacity (as a percentage of the national total). This dual dataset approach facilitates a more detailed analysis by accounting for both the absolute contributions of various factors and their relative importance at a national level. The datasets cover 36 European countries, 333 regions, and span the years 2010 to 2023.

In the normalized dataset, climate data were normalized relative to national averages to highlight significant regional climatic variations. Socioeconomic factors were normalized based on the variable: disposable income per inhabitant and median age were normalized using national means, while aggregate metrics—such as population, disposable income, employment, GDP, gross domestic expenditure on R&D (GERD), and unemployment—were normalized by national totals. Finally, regional solar PV capacities were normalized relative to the national total capacity.

4.3.2 Feature Selection

The feature selection process adhered to the DAMA data quality framework [239], which considers key aspects such as accuracy, completeness, uniqueness, consistency, timeliness, and validity. Features that were complete for the majority of countries and consistent over time were prioritized over those that were only available for certain countries or specific years. Selection was based on two main criteria: correlation with PV capacity and data

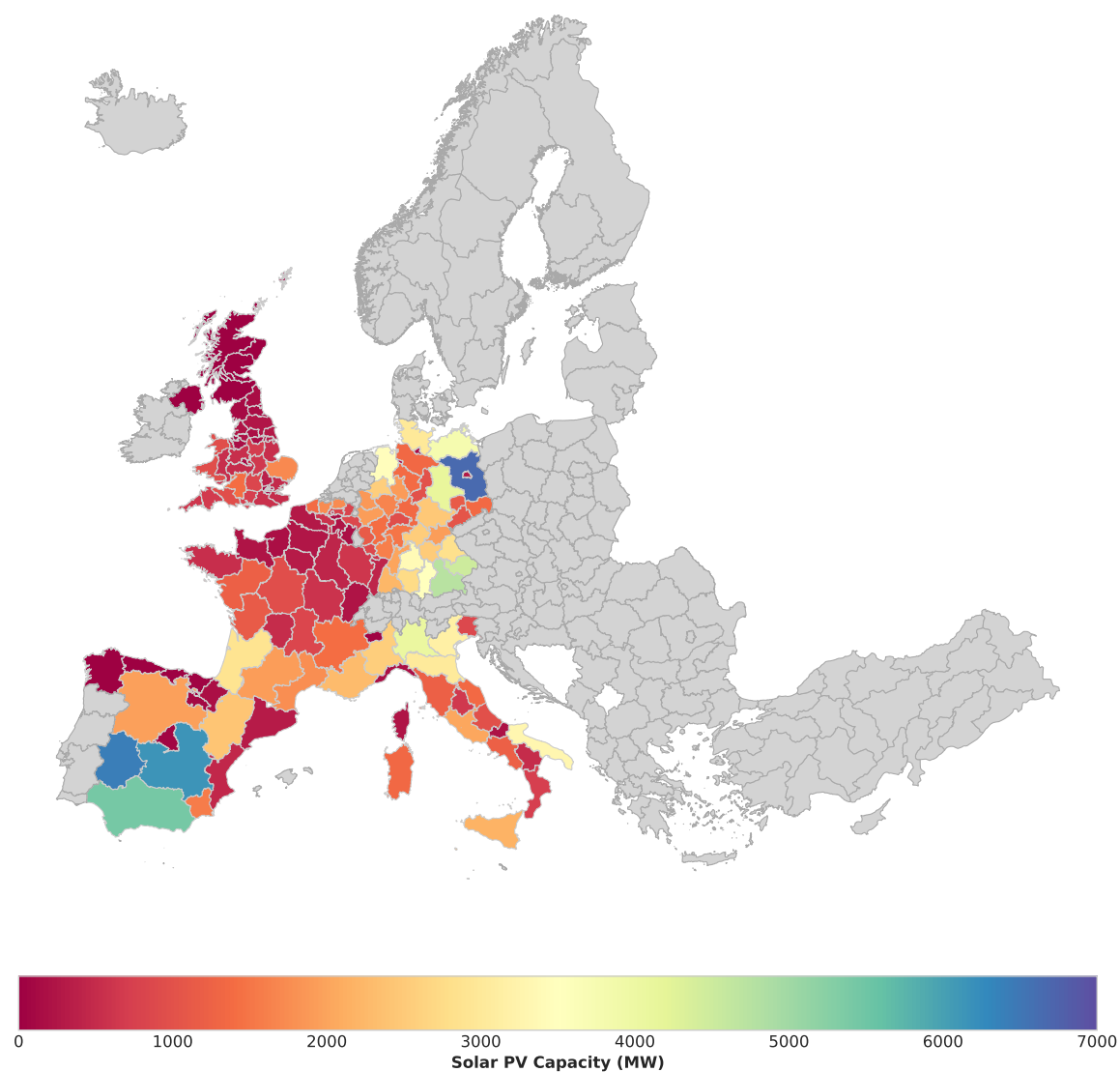


Figure 4.1: Actual regional photovoltaic capacity for 150 NUTS 2 regions in 2023. The capacity data is sourced from the United Kingdom, Italy, Spain, Belgium, Germany, and France.

availability. To account for both linear and non-linear associations, we calculated the average of Pearson and Spearman correlation coefficients. These two metrics capture different types of relationships - linear in the case of Pearson, and monotonic for Spearman - and their combined interpretation can uncover hidden relationships. It is recommended to calculate both Pearson and Spearman correlations when the data includes multiple distributions [240]. Rovetta [240] also highlights a peculiar statistical phenomenon in which monotonic correlations emerge only above certain thresholds. In such cases, Pearson's coefficient may actually outperform Spearman's, despite the data being non-normally distributed. This is because Pearson's method gives more weight to extreme values, which may be especially informative in identifying high capacity regions. Averaging the two coefficients thus provides a composite measure that balances the magnitude sensitivity of Pearson with the distributional robustness of Spearman, enhancing the overall reliability of the feature selection process. To determine the optimal configuration, we tested a range of correlation thresholds from 0.2 to 0.5, with a data availability threshold set at 90%.

4.3.3 Model Training

We selected extreme gradient boosted parallel tree algorithm (XGBoost) for our modelling approach due to several key advantages. Firstly, decision tree-based models such as XGBoost are nonparametric [241]. This is crucial for our study as our data is often not normally distributed, making XGBoost a better fit compared to parametric models such as linear regression. Secondly, XGBoost performs exceptionally well with tabular data [242], which matches the structure of our dataset.

Two XGBoost models were developed: one using absolute data to predict regional capacity (in MW) and another using normalized data to predict each region's share of the national total (in percentage terms). Data were grouped by country and year and modelled using three distinct validation schemes. In the first approach (mixed hold-out), the data were grouped by country and year and then partitioned, with 80% used for training and 20% for testing. In the second approach (hold-out country), the data were grouped by country, and the model

was trained on all countries except one, which was held out for testing. In the third approach (hold-out year), the model was trained on data for the years 2010–2020 and tested on 2021–2023. These approaches jointly assess both temporal and spatial generalisation performance. The mixed hold-out configuration was subsequently used for model applications presented in this study.

Data that are right-skewed, contain zeros, and are non-negative can be effectively modelled using the Tweedie distribution [243, 244]. Since regional solar PV capacity exhibits these characteristics, the Tweedie objective was adopted for the absolute XGBoost model. General linear models (GLMs) employing the Tweedie distribution have shown strong performance for similar types of data, such as rainfall [245], health costs [244], and insurance claims [246]. To further verify this choice, we compare the XGBoost model trained with the Tweedie objective to a standard model using the Gaussian objective under the mixed hold-out configuration, and present results demonstrating the improvement in performance.

As a benchmark, we also trained a multiple linear regression (MLR) model on the same set of features used in XGBoost. The MLR was implemented in `scikit-learn` [214], which provides a suite of algorithms for supervised and unsupervised learning. The model is formulated as:

$$y = \beta_0 + \sum_{i=1}^p \beta_i x_i + \varepsilon, \quad (4.1)$$

where y is the dependent variable, β_0 the intercept, β_i the estimated coefficients, x_i the predictors for $i = 1, \dots, p$, and ε the error term representing unexplained variation. Parameters were estimated using Ordinary Least Squares (OLS). The MLR was implemented as an unscaled model, meaning regional predictions were used directly without adjusting them to match national totals. Scaling requires complete regional coverage for each year, which was unavailable in this dataset, and therefore could not be applied.

We quantify feature importance using Shapley Additive Explanations (SHAP), a unified framework for interpreting model predictions based on principles from cooperative game theory. In this framework, each feature is treated as a “player” that contributes to the model’s overall prediction. The Shapley value represents each feature’s average marginal

contribution to the prediction when considered across all possible combinations of the other features. This ensures that every feature’s influence is fairly assessed, regardless of the order in which it is added to the model.

The SHAP value for a given feature and observation therefore measures how much that feature increases or decreases the prediction relative to a baseline (the average model output). By design, SHAP attributions are additive: the sum of all feature contributions equals the difference between the prediction and the baseline [247, 248].

We compute SHAP values for each observation (local explanations) and summarise overall feature importance as the mean absolute SHAP value across the dataset. For tree ensembles (our XGBoost models), we use SHAP’s `TreeExplainer` (TreeSHAP), which provides additive, locally accurate attributions efficiently for tree-based models. Reported SHAP units correspond to the model output: percentage points for the normalized model and megawatts (MW) for the absolute model. To express each feature’s relative importance, we convert mean absolute SHAP values into percentage contributions using Equation (4.2):

$$\text{Contribution}_i = \frac{|\text{SHAP}_i|}{\sum_{j=1}^n |\text{SHAP}_j|} \times 100 \quad (4.2)$$

where $|\text{SHAP}_i|$ is the mean absolute SHAP value for feature i , and n is the total number of features.

To ensure the model’s predictions align with reality and are applicable to countries with unknown regional capacity values, we scaled the predicted regional capacity so it sums to the national capacity, and then compared the scaled values to actual values. This allows us to confidently apply this methodology to countries where regional capacity is unknown.

To evaluate the model’s performance, we used several error metrics, including Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). We avoid using MAPE when evaluating regional capacity percentage because it can yield large values due to data representation in percentages. Instead, MAPE is only applied to evaluate national capacity.

4.3.4 Model Applications

The models are applied in several ways: to disaggregate national capacity into regional capacity, to allocate capacity with an unknown location to a geographical region, and to serve as benchmarking tools. A Solar PV Deployment Index (SPVDI) was developed previously to assess national capacity in a global context [229]. In this study, we adapt the same concept to evaluate regional PV capacity within a European context.

The SPVDI serves as a benchmarking tool by comparing solar PV deployment in a region relative to others with similar social, economic, climatic, and land use characteristics. It is calculated as shown in Equation (4.3), where t_{start} and t_{end} denote the first and last years of the analysis period, respectively. The index sums the difference between actual and predicted capacity for region i over the specified time range. A positive SPVDI value indicates that a region has deployed more capacity than expected, while a negative value reflects lower-than-expected deployment.

The SPVDI allows for performance comparisons of regions over multiple years and time periods and serves as a tool to rank regions based on their solar PV deployment. When applied to the absolute model, the SPVDI compares solar PV deployment in a region relative to other regions across Europe. In contrast, when applied to the normalized model, the SPVDI evaluates solar PV deployment within a single country, highlighting regions that either exceed or fall short of deployment expectations in a national context.

$$\text{SPVDI}_i = \sum_{j=t_{\text{start}}}^{t_{\text{end}}} (\text{Actual Capacity}_{i,j} - \text{Predicted Capacity}_{i,j}) \quad (4.3)$$

4.4 Results and Discussion

4.4.1 Feature Analysis and Selection

Table 4.1 and Table 4.2 present the top 15 features with the highest average correlation to regional solar photovoltaic capacity in absolute and normalized terms, respectively. The full analysis, including all features along with data availability and relevant literature, is provided

in Appendix 4.6.1. In the absolute data, national PV capacity has the strongest correlation with regional PV capacity. Regions characterized by extensive artificial surfaces, large areas of arable land, coniferous forests, and a higher median age generally show higher PV capacity. In the normalized data, regions with higher percentages of land area, agricultural areas, artificial surfaces, population, disposable income, and employment—each calculated as a proportion of the national total—tend to hold a larger share of national PV capacity.

The strong correlation of artificial surfaces with PV capacity could be explained by the inclusion of energy production and distribution facilities, including solar installations, within this category [231]. These facilities could serve as a proxy for grid connection points, which are essential for installing grid-connected solar PV systems. Regions with greater access to such infrastructure are more likely to support large-scale solar installations. Additionally, the presence of large-scale or industrial agricultural facilities, also categorized under artificial surfaces, may further contribute to the high correlation with PV capacity observed in the results. The strong correlation between agricultural land and PV capacity is expected, as commercial, industrial, and utility-scale solar PV installations are predominantly sited on agricultural land, particularly arable land [249, 119].

Table 4.1: Top 15 features with the highest correlation to regional solar photovoltaic capacity in absolute terms. Displayed are the coefficient of determination (R^2), Pearson correlation, Spearman correlation, and the average of the two correlations. Land cover features are preceded by their CORINE Land Cover (CLC) classification codes (e.g., 1.2.1 for industrial or commercial units).

Feature	R^2	Pearson	Spearman	Corr. Avg.
National solar PV capacity (MWp)	0.31	0.56***	0.63***	0.59
1.2.1 Industrial or commercial units (m ²)	0.21	0.46***	0.55***	0.51
1.2 Industrial, commercial and transport units (m ²)	0.18	0.43***	0.52***	0.47
1 Artificial Surfaces (m ²)	0.15	0.38***	0.51***	0.45
1.1 Urban fabric (m ²)	0.13	0.37***	0.52***	0.45
2.1 Arable land (m ²)	0.13	0.36***	0.54***	0.45
2.1.1 Non-irrigated arable land (m ²)	0.12	0.35***	0.53***	0.44
3.1.2 Coniferous forest (m ²)	0.17	0.41***	0.46***	0.43
Median age (years)	0.15	0.38***	0.46***	0.42
1.1.2 Discontinuous urban fabric (m ²)	0.11	0.33***	0.50***	0.42
2 Agricultural Areas (m ²)	0.12	0.35***	0.47***	0.41
5 Water Bodies (m ²)	0.14	0.38***	0.41***	0.40
5.1.1 Water courses (m ²)	0.13	0.36***	0.43***	0.40
5.1 Inland waters (m ²)	0.15	0.38***	0.41***	0.40
2.2 Permanent crops (m ²)	0.08	0.28***	0.51***	0.40

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 4.2: Top 15 normalized features with the highest correlation to the percentage of regional solar photovoltaic capacity. Features are normalized relative to national values and expressed as percentages. Displayed are the coefficient of determination (R^2), Pearson correlation, Spearman correlation, and the average of the two correlations. Land cover features are preceded by their CORINE Land Cover (CLC) classification codes (e.g., 1.2.1 for industrial or commercial units).

Feature	R^2	Pearson	Spearman	Corr. Avg.
2 Agricultural Areas	0.49	0.70***	0.67***	0.69
Region area	0.38	0.61***	0.63***	0.62
2.1 Arable land	0.32	0.57***	0.63***	0.60
2.4 Heterogeneous agricultural areas	0.41	0.64***	0.52***	0.58
2.1.1 Non-irrigated arable land	0.28	0.53***	0.62***	0.57
1.2 Industrial, commercial and transport units	0.31	0.56***	0.52***	0.54
1.2.1 Industrial or commercial units	0.31	0.56***	0.51***	0.54
2.2 Permanent crops	0.29	0.54***	0.52***	0.53
Disposable income	0.22	0.47***	0.56***	0.52
2.4.2 Complex cultivation patterns	0.27	0.52***	0.52***	0.52
Employment	0.24	0.49***	0.53***	0.51
Population	0.25	0.50***	0.52***	0.51
1 Artificial Surfaces	0.22	0.47***	0.54***	0.51
1.3 Mine, dump and construction sites	0.23	0.48***	0.52***	0.50
1.1 Urban fabric	0.20	0.45***	0.55***	0.50

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

It is surprising that coniferous forests rank among the top features in our analysis. It's possible that the 2018 CORINE land cover data used in this analysis does not reflect recent deforestation that may have occurred to make way for PV systems. Previous research shows that forests rank fourth in terms of PV generation potential [95], with approximately 9.14% of solar PV farms located in forested areas [250]. This is typically observed in regions with high densities of PV installations, where prime locations such as agricultural land are already occupied, forcing PV installations into less ideal areas such as forests. We find a positive correlation between demographics and PV capacity. This result aligns with some studies that suggest a positive correlation between average age and PV installations [109, 251]. However, other research indicates a negative relationship between age and PV adoption [252, 253].

Table 4.3: Regional error metrics for the absolute XGBoost models trained with different objectives and scaling approaches under the mixed hold-out configuration. Models were trained using either a Gaussian (squared-error) or Tweedie objective, with results reported both before and after national scaling. Metrics include the coefficient of determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), reported separately for training and test sets in megawatts (MW).

Model	Train R^2	Test R^2	Train MAE	Test MAE	Train MSE	Test MSE	Train RMSE	Test RMSE
Gaussian objective (unscaled)	1.00	0.92	12	142	297	90,149	17	300
Gaussian objective (scaled)	1.00	0.98	16	77	552	24,308	23	156
Tweedie objective (unscaled)	1.00	0.94	9	132	233	73,795	15	272
Tweedie objective (scaled)	1.00	0.99	12	53	422	10,501	21	102

The selected features for the models are shown in Appendix 4.6.2 and 4.6.3.

4.4.2 Model Performance Evaluation

Table 4.3 compares the absolute XGBoost models trained using Gaussian and Tweedie objectives under the mixed hold-out configuration. Across both scaled and unscaled variants, the Tweedie objective achieves higher test R^2 and lower RMSE, confirming its suitability for modelling skewed and zero-inflated capacity data. Therefore, the Tweedie objective was adopted for training all absolute models throughout this study.

Table 4.4 summarises the performance of the benchmark multiple linear regression (MLR) models trained on the full dataset without cross-validation. The normalized MLR model achieves substantially higher explanatory power than the absolute model ($R^2 = 0.86$ versus 0.50). Compared with the XGBoost models in Table 4.5, the linear models exhibit lower predictive accuracy. The MLR models yield R^2 values between 0.50 and 0.86, mean absolute errors (MAE) between 206 and 380 MW, and root mean squared errors (RMSE) between 331 and 542 MW. In contrast, the XGBoost models achieve higher R^2 values (0.93–0.99) and substantially lower errors, with MAE ranging from 53 to 132 MW and RMSE from 102 to 272 MW. These results confirm that XGBoost provides a more accurate and flexible framework for modelling regional solar PV capacity. Furthermore, the MLR models cannot accommodate scaling adjustments because they require complete regional coverage and do

Table 4.4: Regional error metrics for the benchmark multiple linear regression (MLR) models. The Absolute Model metrics are reported in megawatts (MW), while the Normalized Model metrics are reported in both megawatts (MW) and percentage points (%). Metrics include the coefficient of determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The MLR models were trained using the full dataset without cross-validation.

Model	R^2	MAE	MSE	RMSE
Absolute Model (MW)	0.50	380	294,040	542
Normalized Model (MW)	0.85	206	109,444	331
Normalized Model (%)	0.86	1.08	2.26	1.50

not handle missing data, limiting their applicability. Even when trained on the full dataset, the linear models perform worse than the XGBoost models evaluated on the test set under the mixed hold-out configuration, highlighting the superior generalisation capability of the latter.

Table 4.5 shows the regional error metrics for the absolute and normalized models under the mixed hold-out configuration. The normalized model outperforms the absolute model in estimating the regional capacity. However, this is not the case when scaling the results by the national capacity as the absolute models error reduces. Scaling the results by the national capacity reduces the error by more than 50% in the absolute model and doesn't impact the error in the normalized model. To further examine the robustness of these findings, we assess how the models generalise across space and time.

Tables 4.6 and 4.7 present results from the hold-out country validation, which assesses spatial generalisation by training the model on all but one country and testing on the excluded country. The absolute model shows considerable variation in performance across countries, with some negative test R^2 values, indicating limited transferability in predicting absolute capacity magnitudes. Applying national scaling improves consistency, increasing mean test R^2 from -0.28 to 0.15 . In contrast, the normalized models achieve more stable results across all countries, with mean test R^2 values of 0.2 – 0.4 . These results suggest that modelling regional capacity as a share of national totals reduces cross-country scale effects and improves the model's ability to generalise spatially.

Table 4.5: Regional error metrics for the XGBoost models under the mixed hold-out configuration, reported separately for training and test sets in both megawatts (MW) and percentage points (%). For the scaled models, regional predictions were adjusted to ensure that their sum matched the corresponding national total. Metrics include the coefficient of determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

Model	Train R^2	Test R^2	Train MAE	Test MAE	Train MSE	Test MSE	Train RMSE	Test RMSE
Absolute Model (MW)	1.00	0.94	9	132	233	73,795	15	272
Scaled Absolute Model (MW)	1.00	0.99	12	53	422	10,501	21	102
Normalized Model (MW)	0.99	0.95	48	107	6,722	61,426	82	248
Scaled Normalized Model (MW)	0.99	0.95	48	108	6,918	57,152	83	239
Normalized Model (%)	0.98	0.93	0.34	0.55	0.30	1.30	0.55	1.14
Scaled Normalized Model (%)	0.98	0.93	0.35	0.56	0.32	1.25	0.57	1.12

Model generalisation depends on the similarity between the test country and the countries included in training. Countries with land cover and climatic characteristics that differ substantially from the training set tend to exhibit higher errors, whereas those with similar feature distributions perform better. For this reason, the mixed hold-out configuration—where all countries contribute to model training—was adopted for applications to unseen regions, as it provides more robust and generalisable predictions.

Table 4.8 summarises the results of the hold-out year validation, which tests the models’ ability to generalise over time. All configurations show strong temporal predictive performance, with test MAE values ranging from 136 MW to 266 MW and RMSE values from 276 MW to 431 MW. Scaling regional predictions to match national totals further improved performance, particularly for the absolute model, where the test RMSE decreased from 431 MW to 276 MW. These results indicate that the XGBoost models capture stable temporal relationships between predictors and installed capacity, maintaining robustness when applied to future years.

Figure 4.2a compares the actual and predicted regional capacities for the absolute model under the mixed hold-out configuration. The model systematically underestimates regions with high installed capacity, suggesting limited extrapolation ability at the upper end of the distribution. However, as shown in Figure 4.2b, applying national scaling substantially

Table 4.6: Regional error metrics for the XGBoost absolute model (MW) under the hold-out country configuration. For each iteration, the model was trained on all countries except one, which was used for testing. Metrics are reported separately for training and test sets.

Country	Train R^2	Test R^2	Train MAE	Test MAE	Train MSE	Test MSE	Train RMSE	Test RMSE
France (unscaled)	0.98	-1.07	59	532	11,644	454,794	108	674
Spain (unscaled)	0.86	0.06	179	658	87,323	1,413,287	296	1,189
Germany (unscaled)	0.96	-0.06	71	694	16,451	929,653	128	964
Italy (unscaled)	0.84	0.29	185	429	102,011	474,864	319	689
United Kingdom (unscaled)	1.00	0.03	0.33	187	0.33	83,912	0.57	290
Belgium (unscaled)	0.84	-0.95	190	400	111,172	281,340	333	530
Mean (unscaled)	0.91	-0.28	114	483	54,434	606,975	197	723
France (scaled)	0.99	-0.23	49	374	7,067	269,358	84	519
Spain (scaled)	0.90	0.01	158	741	61,615	1,488,481	248	1,220
Germany (scaled)	0.98	0.08	63	601	10,632	805,911	103	898
Italy (scaled)	0.88	0.54	169	414	77,305	306,712	278	554
United Kingdom (scaled)	1.00	0.05	5	179	148	82,327	12	287
Belgium (scaled)	0.88	0.47	172	222	81,142	76,477	285	277
Mean (scaled)	0.94	0.15	103	422	39,318	504,878	168	626

Table 4.7: Regional error metrics for the XGBoost normalized model (MW) under the hold-out country configuration. For each iteration, the model was trained on all countries except one, which was used for testing. Metrics are reported separately for training and test sets.

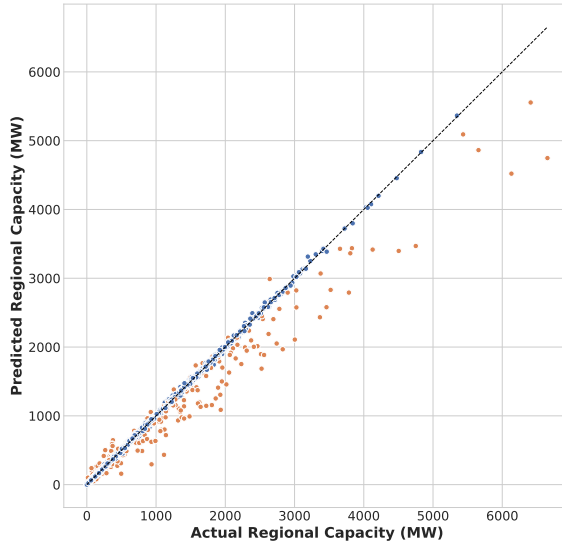
Country	Train R^2	Test R^2	Train MAE	Test MAE	Train MSE	Test MSE	Train RMSE	Test RMSE
France (unscaled)	1.00	0.31	0.83	229	26	150,919	5	388
Spain (unscaled)	0.99	0.29	42	512	6,444	1,059,578	80	1,029
Germany (unscaled)	1.00	-0.65	3	879	154	1,445,534	12	1,202
Italy (unscaled)	0.98	0.21	64	486	14,242	530,604	119	728
United Kingdom (unscaled)	0.97	0.20	82	161	22,032	69,695	148	264
Belgium (unscaled)	0.97	0.35	79	234	20,058	94,023	142	307
Mean (unscaled)	0.99	0.12	45	417	10,493	558,392	84	653
France (scaled)	1.00	0.33	7	243	172	147,314	13	384
Spain (scaled)	0.99	0.58	42	431	6,405	623,377	80	790
Germany (scaled)	1.00	-0.05	12	709	435	918,846	21	959
Italy (scaled)	0.98	0.62	65	400	15,652	251,759	125	502
United Kingdom (scaled)	0.97	0.14	82	151	22,393	74,360	150	273
Belgium (scaled)	0.97	0.43	79	223	20,027	82,889	142	288
Mean (scaled)	0.99	0.34	48	360	10,847	349,758	89	533

Table 4.8: Regional error metrics for the XGBoost models under the hold-out year configuration, reported separately for training and test sets in both megawatts (MW) and percentage points (%). The models were trained on data from 2010–2020 and tested on 2021–2023. Metrics include the coefficient of determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

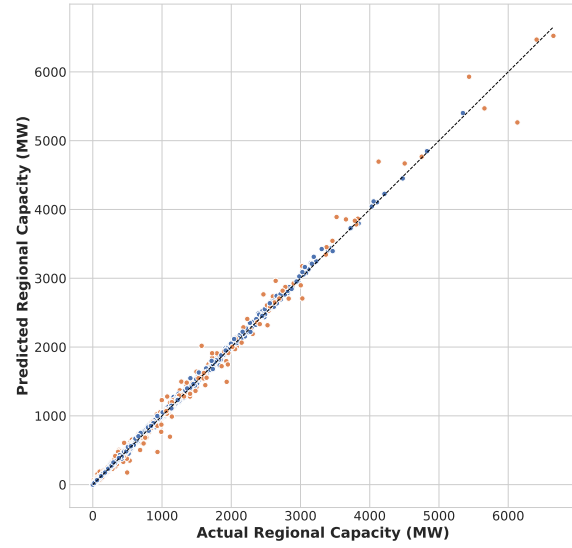
Model	Train R^2	Test R^2	Train MAE	Test MAE	Train MSE	Test MSE	Train RMSE	Test RMSE
Absolute Model (MW)	1.00	0.86	7	266	133	185,517	12	431
Scaled Absolute Model (MW)	1.00	0.94	10	136	269	76,297	16	276
Normalized Model (MW)	0.99	0.92	42	157	4,834	107,434	70	328
Scaled Normalized Model (MW)	0.99	0.92	41	157	4,824	104,709	69	324
Normalized Model (%)	0.98	0.91	0.33	0.64	0.27	1.89	0.52	1.38
Scaled Normalized Model (%)	0.98	0.91	0.33	0.64	0.28	1.83	0.53	1.35

corrects this bias. Figure 4.2c and Figure 4.2d show the corresponding results for the normalized model before and after scaling. In this case, model performance is strong and largely unaffected by scaling, indicating that the normalized representation is inherently more stable across regions.

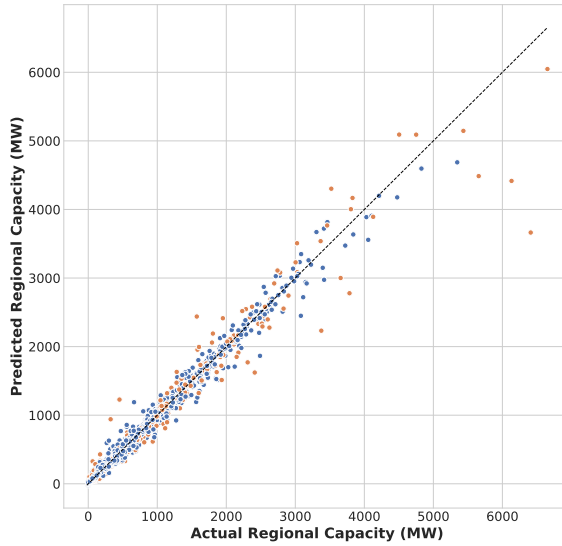
To evaluate model consistency at the national level, the mixed hold-out predictions are aggregated and compared with the observed national totals. Table 4.9 presents the national-level error metrics for both the absolute and normalized XGBoost models under the mixed hold-out configuration. For the training countries, the normalized model achieves a MAPE of 2.5%, while the absolute model performs slightly worse, with a MAPE of 6.4%. However, when the models are applied to all European countries, the error for the normalized model increases to 19.5%, whereas the absolute model yields an unrealistically large value of $5.5 \times 10^{15}\%$. This behaviour is a known artefact of the MAPE formula, which divides the absolute error by the actual value. When the actual national PV capacity is zero or close to zero, the denominator becomes very small. To prevent division by zero, `scikit-learn` replaces zeros with machine epsilon (ϵ), the smallest positive number that can be added to 1.0 such that the result is distinguishable from 1.0 in floating-point arithmetic (approximately 2.22×10^{-16}) [254]. As a result, even minor absolute errors can inflate into extremely large percentage errors. Excluding countries with zero observed capacity reduces the MAPE



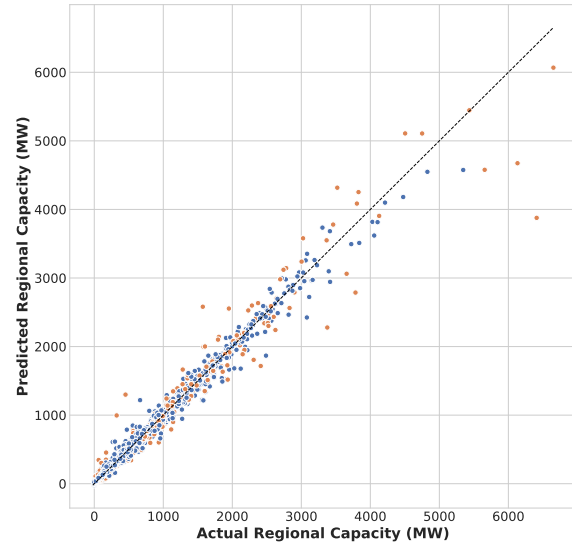
(a) Absolute model.



(b) Absolute model after scaling.



(c) Normalized model.



(d) Normalized model after scaling.

Figure 4.2: Actual versus predicted regional photovoltaic capacity for 150 NUTS 2 regions using the absolute and normalized XGBoost models under the mixed hold-out configuration (2010–2023). Blue points represent training data and orange points represent test data. The actual capacity data are sourced from the United Kingdom, Italy, Spain, Belgium, Germany, and France.

Table 4.9: National-level error metrics for the absolute and normalized XGBoost models under the mixed hold-out configuration. Metrics compare the actual national capacity with the estimated national capacity, obtained by summing predicted regional values. Results are shown for both the training countries and the full European dataset. Training countries include the United Kingdom, Italy, Spain, Belgium, Germany, and France. Scaling refers to adjusting regional capacities so that their totals match the observed national capacity.

Model	R^2	MAE (MW)	MSE (MW ²)	RMSE (MW)	MAPE (%)
Absolute model (training)	0.97	964	1.04×10^7	3233	6.4
Absolute model (scaled, training)	1.00	0	0	0	0.0
Absolute model (entire dataset)	0.95	1359	1.09×10^7	3313	5.5×10^{15}
Absolute model (scaled, entire dataset)	1.00	0	0	0	0.0
Normalized model (training)	1.00	335	2.10×10^5	459	2.5
Normalized model (scaled, training)	1.00	0	0	0	0.0
Normalized model (entire dataset)	1.00	331	2.67×10^5	517	19.5
Normalized model (scaled, entire dataset)	1.00	0	0	0	0.0

for the absolute model to 58%. Given this sensitivity to small denominators, we conclude that the absolute model does not generalise reliably beyond the training countries. Therefore, the normalized model is used for disaggregating national capacity in countries lacking regional data. Nonetheless, because the absolute model performs well within the training set, it remains useful for allocating capacity that is not associated with a specific geographic location.

4.4.3 Model Interpretation and Applications

Feature Importance

The feature importance analysis based on SHAP values for the mixed hold-out models highlights differences in how features contribute to the model predictions in the absolute and normalized settings. In the absolute model, artificial surfaces have the most significant impact, accounting for 55% of the total SHAP values. This is followed by agricultural land, contributing 18%, national PV capacity at 17%, forests at 6%, and median age at 3%. In the normalized model, agricultural land emerges as the most influential factor, representing 41% of the total SHAP values. Artificial surfaces contribute 19%, with water bodies at 15%,

forests and seminatural areas at 11%, and GHI at 10%, with all other factors contributing less than 5%. A full breakdown of SHAP values and their corresponding percentage contributions for both models is provided in Appendix 4.6.4. Water bodies being the 3rd most important feature in the normalized model is unexpected. While there has been a rapid rise in floating solar PV installations, with Europe having 451 MW of grid-connected floating solar PVs as of 2022—280 MW in the Netherlands, 80 MW in France, 25 MW each in Austria and Germany, 10 MW in Spain, and 6.2 MW in Portugal [255]—this alone does not fully explain why water bodies rank so high. A possible explanation lies in the strong Spearman correlations observed between water bodies and agricultural areas (0.65), and between water bodies and artificial surfaces (0.54), both of which are highly correlated with the percentage of solar PV capacity.

Regional Capacity Estimation and Disaggregation

Figure 4.1 illustrates the actual regional capacities for the regions used in training and testing the models for 2023. Figure 4.3b presents the estimated regional capacities for countries lacking regional data in 2023, generated using the scaled normalized model trained under the mixed hold-out configuration. These estimates may support efforts to monitor and model PV generation. However, estimating or forecasting electricity generation from solar PV is beyond the scope of this study. Our focus is on modelling installed capacity using open, accessible, and largely complete data sources. This provides a foundation for others to incorporate additional factors such as weather variability and anthropogenic pollution when modelling actual generation. In addition, the regional capacity estimates may serve as valuable inputs to broader sustainability modelling efforts, including assessments of energy self-sufficiency, land-use impacts, and infrastructure planning.

In some cases, the locations of installed PV systems are not recorded. In 2023, unlocated systems accounted for 112 MW in Germany, 568 MW in Spain, 227 MW in France, 676 MW in Italy, and 820 MW in the United Kingdom. The scaled absolute model trained under the mixed hold-out configuration was used to spatially allocate these capacities, resulting in updated regional capacity estimates, as shown in Figure 4.3a.

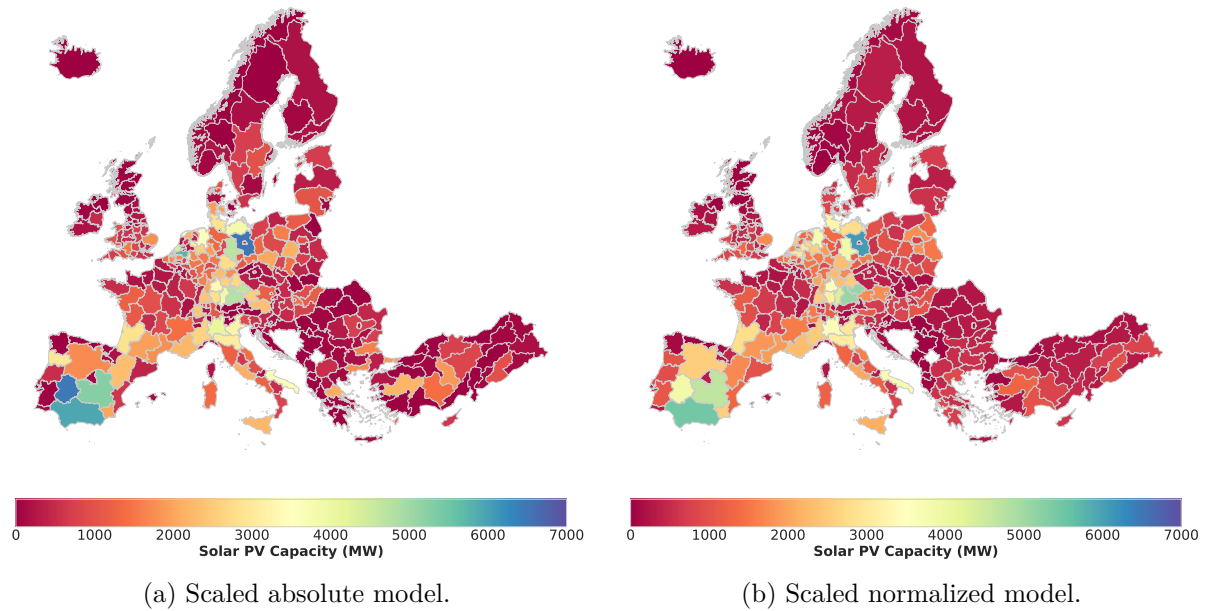


Figure 4.3: Predicted regional solar photovoltaic capacity for 36 European countries and 333 NUTS 2 regions in 2023 using the scaled XGBoost models trained under the mixed hold-out configuration. The corresponding regional capacity values are provided in Appendix 4.6.5. (a) The scaled absolute model is more effective for allocating unknown national capacity to geographic locations in the training countries, which include the United Kingdom, Italy, Spain, Belgium, Germany, and France. (b) The scaled normalized model is better suited for estimating PV capacity in countries where regional data is unavailable. Although the model predicts capacity as a percentage of the national total (i.e., normalized output), these percentages are subsequently multiplied by national PV capacity to produce final capacity estimates in megawatts.

Benchmarking Regional Deployment

Another application of the mixed hold-out models is as benchmarking tools. The absolute model is used to assess and compare regional capacities within the countries where the model was trained, as illustrated in Figure 4.4a. This approach allows for evaluating how different regions perform relative to similar regions. For instance, regions in Germany and Spain often exhibit higher PV capacity than expected when compared to regions in Italy, the United Kingdom, France, and Belgium. In particular, Brandenburg (DE40) has 1,905 MW more capacity than expected for comparable European regions, followed by Castilla-La Mancha (ES42) with 1,612 MW, Oberbayern (DE21) with 1,282 MW, Niederbayern (DE22) with 1,107 MW, Mecklenburg-Vorpommern (DE80) with 993 MW, Stuttgart (DE11) with 933

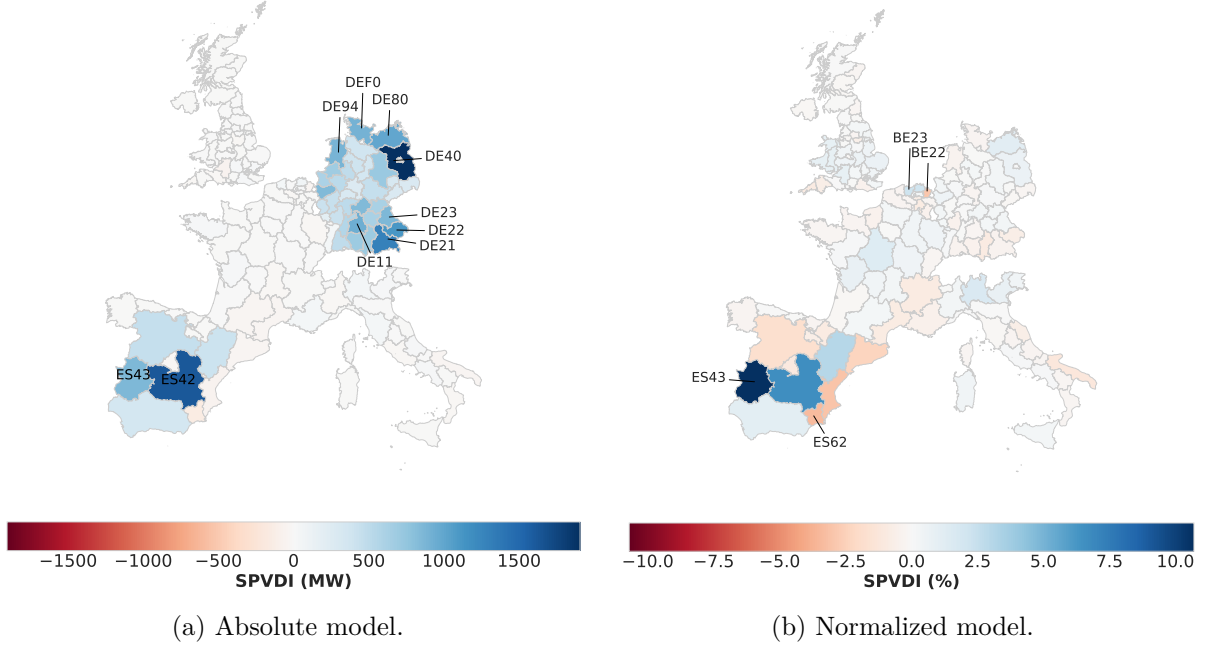


Figure 4.4: The solar PV deployment index results for the year 2023 derived from the XG-Boost models trained under the mixed hold-out configuration. Positive values indicate regions where the actual capacity exceeds the expected capacity, while negative values indicate regions where the actual capacity is less than expected. (a) The absolute model benchmarks deployment compared to similar regions in Europe. (b) The normalized model benchmarks deployment compared to similar regions within a country. The regions included are Brandenburg (DE40), Castilla-La Mancha (ES42), Oberbayern (DE21), Niederbayern (DE22), Mecklenburg-Vorpommern (DE80), Stuttgart (DE11), Schleswig-Holstein (DEF0), Weser-Ems (DE94), Oberpfalz (DE23), Extremadura (ES43), Murcia (ES62), East Flanders (BE23), and Limburg (BE22).

MW, Schleswig-Holstein (DEF0) with 895 MW, Weser-Ems (DE94) with 882 MW, Oberpfalz (DE23) with 862 MW, and Extremadura (ES43) with 855 MW.

Brandenburg’s agricultural landscape makes it appealing to solar PV investors seeking large tracts of land for solar farms. The region’s vast areas of suitable land combined with rising land prices have created competition between agricultural and energy sectors. The lack of strict regulations limiting the size of solar farms exacerbates this issue, allowing large-scale solar projects to dominate the landscape [96].

The normalized model is employed to benchmark capacity within individual countries, comparing regional performance relative to other regions within the same country, as shown

in Figure 4.4b. For example, in Spain in 2023, the Extremadura region (ES43) exceeded expectations for similar regions by 2,747 MW (10.7%), while the Region of Murcia (ES62) fell short by 866 MW (3.4%). The region of Extremadura has its own Integrated Energy and Climate Plan (Plan Extremeño Integrado de Energía y Clima, PEIEC) for the 2021-2030 period. The plan aims for Extremadura to become 100% powered by renewable energy by 2030, with a target of 10.36 GW of renewable capacity, of which 8 GW is allocated to solar PV [256]. Curtailment in Spain is unevenly distributed across regions, with higher curtailment occurring in regions where renewable generation capacity exceeds local demand and where transmission infrastructure is insufficient to transfer excess power to other regions. This situation puts some generators at financial risk, as they are increasingly instructed to reduce output without compensation [257]. Such curtailment could help explain the differences in solar PV deployment between regions.

In Belgium, East Flanders (BE23) exhibits a PV capacity that is 196 MW (2.2%) higher than expected, while Limburg (BE22) shows a deficit of 260 MW (3%) compared to expectations. Both regions are in Flanders, where residential solar PV has gained popularity due to generous support policies [258]. For instance, the Flemish social housing sector launched Access to Sustainability for Tenants through Energy Effective Retrofit (ASTER), an initiative focused on improving the sustainability of energy and insulation in social housing. ASTER supports this by purchasing solar panels and providing energy to benefit Flemish social housing providers and their tenants [259, 260]. The notable difference in PV capacity between these two regions could be due to several factors, including easier grid connections, the presence of enthusiastic investors who own land, or the influence of individual large-scale solar installations.

Forecasting Future Capacity

The scaled models also serve as valuable tools for forecasting future PV capacity distribution across regions. By applying these models, one can predict where PV capacity is likely to expand, which is crucial for strategic grid planning and policy development. As detailed in

Appendix 4.6.6, the scaled normalized model demonstrates strong forecasting performance when evaluated on historical data (2017–2023), providing confidence in its application for forward projections. To demonstrate its practical use, we apply the scaled normalized model trained under the mixed hold-out configuration to forecast the spatial distribution of solar PV capacity in Germany for the year 2030, based on the national target of 215 GW [261]. Figure 4.5 shows how this capacity is likely to be allocated across regions, using 2018 CORINE land cover data and assuming climatic conditions consistent with 2023. Brandenburg is forecasted to have the highest share of installed capacity, reaching 15,725 MW by 2030. This result is particularly relevant given the region’s recent socio-economic dynamics. In Brandenburg, tensions between farmers and solar PV developers have been fuelled by a sharp increase in land prices, particularly since 2007. Between 2010 and 2020, rental prices for agricultural land surged by 62% [262]. This, coupled with enterprise groups controlling 52% of agricultural land [262], has made it more difficult for smaller farmers to compete financially with large solar energy developers. The resulting land acquisition by solar investors has contributed to a growing sense of distrust toward large-scale solar PV projects, which may hinder future deployment in the absence of a proper legal and policy framework [96]. By forecasting where solar PV systems are likely to be installed, our model can inform policy interventions aimed at mitigating land use conflicts. Such interventions may include the promotion of agrivoltaics, which integrate solar energy generation with agricultural activity, offering a more balanced approach to land use in agricultural areas.

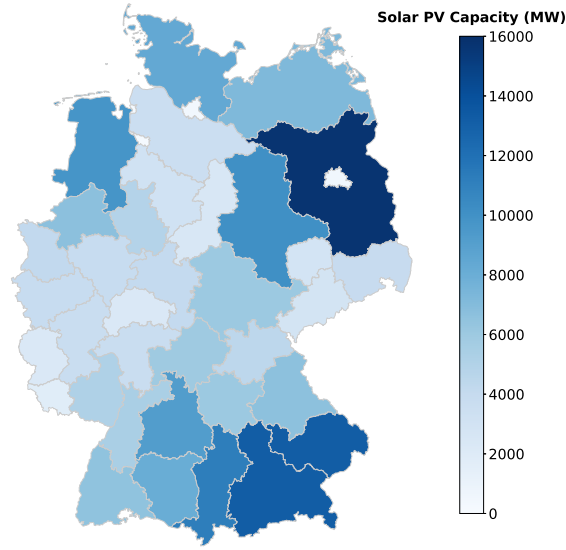


Figure 4.5: Forecasted regional solar PV capacity in Germany for 2030, based on the national target of 215 GW [261]. The distribution across regions is estimated using the scaled normalized XGBoost model trained under the mixed hold-out configuration, which allocates national capacity proportionally according to regional characteristics. The forecast is based on 2018 CORINE land-cover data, with climatic variable values set to those observed in 2023.

4.4.4 Limitations

This study makes several assumptions and is subject to a number of limitations and uncertainties. First, the models assume that the geographic, socio-economic, and climatic factors used as input features are both representative and sufficiently explanatory of regional solar PV capacity. While these factors were selected based on empirical correlations and data availability, their influence may vary across national contexts and over time. Second, the accuracy of the model outputs depends on the reliability of the underlying data sources. In particular, the CORINE land cover dataset, although widely used, has known thematic accuracy limitations that could lead to misclassification of land use types. To mitigate this, we incorporate multiple classification levels (Levels 1–3) to balance detail with robustness. Third, the models were trained and validated using data from a subset of countries with relatively complete regional capacity information. This introduces uncertainty when generalising predictions to countries with missing or incomplete data, especially in the absolute model,

which was shown to be sensitive to near-zero denominators in percentage error metrics. The use of machine epsilon to handle division by zero may inflate model error in such cases. Finally, while the models are effective for disaggregation, benchmarking, and forecasting, they do not account for dynamic policy changes, market conditions, or future shifts in deployment strategies. As such, their predictive accuracy may diminish when applied far outside the historical or geographic range of the training data.

4.5 Conclusion

This study provides a comprehensive analysis of regional solar photovoltaic (PV) capacity across Europe, emphasising the critical role of geographical factors in determining deployment. The findings reveal that artificial surfaces and agricultural areas are key predictors of regional solar PV deployment. Our models, designed to estimate regional PV capacity, achieve a root mean squared error of less than 272 MW and explain over 93% of the variation across 150 NUTS 2 EU regions. These models are subsequently applied to estimate the capacity of 333 NUTS EU regions for the years 2010 to 2023.

The models serve three main purposes. First, the disaggregation of national PV capacity, where the scaled models (both absolute and normalized) are the most appropriate. Second, benchmarking of regional capacities both within and across countries, where the absolute model is used for comparisons across countries, while the normalized model is used for comparing regions within a country. Finally, for forecasting future PV capacity distribution, the scaled models are again the most suitable.

The insights gained from this research are particularly valuable for informing policy-making and grid management. By addressing the gaps in regional capacity data and offering a robust method for estimating PV deployment, we can better anticipate land use conflicts, especially in areas where competition between agricultural and energy sectors is intense. Policymakers can use these models to craft targeted interventions that promote solar PV expansion while balancing agricultural needs. Additionally, grid operators can use these models to forecast potential areas of capacity expansion, aiding in effective planning for grid

development and extension. These models also provide a foundation for developing enhanced solar PV monitoring services, ensuring a more reliable integration of renewable energy into the grid.

Data Availability

All datasets and source code used in this study are publicly available via Zenodo at the following DOI: <https://doi.org/10.5281/zenodo.15366956>.

4.6 Appendix

4.6.1 Full Feature Analysis Tables

Table 4.10: Absolute data analysis. Features considered for modelling regional solar photovoltaic capacity. The data availability, coefficient of determination (R^2), Pearson correlation, Spearman correlation, correlation average are shown. Relevant literature that explores similar features is cited. Land cover features are preceded by their CORINE Land Cover (CLC) classification codes (e.g., 1.2.1 for industrial or commercial units).

Feature	Data Availabil- ity (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
National solar PV capacity (MWp)	100	0.31	0.56***	0.63***	0.59	
1.2.1 Industrial or commercial units (m ²)	98	0.21	0.46***	0.55***	0.51	
1.2 Industrial, commercial and transport units (m ²)	98	0.18	0.43***	0.52***	0.47	
1 Artificial Surfaces (m ²)	98	0.15	0.38***	0.51***	0.45	
1.1 Urban fabric (m ²)	98	0.13	0.37***	0.52***	0.45	
2.1 Arable land (m ²)	98	0.13	0.36***	0.54***	0.45	[249, 119]
2.1.1 Non-irrigated arable land (m ²)	98	0.12	0.35***	0.53***	0.44	[249, 119]
3.1.2 Coniferous forest (m ²)	98	0.17	0.41***	0.46***	0.43	
Median age (years)	92	0.15	0.38***	0.46***	0.42	[109, 252, 253, 251]
1.1.2 Discontinuous urban fabric (m ²)	98	0.11	0.33***	0.5***	0.42	
2 Agricultural Areas (m ²)	98	0.12	0.35***	0.47***	0.41	
5 Water Bodies (m ²)	98	0.14	0.38***	0.41***	0.40	[255]

Feature	Data Availabil- ity (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
5.1.1 Water courses (m ²)	98	0.13	0.36***	0.43***	0.40	
5.1 Inland waters (m ²)	98	0.15	0.38***	0.41***	0.40	
2.2 Permanent crops (m ²)	98	0.08	0.28***	0.51***	0.40	[119]
3.1 Forests (m ²)	98	0.11	0.32***	0.47***	0.40	[250]
Employment (thousand persons)	86	0.10	0.32***	0.45***	0.39	
1.3.1 Mineral extraction sites (m ²)	98	0.15	0.39***	0.38***	0.39	
3.1.3 Mixed forest (m ²)	98	0.08	0.29***	0.49***	0.39	
1.3.2 Dump sites (m ²)	98	0.11	0.33***	0.43***	0.38	
1.3 Mine, dump and construction sites (m ²)	98	0.14	0.37***	0.38***	0.38	
Disposable income (million euros)	64	0.09	0.29***	0.47***	0.38	
Region area (m ²)	98	0.11	0.33***	0.43***	0.38	
2.2.2 Fruit trees and berry plantations (m ²)	98	0.02	0.16***	0.58***	0.37	[119]
Population	67	0.08	0.29***	0.43***	0.36	
5.1.2 Water bodies (m ²)	98	0.12	0.35***	0.35***	0.35	
1.1.1 Continuous urban fabric (m ²)	98	0.07	0.27***	0.38***	0.33	
GDP (million euros)	81	0.04	0.21***	0.45***	0.33	
Wind speed (m s ⁻¹)	93	0.09	-0.3***	-0.37***	-0.33	
GERD (million euros)	51	0.07	0.26***	0.38***	0.32	
1.2.4 Airports (m ²)	98	0.08	0.29***	0.34***	0.32	
3 Forest And Seminatural Areas (m ²)	98	0.07	0.26***	0.36***	0.31	
GHI (W m ⁻²)	93	0.07	0.25***	0.37***	0.31	

Feature	Data Availabil- ity (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
Mean sea level pressure (hPa)	93	0.06	0.24***	0.33***	0.29	
Year	100	0.08	0.29***	0.29***	0.29	
2.2.1 Vineyards (m ²)	98	0.05	0.23***	0.34***	0.29	[119]
Disposable income per inhabitant (euro)	64	0.07	0.26***	0.32***	0.29	
3.1.1 Broad-leaved forest (m ²)	98	0.02	0.15***	0.41***	0.28	
2.4 Heterogeneous agricultural areas (m ²)	98	0.05	0.21***	0.31***	0.26	
Electricity consumption (billion kWh)	75	0.06	0.25***	0.23***	0.24	
2.1.3 Rice fields (m ²)	98	0.05	0.23***	0.26***	0.24	[119]
3.3.1 Beaches, dunes, sands (m ²)	98	0.04	0.21***	0.26***	0.23	
3.2.4 Transitional woodland-scrub (m ²)	98	0.05	0.22***	0.23***	0.23	
4.1.1 Inland marshes (m ²)	98	0.04	0.2***	0.25***	0.23	
2.4.3 Land principally occupied by agriculture, with significant areas of natural vegetation (m ²)	98	0.04	0.2***	0.26***	0.23	[119]
2.4.2 Complex cultivation patterns (m ²)	98	0.01	0.09***	0.37***	0.23	[119]
2.2.3 Olive groves (m ²)	98	0.05	0.23***	0.22***	0.23	[119]
Air temperature (K)	93	0.04	0.21***	0.23***	0.22	
4.2.2 Salines (m ²)	98	0.04	0.21***	0.23***	0.22	
4.2.3 Intertidal flats (m ²)	98	0.01	-0.12***	-0.3***	-0.21	
2.4.4 Agro-forestry areas (m ²)	98	0.04	0.21***	0.19***	0.20	
2.4.1 Annual crops associated with permanent crops (m ²)	98	0.01	0.1***	0.25***	0.17	[119]
3.2.3 Sclerophyllous vegetation (m ²)	98	0.03	0.17***	0.16***	0.17	

Feature	Data Availabil- ity (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
3.2.1 Natural grassland (m ²)	98	0.02	0.16***	0.17***	0.17	[253]
Unemployment (thousand persons)	85	0.02	0.14***	0.18***	0.16	
5.2.1 Coastal lagoons (m ²)	98	0.02	0.15***	0.15***	0.15	
3.3.4 Burnt areas (m ²)	98	0.02	0.16***	0.14***	0.15	[119]
1.2.2 Road and rail networks and associated land (m ²)	98	0.02	0.12***	0.19***	0.15	
3.3 Open spaces with little or no vegetation (m ²)	98	0.03	0.17***	0.1***	0.14	
2.1.2 Permanently irrigated land (m ²)	98	0.03	0.16***	0.11***	0.14	[119]
3.2.2 Moors and heathland (m ²)	98	0.02	-0.13***	-0.14***	-0.14	
3.2 Scrub and/or herbaceous associations (m ²)	98	0.02	0.14***	0.12***	0.13	
Country area (m ²)	100	0.00	0.02	0.24***	0.13	
5.2.2 Estuaries (m ²)	98	0.00	0.04	-0.3***	-0.13	
3.3.2 Bare rocks (m ²)	98	0.02	0.14***	0.09***	0.12	
3.3.5 Glaciers and perpetual snow (m ²)	98	0.00	0.04	0.2***	0.12	
2.3.1 Pastures (m ²)	98	0.00	0.06**	0.17***	0.12	
2.3 Pastures (m ²)	98	0.00	0.06**	0.17***	0.12	
4.1.2 Peat bogs (m ²)	98	0.01	-0.09***	-0.14***	-0.12	
3.3.3 Sparsely vegetated areas (m ²)	98	0.02	0.15***	0.07**	0.11	
1.3.3 Construction sites (m ²)	98	0.02	0.13***	0.06*	0.10	
4 Wetlands (m ²)	98	0.01	-0.08***	-0.06*	-0.07	
1.4.1 Green urban areas (m ²)	98	0.01	-0.08***	-0.06**	-0.07	
4.2.1 Salt marshes (m ²)	98	0.03	0.18***	-0.06**	0.06	

Feature	Data Availabil- ity (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
4.1 Inland wetlands (m ²)	98	0.01	-0.09***	-0.03	-0.06	
4.2 Marine wetlands (m ²)	98	0.03	0.17***	-0.1***	0.04	
5.2 Marine waters (m ²)	98	0.02	0.13***	-0.07**	0.03	
Total precipitation (m)	98	0.00	0.06*	0.01	0.03	
1.4.2. Sport and leisure facilities (m ²)	98	0.00	0.04	0.02	0.03	
R&D personnel and researchers (percentage)	47	0.00	-0.01	0.06	0.02	
5.2.3 Sea and ocean (m ²)	98	0.00	0.0	-0.04	-0.02	
1.4 Artificial, non-agricultural vegetated areas (m ²)	98	0.00	0.01	-0.03	-0.01	
1.2.3 Port areas (m ²)	98	0.00	-0.0	-0.0	-0.00	

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 4.11: Normalized data analysis. Features considered for modelling the percentage of regional solar photovoltaic capacity. All features presented are normalized relative to national values and expressed as percentages. The data availability, coefficient of determination (R^2), Pearson correlation, Spearman correlation, correlation average are shown. Relevant literature that explores similar features is cited. Land cover features are preceded by their CORINE Land Cover (CLC) classification codes (e.g., 1.2.1 for industrial or commercial units).

Feature	Data Availabil- ity (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
2 Agricultural Areas	98	0.49	0.7***	0.67***	0.69	

Feature	Data Availabil- ity (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
Region area	98	0.38	0.61***	0.63***	0.62	
2.1 Arable land	98	0.32	0.57***	0.63***	0.60	[249]
2.4 Heterogeneous agricultural areas	98	0.41	0.64***	0.52***	0.58	
2.1.1 Non-irrigated arable land	98	0.28	0.53***	0.62***	0.57	[249]
1.2 Industrial, commercial and transport units	98	0.31	0.56***	0.52***	0.54	
1.2.1 Industrial or commercial units	98	0.31	0.56***	0.51***	0.54	
2.2 Permanent crops	98	0.29	0.54***	0.52***	0.53	[119]
Disposable income	64	0.22	0.47***	0.56***	0.52	
2.4.2 Complex cultivation patterns	98	0.27	0.52***	0.52***	0.52	[119]
Employment	86	0.24	0.49***	0.53***	0.51	
Population	67	0.25	0.5***	0.52***	0.51	
1 Artificial Surfaces	98	0.22	0.47***	0.54***	0.51	
1.3 Mine, dump and construction sites	98	0.23	0.48***	0.52***	0.50	
1.1 Urban fabric	98	0.20	0.45***	0.55***	0.50	
2.4.3 Land principally occupied by agriculture, with significant areas of natural vegetation	98	0.24	0.49***	0.49***	0.49	[119]
5.1.1 Water courses	98	0.32	0.57***	0.4***	0.48	
1.1.2 Discontinuous urban fabric	98	0.18	0.43***	0.51***	0.47	
1.2.4 Airports	98	0.21	0.45***	0.5***	0.47	
1.3.1 Mineral extraction sites	98	0.16	0.4***	0.52***	0.46	
2.2.2 Fruit trees and berry plantations	98	0.12	0.34***	0.58***	0.46	[119]

Feature	Data Availabil- ity (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
2.2.1 Vineyards	98	0.27	0.52***	0.38***	0.45	[119]
GDP	81	0.12	0.35***	0.52***	0.43	
GHI	93	0.14	0.37***	0.48***	0.42	
3.1 Forests	98	0.15	0.39***	0.45***	0.42	[250]
Unemployment	85	0.17	0.41***	0.42***	0.41	[253]
5 Water Bodies	98	0.13	0.35***	0.47***	0.41	[255]
3 Forest And Seminatural Areas	98	0.18	0.42***	0.38***	0.40	
4.1.1 Inland marshes	98	0.11	0.33***	0.45***	0.39	
3.1.3 Mixed forest	98	0.06	0.25***	0.53***	0.39	
1.2.2 Road and rail networks and associated land	98	0.20	0.45***	0.33***	0.39	
4.2.2 Salines	98	0.20	0.44***	0.33***	0.39	
5.1 Inland waters	98	0.10	0.32***	0.43***	0.38	
3.2.4 Transitional woodland-scrub	98	0.23	0.48***	0.28***	0.38	
3.1.1 Broad-leaved forest	98	0.11	0.32***	0.44***	0.38	
1.3.3 Construction sites	98	0.23	0.48***	0.27***	0.38	
2.2.3 Olive groves	98	0.18	0.43***	0.31***	0.37	[119]
3.3.1 Beaches, dunes, sands	98	0.08	0.28***	0.44***	0.36	
1.1.1 Continuous urban fabric	98	0.16	0.4***	0.3***	0.35	
2.4.4 Agro-forestry areas	98	0.16	0.4***	0.3***	0.35	
1.3.2 Dump sites	98	0.14	0.37***	0.32***	0.34	
Air temperature	93	0.09	0.3***	0.36***	0.33	

Feature	Data Availabil- ity (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
5.1.2 Water bodies	98	0.06	0.25***	0.4***	0.33	[119]
2.1.3 Rice fields	98	0.08	0.27***	0.38***	0.33	
3.2.3 Sclerophyllous vegetation	98	0.13	0.36***	0.27***	0.32	
3.3 Open spaces with little or no vegetation	98	0.08	0.28***	0.31***	0.30	
3.2 Scrub and/or herbaceous associations	98	0.10	0.31***	0.3***	0.30	[119]
2.1.2 Permanently irrigated land	98	0.16	0.4***	0.2***	0.30	
5.2.1 Coastal lagoons	98	0.11	0.33***	0.26***	0.30	
3.2.1 Natural grassland	98	0.11	0.33***	0.27***	0.30	
3.1.2 Coniferous forest	98	0.08	0.29***	0.32***	0.30	
3.3.2 Bare rocks	98	0.09	0.3***	0.27***	0.29	
4.2.1 Salt marshes	98	0.17	0.41***	0.14***	0.28	
GERD	51	0.02	0.14***	0.36***	0.25	
4.2 Marine wetlands	98	0.15	0.39***	0.11***	0.25	
3.3.4 Burnt areas	98	0.07	0.26***	0.23***	0.24	
3.3.3 Sparsely vegetated areas	98	0.05	0.23***	0.19***	0.21	
1.2.3 Port areas	98	0.11	0.33***	0.09***	0.21	
2.4.1 Annual crops associated with permanent crops	98	0.01	0.11***	0.31***	0.21	[119]
5.2 Marine waters	98	0.09	0.3***	0.09***	0.20	
Mean sea level pressure	93	0.01	0.12***	0.24***	0.18	
1.4.2 Sport and leisure facilities	98	0.05	0.22***	0.15***	0.18	
3.3.5 Glaciers and perpetual snow	98	0.01	0.11***	0.26***	0.18	

Feature	Data Availabil- ity (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
1.4 Artificial, non-agricultural vegetated areas	98	0.04	0.19***	0.1***	0.15	
2.3.1 Pastures	98	0.00	0.05*	0.23***	0.14	
2.3 Pastures	98	0.00	0.05*	0.23***	0.14	
4.1.2 Peat bogs	98	0.01	-0.1***	-0.17***	-0.14	
Tertiary education	85	0.01	-0.1***	-0.16***	-0.13	
Median age	92	0.01	0.11***	0.1***	0.11	[109, 252, 253, 251]
4.2.3 Intertidal flats	98	0.01	-0.08***	-0.13***	-0.11	
5.2.3 Sea and ocean	98	0.00	0.02	0.13***	0.07	
Wind speed percentage	93	0.00	-0.06**	-0.07**	-0.07	
Disposable income per inhabitant	64	0.00	-0.01	0.1***	0.05	
R&D personnel and researchers	47	0.00	-0.07*	-0.02	-0.05	
5.2.2 Estuaries	98	0.05	0.23***	-0.15***	0.04	
Total precipitation	97	0.00	-0.01	-0.07**	-0.04	
4 Wetlands	98	0.01	-0.09***	0.15***	0.03	
4.1 Inland wetlands	98	0.01	-0.1***	0.14***	0.02	
3.2.2 Moors and heathland	98	0.01	-0.1***	0.12***	0.01	
1.4.1 Green urban areas	98	0.00	-0.0	-0.0	-0.00	

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

4.6.2 Input Features and Predicted Outputs of the Absolute Model

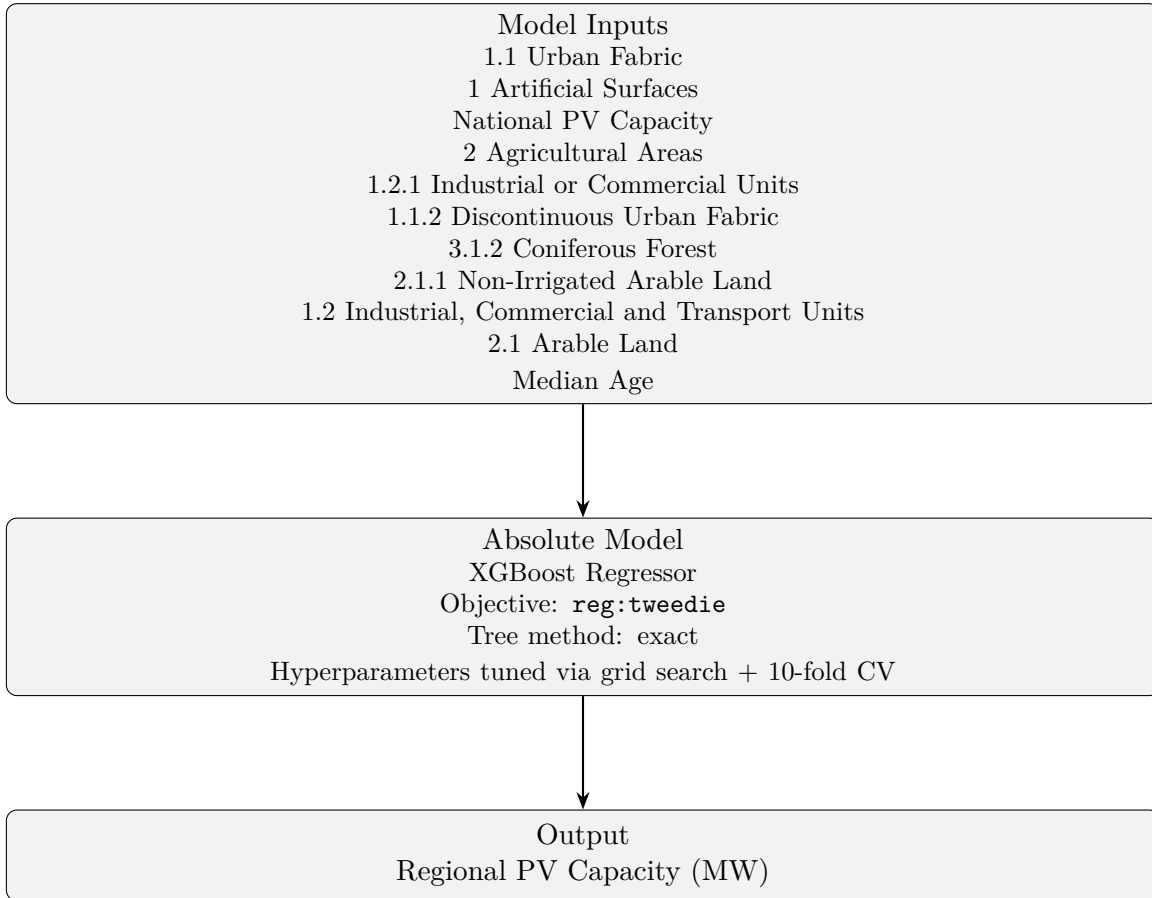


Figure 4.6: Flow chart of the absolute XGBoost model trained under the mixed hold-out configuration, showing all input features, model configuration, and predicted output.

4.6.3 Input Features and Predicted Outputs of the Normalized Model

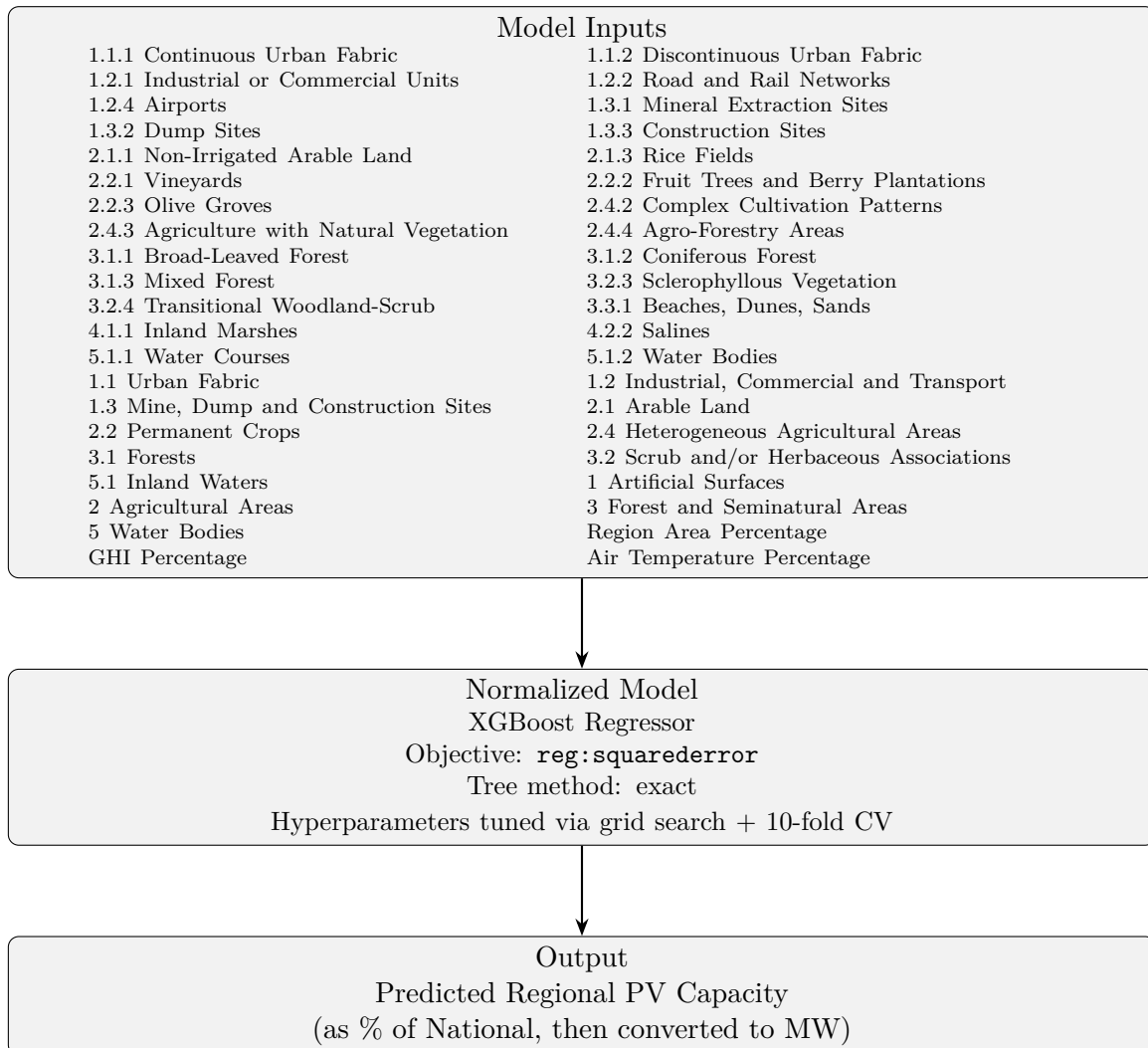


Figure 4.7: Flow chart of the normalized XGBoost model trained under the mixed hold-out configuration, showing all input features, model configuration, and predicted output.

4.6.4 Feature Importance Based on SHAP Values

Table 4.12: SHAP value contributions of features in the absolute XGBoost model trained under the mixed hold-out configuration. The mean SHAP value represents the average absolute SHAP value for each feature across all observations. The percentage contribution is calculated by dividing the mean absolute SHAP value of each feature by the sum of mean absolute SHAP values for all features, and then multiplying by 100 as shown in Equation (4.2). Percentage contributions are rounded to one decimal place and may not sum to exactly 100% due to rounding.

Feature	Mean SHAP Value	SHAP Contribution (%)
1.1 Urban Fabric (m ²)	0.92	21.3
National PV Capacity (MWp)	0.75	17.2
1 Artificial Surfaces (m ²)	0.71	16.4
2 Agricultural Areas (m ²)	0.44	10.2
1.2.1 Industrial or Commercial Units (m ²)	0.29	6.6
3.1.2 Coniferous Forest (m ²)	0.28	6.5
1.1.2 Discontinuous Urban Fabric (m ²)	0.27	6.3
2.1.1 Non-Irrigated Arable Land (m ²)	0.23	5.3
1.2 Industrial, Commercial and Transport Units (m ²)	0.17	3.9
Median Age (years)	0.14	3.3
2.1 Arable Land (m ²)	0.12	2.9

Table 4.13: SHAP value contributions of features in the normalized XGBoost model trained under the mixed hold-out configuration. All features presented are normalized relative to national values and expressed as percentages. The mean SHAP value represents the average absolute SHAP value for each feature across all observations. The percentage contribution is calculated by dividing the mean absolute SHAP value of each feature by the sum of mean absolute SHAP values for all features, and then multiplying by 100 as shown in Equation (4.2). Percentage contributions are rounded to one decimal place and may not sum to exactly 100% due to rounding.

Feature	Mean SHAP Value	SHAP Contribution (%)
2.4 Heterogeneous agricultural areas	0.81	17.2
GHI	0.49	10.5
2 Agricultural Areas	0.47	10.0
5 Water Bodies	0.46	9.8
1 Artificial Surfaces	0.20	4.3
2.2.1 Vineyards	0.20	4.3
5.1.1 Water courses	0.19	4.1
3.1.3 Mixed forest	0.17	3.6
1.3.1 Mineral extraction sites	0.14	3.0
2.1.1 Non-irrigated arable land	0.14	2.9
1.1.1 Continuous urban fabric	0.13	2.7
Air temperature	0.12	2.6
3.1.1 Broad-leaved forest	0.11	2.4
2.1 Arable land	0.11	2.3
3.2 Scrub and/or herbaceous associations	0.10	2.2
1.3.2 Dump sites	0.10	2.2
1.1.2 Discontinuous urban fabric	0.07	1.5
1.2.1 Industrial or commercial units	0.07	1.4
1.3.3 Construction sites	0.06	1.3

Feature	Mean	SHAP
	SHAP Value	Contribution (%)
1.2.4 Airports	0.06	1.2
2.4.2 Complex cultivation patterns	0.05	1.1
3.1.2 Coniferous forest	0.05	1.1
3 Forest And Seminatural Areas	0.05	1.0
2.2.2 Fruit trees and berry plantations	0.04	0.9
3.2.4 Transitional woodland-scrub	0.04	0.8
2.1.3 Rice fields	0.03	0.7
1.2.2 Road and rail networks and associated land	0.03	0.7
2.4.3 Land principally occupied by agriculture with natural vegetation	0.03	0.6
1.2 Industrial, commercial and transport units	0.02	0.5
Region area	0.02	0.5
2.4.4 Agro-forestry areas	0.02	0.4
4.1.1 Inland marshes	0.02	0.4
5.1 Inland waters	0.02	0.3
2.2.3 Olive groves	0.01	0.3
5.1.2 Water bodies	0.01	0.3
1.1 Urban fabric	0.01	0.3
2.2 Permanent crops	0.01	0.3
1.3 Mine, dump and construction sites	0.01	0.1
3.1 Forests	0.01	0.1
4.2.2 Salines	0.01	0.1
3.3.1 Beaches, dunes, sands	0.00	0.1
3.2.3 Sclerophyllous vegetation	0.00	0.0

4.6.5 Regional PV Capacity Predictions for 2023

Table 4.14: Actual and predicted regional solar PV capacity in 2023 using the XGBoost models trained under the mixed hold-out configuration. Results are shown for both the absolute and normalized models. Percentages represent each region’s share of its national total, while MW values denote capacity estimates in megawatts. “Scaled” values are adjusted to ensure that regional totals match national capacity. Blank cells indicate regions with unavailable data, whereas 0.0 denotes regions with no installed capacity within countries for which regional data exist.

NUTS	Region	Actual (%)	Actual (MW)	Pred. Norm (%)	Scaled Pred. Norm (%)	Pred. Norm (MW)	Scaled Pred. Norm (MW)	Pred. Abs (MW)	Scaled Pred. Abs (MW)
AL01	Veri			10.31	28.77	16.8	46.9	3.9	72.2
AL02	Qender			11.96	33.38	19.5	54.4	3.4	63.5
AL03	Jug			13.56	37.85	22.1	61.7	1.5	27.3
AT11	Burgenland			3.61	6.65	246.6	454.4	62.3	392.5
AT22	Steiermark			5.62	10.36	384.0	707.5	377.0	2374.3
AT33	Tirol			3.33	6.14	227.5	419.3	0.8	4.9
AT12	Niederösterreich			13.95	25.71	953.1	1756.1	188.1	1184.7
AT34	Vorarlberg			1.77	3.26	120.7	222.4	22.7	143.1
AT31	Oberösterreich			13.83	25.48	944.7	1740.7	408.2	2570.5
AT32	Salzburg			2.33	4.3	159.3	293.6	2.4	15.1
AT21	Kärnten			8.69	16.02	593.8	1094.1	1.2	7.5
AT13	Wien			1.14	2.1	77.9	143.6	22.1	139.2
BE24	Prov. Vlaams-Brabant	9.19	808.0	8.89	8.94	781.4	785.8	778.5	783.5

NUTS	Region	Actual	Actual	Pred.	Scaled	Pred.	Scaled	Pred.	Scaled	
		(%)	(MW)	Norm (%)	Pred. Norm (%)	Norm (MW)	Pred. Norm (MW)	Abs (MW)	Pred. Abs (MW)	
BE25	Prov. West-Vlaanderen	15.62	1373.1	15.88	15.96	1395.2	1403.0	1343.2	1351.7	CHAPTER 4. EUROPEAN MODEL
BE22	Prov. Limburg (BE)	9.06	796.7	12.02	12.09	1056.5	1062.4	827.2	832.4	
BE23	Prov. Oost-Vlaanderen	19.77	1737.2	17.54	17.64	1541.4	1550.0	1735.9	1746.9	
BE10	Région de Bruxelles-Capitale/ Brussels Hoofdstede...	3.42	300.5	1.8	1.81	158.4	159.3	300.1	302.0	
BE21	Prov. Antwerpen	19.72	1733.5	17.6	17.7	1547.1	1555.7	1669.9	1680.5	
BE33	Prov. Liège	9.69	851.7	9.93	9.98	872.5	877.4	857.1	862.5	
BE31	Prov. Brabant Wallon	1.82	159.5	2.55	2.56	224.2	225.4	158.6	159.6	
BE32	Prov. Hainaut	6.83	600.4	7.02	7.06	617.0	620.4	626.8	630.8	
BE34	Prov. Luxembourg (BE)	2.16	189.9	2.62	2.63	230.0	231.3	191.7	192.9	
BE35	Prov. Namur	2.71	238.3	3.6	3.62	316.1	317.9	244.5	246.1	
BG33	Severoiztochen			8.28	12.7	243.1	373.1	64.0	715.0	
BG34	Yugoiztochen			11.45	17.58	336.4	516.3	145.6	1627.7	
BG41	Yugozapaden			11.64	17.87	342.0	525.0	8.5	94.9	
BG32	Severen tsentralen			9.52	14.62	279.8	429.4	15.9	177.6	
BG42	Yuzhen tsentralen			12.23	18.76	359.1	551.2	15.0	167.2	
BG31	Severozapaden			12.03	18.46	353.3	542.2	13.8	154.8	
CH05	Ostschweiz			11.19	20.2	653.4	1179.5	106.2	1078.6	
CH01	Région lémanique			14.9	26.89	869.9	1570.4	47.8	485.3	
CH07	Ticino			2.5	4.52	146.3	264.0	4.1	41.4	

NUTS	Region	Actual (%)	Actual (MW)	Pred.	Scaled	Pred.	Scaled	Pred.	Scaled	CHAPTER 4. EUROPEAN MODEL
				Norm (%)	Pred. Norm (%)	Norm (MW)	Pred. Norm (MW)	Abs (MW)	Pred. Abs (MW)	
CH06	Zentralschweiz			7.87	14.2	459.4	829.2	26.7	271.0	
CH02	Espace Mittelland			14.67	26.49	856.9	1546.9	242.6	2463.6	
CH03	Nordwestschweiz			2.06	3.72	120.4	217.3	73.7	748.6	
CH04	Zürich			2.21	3.98	128.8	232.5	74.0	751.5	
CY00	Kýpros			15.64	100.0	94.8	606.0	0.3	606.0	
CZ02	Střední Čechy			10.44	16.0	260.9	399.9	2.1	107.1	
CZ03	Jihozápad			9.85	15.1	246.2	377.3	7.2	372.6	
CZ01	Praha			3.14	4.8	78.3	120.1	11.7	606.6	
CZ06	Jihovýchod			10.34	15.86	258.6	396.3	4.5	234.0	
CZ05	Severovýchod			9.18	14.08	229.6	351.9	5.8	300.0	
CZ08	Moravskoslezsko			4.46	6.84	111.5	170.8	6.9	357.2	
CZ04	Severozápad			10.6	16.24	264.9	406.0	0.1	2.7	
CZ07	Střední Morava			7.23	11.08	180.7	277.0	10.0	519.1	
DE12	Karlsruhe	2.43	2015.4	2.5	2.51	2073.7	2080.6	1457.0	2002.3	142
DE14	Tübingen	3.3	2741.7	3.75	3.76	3109.8	3120.1	2051.4	2819.2	
DE13	Freiburg	2.69	2231.3	3.04	3.05	2518.5	2526.9	1752.6	2408.5	
DE21	Oberbayern	5.73	4750.8	6.14	6.16	5090.7	5107.6	3469.0	4767.3	
DEF0	Schleswig-Holstein	3.62	3003.8	3.89	3.9	3228.4	3239.1	2108.3	2897.4	
DE30	Berlin	0.34	283.2	0.29	0.29	238.8	239.6	166.4	228.7	
DEA1	Düsseldorf	2.24	1861.0	1.96	1.97	1627.6	1633.0	1252.7	1721.5	

NUTS	Region	Actual	Actual	Pred.	Scaled	Pred.	Scaled	Pred.	Scaled
		(%)	(MW)	Norm (%)	Pred. Norm (%)	Norm (MW)	Pred. Norm (MW)	Abs (MW)	Pred. Abs (MW)
DEG0	Thüringen	2.97	2464.5	2.81	2.82	2332.0	2339.8	2012.5	2765.7
DEA2	Köln	2.33	1933.9	1.82	1.83	1513.0	1518.0	1087.4	1494.3
DE25	Mittelfranken	3.03	2515.4	2.82	2.82	2335.2	2342.9	1897.1	2607.0
DED2	Dresden	1.63	1350.7	1.85	1.85	1532.1	1537.2	1056.9	1452.5
DE50	Bremen	0.12	98.2	0.14	0.14	114.7	115.1	50.5	69.4
DE60	Hamburg	0.17	138.0	0.23	0.24	194.2	194.9	78.5	107.9
DEC0	Saarland	1.05	869.6	0.87	0.87	719.2	721.6	663.1	911.3
DEA4	Detmold	2.28	1895.7	2.27	2.28	1883.8	1890.0	1409.6	1937.2
DE72	Gießen	1.12	932.1	1.1	1.1	908.3	911.3	622.4	855.3
DEE0	Sachsen-Anhalt	4.98	4127.9	4.69	4.71	3892.6	3905.5	3416.7	4695.5
DED5	Leipzig	1.35	1122.2	1.36	1.36	1128.9	1132.7	801.2	1101.1
DED4	Chemnitz	1.28	1062.1	1.37	1.37	1135.1	1138.9	778.0	1069.2
DE22	Niederbayern	5.43	4504.6	6.14	6.16	5091.6	5108.5	3397.3	4668.7
DEB3	Rhein Hessen-Pfalz	1.93	1604.5	2.4	2.41	1994.8	2001.5	1181.0	1623.0
DE11	Stuttgart	4.06	3366.4	4.26	4.28	3537.7	3549.5	2433.2	3343.8
DE40	Brandenburg	8.02	6652.6	7.29	7.31	6047.4	6067.5	4747.2	6523.9
DE26	Unterfranken	3.04	2523.6	2.76	2.77	2293.5	2301.1	1685.8	2316.8
DE93	Lüneburg	1.59	1318.3	1.66	1.66	1374.5	1379.1	933.1	1282.3
DE94	Weser-Ems	4.17	3461.4	4.54	4.56	3766.7	3779.2	2579.0	3544.2
DEA5	Arnsberg	1.98	1643.3	1.82	1.82	1506.5	1511.5	1130.7	1553.9

NUTS	Region	Actual (%)	Actual (MW)	Pred. Norm (%)	Scaled Pred. Norm (%)	Pred. Norm (MW)	Scaled Pred. Norm (MW)	Pred. Abs (MW)	Scaled Pred. Abs (MW)
DE24	Oberfranken	2.32	1927.9	2.08	2.08	1720.9	1726.6	1306.6	1795.6
DEB2	Trier	1.38	1142.1	1.08	1.08	897.4	900.3	719.8	989.1
DE91	Braunschweig	1.19	990.8	1.18	1.19	982.9	986.2	634.7	872.2
DE80	Mecklenburg-Vorpommern	4.56	3784.8	3.35	3.36	2778.3	2787.5	2791.6	3836.4
DE23	Oberpfalz	3.41	2829.4	3.08	3.09	2554.1	2562.6	1967.6	2704.0
DEA3	Münster	3.09	2560.0	3.11	3.12	2580.2	2588.8	1884.9	2590.4
DE27	Schwaben	4.24	3521.1	5.19	5.2	4302.6	4316.9	2831.0	3890.6
DE71	Darmstadt	1.79	1483.3	1.72	1.73	1428.9	1433.7	991.6	1362.8
DE73	Kassel	1.64	1356.7	1.9	1.91	1578.1	1583.3	1078.7	1482.5
DE92	Hannover	1.65	1369.3	1.47	1.48	1222.0	1226.1	977.3	1343.1
DEB1	Koblenz	1.69	1405.4	1.74	1.75	1445.9	1450.7	961.0	1320.6
DK01	Hovedstaden			10.13	15.93	357.4	562.1	63.7	381.4
DK02	Sjælland			13.07	20.56	461.2	725.5	3.2	19.4
DK03	Syddanmark			15.14	23.81	534.1	840.2	323.5	1937.3
DK04	Midtjylland			13.95	21.94	492.2	774.2	28.6	171.3
DK05	Nordjylland			11.29	17.76	398.5	626.9	170.3	1019.7
EE00	Eesti			12.77	100.0	88.1	690.0	3.0	690.0
EL53	Dytiki Makedonia			6.27	6.52	440.8	458.0	20.6	225.7
EL64	Stereia Elláda			7.44	7.73	522.7	543.2	9.3	101.4
EL54	Ipeiros			8.22	8.54	577.7	600.4	1.0	11.2

NUTS	Region	Actual (%)	Actual (MW)	Pred. Norm (%)	Scaled Pred. Norm (%)	Pred. Norm (MW)	Scaled Pred. Norm (MW)	Pred. Abs (MW)	Scaled Pred. Abs (MW)
EL43	Kriti			3.39	3.52	238.2	247.5	11.7	128.2
EL65	Peloponnisos			10.01	10.4	703.7	731.3	2.2	24.2
EL30	Attiki			3.82	3.97	268.3	278.8	22.5	246.1
EL51	Anatoliki Makedonia, Thraki			7.8	8.1	548.2	569.7	155.7	1705.2
EL41	Voreio Aigaio			6.46	6.72	454.3	472.1	46.5	509.6
EL62	Ionia Nisia			2.09	2.17	146.8	152.6	51.9	568.4
EL63	Dytiki Elláda			11.53	11.98	810.4	842.1	0.4	4.5
EL61	Thessalia			9.81	10.2	689.7	716.7	192.1	2103.7
EL52	Kentriki Makedonia			12.23	12.71	859.9	893.6	8.5	92.6
EL42	Notio Aigaio			7.17	7.45	504.3	524.0	119.5	1309.2
ES53	Illes Balears			1.03	1.1	265.6	281.1	113.6	132.2
ES61	Andalucía	21.17	5434.6	20.05	21.21	5146.0	5446.1	5091.5	5929.4
ES62	Región de Murcia	6.12	1572.3	9.5	10.05	2438.6	2580.9	1732.9	2018.1
ES63	Ciudad de Ceuta			0.47	0.5	121.5	128.6	21.7	25.3
ES64	Ciudad de Melilla			0.4	0.42	101.5	107.4	23.6	27.5
ES11	Galicia	0.07	18.9	0.41	0.43	104.1	110.2	21.5	25.0
ES13	Cantabria	0.02	5.0	0.04	0.05	11.6	12.3	4.1	4.8
ES42	Castilla-La Mancha	23.89	6132.5	17.2	18.21	4416.2	4673.7	4520.9	5264.9
ES43	Extremadura	24.97	6410.0	14.27	15.1	3663.5	3877.1	5554.9	6469.1
ES41	Castilla y León	7.6	1950.3	9.4	9.95	2413.7	2554.5	1500.2	1747.1

NUTS	Region	Actual (%)	Actual (MW)	Pred. Norm (%)	Scaled Pred. Norm (%)	Pred. Norm (MW)	Scaled Pred. Norm (MW)	Pred. Abs (MW)	Scaled Pred. Abs (MW)
ES22	Comunidad Foral de Navarra	0.67	171.6	1.67	1.77	428.8	453.8	174.4	203.1
ES12	Principado de Asturias	0.0	1.2	-0.02	-0.02	-4.6	-4.9	1.7	1.9
ES52	Comunitat Valenciana	1.76	452.6	4.78	5.06	1227.7	1299.3	521.3	607.1
ES21	País Vasco	0.21	53.5	0.48	0.51	123.6	130.8	63.9	74.5
ES24	Aragón	9.4	2411.9	6.32	6.69	1622.0	1716.6	2003.8	2333.5
ES30	Comunidad de Madrid	0.25	63.5	1.27	1.35	326.6	345.6	78.7	91.7
ES70	Canarias			2.43	2.58	624.5	660.9	153.5	178.8
ES51	Cataluña	1.26	324.1	3.66	3.88	940.5	995.4	357.6	416.5
ES23	La Rioja	0.4	102.2	1.11	1.18	285.8	302.5	104.3	121.5
FI1B	Helsinki-Uusimaa			3.34	12.37	30.1	111.4	2.9	175.9
FI20	Åland			1.66	6.15	14.9	55.3	3.3	198.1
FI1C	Etelä-Suomi			8.13	30.12	73.2	271.1	3.5	209.4
FI1D	Pohjois- ja Itä-Suomi			6.91	25.58	62.2	230.3	3.5	211.9
FI19	Länsi-Suomi			6.96	25.77	62.6	232.0	1.7	104.7
FR10	Ile-de-France	1.63	321.1	1.68	1.63	330.2	321.6	316.6	315.1
FRB0	Centre — Val de Loire	4.86	956.6	3.45	3.36	678.5	661.0	965.6	961.0
FRI2	Limousin	2.62	515.1	2.22	2.17	437.7	426.4	510.8	508.4
FRI3	Poitou-Charentes	5.77	1135.4	4.98	4.85	980.2	955.0	1113.8	1108.5
FRD2	Haute-Normandie	0.87	172.1	1.32	1.28	259.8	253.1	179.7	178.8
FRD1	Basse-Normandie	1.17	230.0	1.45	1.41	285.6	278.2	242.9	241.8

NUTS	Region	Actual	Actual	Pred.	Scaled	Pred.	Scaled	Pred.	Scaled	CHAPTER 4. EUROPEAN MODEL
		(%)	(MW)	Norm (%)	Pred. Norm (%)	Norm (MW)	Pred. Norm (MW)	Abs (MW)	Pred. Abs (MW)	
FRE2	Picardie	1.31	257.8	1.08	1.05	212.9	207.4	250.6	249.4	
FRK2	Rhône-Alpes	7.14	1406.1	8.14	7.93	1602.2	1560.9	1415.6	1408.9	
FRM0	Corse	1.17	230.4	1.42	1.38	279.0	271.8	228.2	227.1	
FRF1	Alsace	2.13	418.8	2.03	1.98	400.2	389.9	379.9	378.1	
FRE1	Nord-Pas de Calais	1.49	293.0	1.82	1.77	357.4	348.1	297.0	295.6	
FRY4	La Réunion	1.22	241.0	1.11	1.08	218.2	212.6	243.5	242.3	
FRY5	Mayotte			1.11	1.08	218.2	212.6	71.1	70.7	
FRJ2	Midi-Pyrénées	9.75	1919.3	9.57	9.32	1883.3	1834.7	1959.2	1949.9	
FRG0	Pays de la Loire	6.22	1224.8	6.06	5.91	1193.9	1163.1	1235.2	1229.3	
FRY3	Guyane			1.11	1.08	218.2	212.6	67.1	66.8	
FRF2	Champagne-Ardenne	2.12	417.3	1.72	1.68	339.5	330.7	409.2	407.2	
FRL0	Provence-Alpes-Côte d’Azur	11.56	2276.4	12.11	11.8	2384.7	2323.2	2231.8	2221.2	
FRY1	Guadeloupe			1.11	1.08	218.2	212.6	100.6	100.1	
FRY2	Martinique			1.11	1.08	218.2	212.6	88.8	88.4	
FRI1	Aquitaine	14.58	2871.1	14.67	14.3	2888.7	2814.2	2858.3	2844.7	
FRF3	Lorraine	3.11	611.8	2.68	2.61	528.2	514.6	638.6	635.6	
FRC2	Franche-Comté	1.13	222.0	1.35	1.32	265.8	258.9	213.0	211.9	
FRK1	Auvergne	4.29	845.0	4.01	3.91	789.9	769.6	868.4	864.3	
FRC1	Bourgogne	3.01	592.4	2.51	2.45	494.8	482.0	579.4	576.6	
FRJ1	Languedoc-Roussillon	8.93	1758.5	9.84	9.59	1937.3	1887.3	1789.5	1781.0	

NUTS	Region	Actual (%)	Actual (MW)	Pred. Norm (%)	Scaled Pred. Norm (%)	Pred. Norm (MW)	Scaled Pred. Norm (MW)	Pred. Abs (MW)	Scaled Pred. Abs (MW)
FRH0	Bretagne	2.76	543.8	2.98	2.91	587.6	572.4	527.0	524.5
HR02	Panonska Hrvatska			13.2	34.25	60.8	157.8	2.0	206.9
HR03	Jadranska Hrvatska			11.98	31.07	55.2	143.1	0.3	37.0
HR06	Sjeverna Hrvatska			10.89	28.26	50.2	130.2	2.0	207.8
HR05	Grad Zagreb			2.48	6.43	11.4	29.6	0.1	9.0
HU22	Nyugat-Dunántúl			8.57	12.38	500.4	722.1	105.5	442.4
HU31	Észak-Magyarország			9.13	13.17	532.5	768.6	215.7	904.4
HU23	Dél-Dunántúl			9.56	13.79	557.7	804.8	115.0	482.3
HU12	Pest			5.83	8.41	340.1	490.8	341.9	1433.7
HU11	Budapest			1.97	2.84	114.8	165.7	8.3	34.8
HU21	Közép-Dunántúl			9.35	13.49	545.5	787.3	149.9	628.7
HU32	Észak-Alföld			12.28	17.73	716.8	1034.6	279.0	1169.9
HU33	Dél-Alföld			12.6	18.18	735.2	1061.1	176.2	738.8
IE06	Eastern and Midland			12.85	32.92	94.8	243.0	0.5	506.5
IE04	Northern and Western			10.02	25.68	74.0	189.6	0.0	29.2
IE05	Southern			16.16	41.4	119.2	305.5	0.2	202.4
IS00	Ísland			3.7	100.0	0.3	7.0	0.0	7.0
ITG2	Sardegna	4.42	1337.0	4.07	4.14	1232.5	1254.0	1339.8	1370.7
ITH1	Provincia Autonoma di Bolzano/Bozen			0.4	0.4	119.7	121.8	2.0	2.1
ITH4	Friuli-Venezia Giulia	2.89	873.8	2.5	2.55	757.9	771.1	873.0	893.1

NUTS	Region	Actual	Actual	Pred.	Scaled	Pred.	Scaled	Pred.	Scaled
		(%)	(MW)	Norm (%)	Pred. Norm (%)	Norm (MW)	Pred. Norm (MW)	Abs (MW)	Pred. Abs (MW)
ITI1	Toscana	4.04	1223.3	4.13	4.21	1251.7	1273.5	1189.0	1216.4
ITI2	Umbria	2.08	631.5	2.38	2.42	721.3	733.9	627.7	642.2
ITH5	Emilia-Romagna	10.0	3027.1	9.99	10.17	3026.6	3079.4	3020.1	3089.7
ITG1	Sicilia	7.15	2163.9	6.92	7.04	2095.3	2131.9	2172.5	2222.5
ITI4	Lazio	6.74	2041.4	6.63	6.75	2008.8	2043.9	2068.7	2116.3
ITH3	Veneto	10.45	3164.0	9.65	9.82	2921.3	2972.4	3138.7	3211.0
ITH2	Provincia Autonoma di Trento			0.8	0.82	242.4	246.7	1.1	1.1
ITF5	Basilicata	1.66	501.5	1.9	1.94	576.1	586.2	510.7	522.5
ITF6	Calabria	2.41	730.8	2.6	2.65	787.7	801.4	752.6	769.9
ITC3	Liguria	0.62	187.3	0.55	0.56	167.8	170.7	188.6	192.9
ITC2	Valle d'Aosta/Vallée d'Aoste	0.12	34.9	0.19	0.19	57.5	58.5	34.6	35.4
ITF1	Abruzzo	3.21	972.9	3.47	3.53	1051.1	1069.5	955.8	977.8
ITF2	Molise	0.69	208.2	0.91	0.92	274.2	279.0	205.2	209.9
ITC4	Lombardia	13.39	4056.1	11.75	11.95	3557.1	3619.2	4024.3	4116.9
ITF3	Campania	4.05	1226.2	3.91	3.98	1185.4	1206.1	1203.5	1231.2
ITC1	Piemonte	8.44	2557.0	8.16	8.3	2471.0	2514.2	2577.4	2636.7
ITI3	Marche	4.5	1362.6	5.24	5.33	1585.7	1613.4	1367.0	1398.5
ITF4	Puglia	10.92	3306.1	12.12	12.33	3670.5	3734.6	3347.9	3425.0
LT02	Vidurio ir vakarų Lietuvos regionas			12.48	50.46	145.4	587.9	6.2	1053.9
LT01	Sostinės regionas			12.26	49.54	142.8	577.1	0.7	111.1

NUTS	Region	Actual	Actual	Pred.	Scaled	Pred.	Scaled	Pred.	Scaled
		(%)	(MW)	Norm (%)	Pred. Norm (%)	Norm (MW)	Pred. Norm (MW)	Abs (MW)	Pred. Abs (MW)
LU00	Luxembourg			14.62	100.0	63.1	431.6	2.4	431.6
LV00	Latvija			12.76	100.0	45.1	353.0	0.3	353.0
ME00	Crna Gora			12.03	100.0	5.0	41.8	0.0	41.8
MK00	Severna Makedonija			12.9	100.0	69.0	534.6	0.1	534.6
MT00	Malta			15.18	100.0	35.1	231.0	0.5	231.0
NL12	Friesland (NL)			5.8	5.83	1387.1	1394.5	186.0	2141.4
NL13	Drenthe			6.24	6.28	1492.6	1500.5	34.4	396.6
NL11	Groningen			6.34	6.37	1514.4	1522.5	92.1	1060.5
NL41	Noord-Brabant			12.44	12.5	2973.0	2988.9	499.2	5748.7
NL42	Limburg (NL)			11.75	11.81	2809.1	2824.1	344.3	3965.2
NL22	Gelderland			12.47	12.54	2981.3	2997.2	21.5	247.3
NL34	Zeeland			5.88	5.91	1405.7	1413.2	95.0	1094.3
NL23	Flevoland			4.44	4.46	1061.6	1067.2	49.1	564.9
NL31	Utrecht			5.46	5.49	1306.3	1313.3	91.0	1047.5
NL32	Noord-Holland			10.28	10.33	2457.1	2470.2	22.4	258.5
NL33	Zuid-Holland			7.18	7.21	1715.4	1724.6	384.1	4423.5
NL21	Overijssel			11.18	11.24	2673.7	2688.0	256.7	2955.6
NO02	Innlandet			3.55	15.62	21.8	96.2	0.0	0.2
NO06	Trøndelag			7.12	31.36	43.8	193.1	0.0	3.1
NO07	Nord-Norge			4.28	18.84	26.3	116.0	0.0	4.4

NUTS	Region	Actual	Actual	Pred.	Scaled	Pred.	Scaled	Pred.	Scaled
		(%)	(MW)	Norm (%)	Pred. Norm (%)	Norm (MW)	Pred. Norm (MW)	Abs (MW)	Pred. Abs (MW)
NO08	Oslo og Viken			1.84	8.11	11.3	50.0	1.2	232.9
NO09	Agder og Sør-Østlandet			1.0	4.4	6.2	27.1	0.3	70.0
NO0A	Vestlandet			3.81	16.78	23.5	103.3	0.6	121.8
NO0B	Jan Mayen and Svalbard			1.11	4.88	6.8	30.1	0.9	183.3
PL22	Śląskie			5.43	6.06	858.2	958.9	666.2	1437.5
PL61	Kujawsko-pomorskie			3.29	3.68	520.4	581.5	735.3	1586.6
PL41	Wielkopolskie			5.69	6.35	899.0	1004.5	416.3	898.3
PL62	Warmińsko-mazurskie			3.04	3.4	481.4	537.9	522.2	1126.8
PL42	Zachodniopomorskie			5.67	6.33	896.3	1001.4	344.7	743.9
PL51	Dolnośląskie			5.36	5.99	847.6	947.0	948.8	2047.5
PL52	Opolskie			3.64	4.07	575.8	643.4	369.5	797.4
PL71	Łódzkie			6.09	6.8	962.4	1075.3	991.6	2139.9
PL63	Pomorskie			3.93	4.39	621.9	694.9	183.8	396.6
PL43	Lubuskie			3.29	3.67	519.8	580.8	602.1	1299.2
PL92	Mazowiecki regionalny			9.91	11.07	1566.6	1750.4	197.8	426.7
PL72	Świętokrzyskie			3.38	3.78	534.2	596.9	336.3	725.7
PL82	Podkarpackie			5.38	6.01	850.4	950.2	150.5	324.9
PL81	Lubelskie			8.55	9.55	1351.0	1509.6	228.9	493.9
PL21	Małopolskie			5.04	5.63	797.0	890.6	190.7	411.5
PL84	Podlaskie			7.49	8.36	1183.6	1322.5	8.7	18.8

NUTS	Region	Actual	Actual	Pred.	Scaled	Pred.	Scaled	Pred.	Scaled	CHAPTER 4. EUROPEAN MODEL
		(%)	(MW)	Norm (%)	Pred. Norm (%)	Norm (MW)	Pred. Norm (MW)	Abs (MW)	Pred. Abs (MW)	
PL91	Warszawski stołeczny			4.32	4.83	683.4	763.6	432.9	934.2	
PT17	Área Metropolitana de Lisboa			4.28	7.01	165.9	271.7	5.6	26.5	
PT16	Centro (PT)			14.87	24.36	576.4	944.2	115.7	546.2	
PT18	Alentejo			16.88	27.66	654.4	1072.0	10.2	48.2	
PT15	Algarve			4.33	7.09	167.8	274.9	18.2	86.0	
PT11	Norte			16.67	27.3	646.0	1058.2	626.5	2956.7	
PT20	Região Autónoma dos Açores			2.32	3.8	89.8	147.1	39.8	187.7	
PT30	Região Autónoma da Madeira			1.7	2.79	66.0	108.1	5.3	25.0	
RO21	Nord-Est			8.11	12.96	155.5	248.4	1.3	12.9	
RO31	Sud-Muntenia			10.08	16.11	193.3	308.8	72.3	734.6	
RO22	Sud-Est			10.64	17.0	204.0	325.9	43.1	438.5	
RO12	Centru			7.99	12.77	153.2	244.8	5.2	52.8	
RO11	Nord-Vest			7.43	11.87	142.4	227.5	4.2	42.8	
RO32	București-Ilfov			1.48	2.37	28.4	45.4	24.4	248.1	
RO41	Sud-Vest Oltenia			9.11	14.55	174.5	278.8	31.7	322.3	
RO42	Vest			7.75	12.38	148.5	237.2	6.4	65.1	
RS21	Region Šumadije i Zapadne Srbije			10.69	25.86	14.6	35.4	0.5	11.5	
RS12	Autonomous Province of Vojvodina			11.41	27.59	15.6	37.8	1.0	25.2	
RS22	Region Južne i Istočne Srbije			10.87	26.3	14.9	36.0	3.3	79.5	
RS11	City of Belgrade			8.37	20.25	11.5	27.7	0.8	20.7	

NUTS	Region	Actual	Actual	Pred.	Scaled	Pred.	Scaled	Pred.	Scaled	CHAPTER 4. EUROPEAN MODEL
		(%)	(MW)	Norm (%)	Pred. Norm (%)	Norm (MW)	Pred. Norm (MW)	Abs (MW)	Pred. Abs (MW)	
SE11	Stockholm			2.33	7.49	81.3	261.2	27.9	183.4	
SE12	Östra Mellansverige			4.17	13.4	145.5	467.4	151.4	995.8	
SE21	Småland med öarna			8.09	25.99	282.1	906.5	18.5	121.7	
SE23	Västssverige			4.51	14.48	157.2	505.2	107.6	707.2	
SE33	Övre Norrland			3.16	10.16	110.3	354.4	0.4	2.4	
SE31	Norra Mellansverige			2.07	6.65	72.2	231.8	108.2	711.2	
SE32	Mellersta Norrland			2.84	9.13	99.1	318.5	26.1	171.3	
SE22	Sydsverige			3.95	12.7	137.9	442.9	90.5	595.0	
SI04	Zahodna Slovenija			12.35	48.0	127.7	496.4	1.0	578.6	
SI03	Vzhodna Slovenija			13.38	52.0	138.4	537.8	0.8	455.6	
SK03	Stredné Slovensko			10.95	24.94	69.1	157.4	5.1	50.3	
SK04	Východné Slovensko			9.92	22.59	62.6	142.5	1.3	12.7	
SK02	Západné Slovensko			14.82	33.76	93.5	213.0	48.0	471.8	
SK01	Bratislavský kraj			8.22	18.72	51.9	118.1	9.8	96.3	
TR82	Kastamonu, Çankırı, Sinop			1.58	1.7	177.9	191.9	4.4	46.8	
TRC2	Şanlıurfa, Diyarbakır			7.86	8.48	887.7	957.5	90.2	963.5	
TRC3	Mardin, Batman, Şırnak, Siirt			2.67	2.88	301.9	325.7	12.3	131.2	
TR81	Zonguldak, Karabük, Bartın			0.6	0.64	67.2	72.5	0.1	0.6	
TRB2	Van, Muş, Bitlis, Hakkari			6.64	7.17	750.4	809.4	17.4	186.3	
TR83	Samsun, Tokat, Çorum, Amasya			2.24	2.42	253.0	272.9	70.0	747.5	

NUTS	Region	Actual (%)	Actual (MW)	Pred. Norm (%)	Scaled Pred. Norm (%)	Pred. Norm (MW)	Scaled Pred. Norm (MW)	Pred. Abs (MW)	Scaled Pred. Abs (MW)
TR90	Trabzon, Ordu, Giresun, Rize, Artvin, Gümüşhane			1.36	1.47	154.2	166.3	0.3	3.6
TR71	Kırıkkale, Aksaray, Niğde, Nevşehir, Kırşehir			3.52	3.8	397.7	429.0	176.3	1882.2
TRC1	Gaziantep, Adıyaman, Kilis			2.37	2.56	267.4	288.5	7.8	83.8
TRA2	Ağrı, Kars, Iğdır, Ardahan			1.6	1.73	181.0	195.3	5.8	62.4
TRB1	Malatya, Elazığ, Bingöl, Tunceli			6.0	6.48	678.0	731.4	1.4	15.2
TRA1	Erzurum, Erzincan, Bayburt			2.38	2.57	268.6	289.8	6.3	67.3
TR63	Hatay, Kahramanmaraş, Osmaniye			3.29	3.55	371.5	400.7	0.7	7.0
TR72	Kayseri, Sivas, Yozgat			2.74	2.95	309.2	333.5	78.2	835.3
TR61	Antalya, Isparta, Burdur			8.51	9.18	961.0	1036.6	0.0	0.5
TR62	Adana, Mersin			3.91	4.21	441.0	475.7	0.4	4.7
TR33	Manisa, Afyonkarahisar, Kütahya, Uşak			10.39	11.2	1172.9	1265.1	205.5	2194.0
TR41	Bursa, Eskişehir, Bilecik			7.97	8.6	899.7	970.5	0.5	4.9
TR21	Tekirdağ, Edirne, Kırklareli			2.54	2.73	286.2	308.7	33.6	358.8
TR42	Kocaeli, Sakarya, Düzce, Bolu, Yalova			0.98	1.06	111.0	119.8	1.2	13.1
TR22	Balıkesir, Çanakkale			0.66	0.71	74.2	80.0	26.4	282.1
TR31	İzmir			1.18	1.27	132.7	143.1	0.0	0.2
TR10	İstanbul			0.18	0.19	19.9	21.5	176.0	1880.0
TR32	Aydın, Denizli, Muğla			2.3	2.48	259.2	279.6	0.8	8.9

NUTS	Region	Actual (%)	Actual (MW)	Pred.	Scaled	Pred.	Scaled	Pred.	Scaled
				Norm (%)	Pred. Norm (%)	Norm (MW)	Pred. Norm (MW)	Abs (MW)	Pred. Abs (MW)
TR51	Ankara			2.53	2.73	286.0	308.5	9.8	105.0
TR52	Konya, Karaman			6.72	7.24	758.4	818.1	131.7	1406.7
UKG1	Herefordshire, Worcestershire and Warwickshire	3.14	517.2	2.58	2.77	423.4	455.0	503.6	528.3
UKL2	East Wales	3.22	530.3	2.32	2.49	380.9	409.3	533.7	560.0
UKM5	North Eastern Scotland	0.0	0.0	0.18	0.2	30.4	32.6	0.0	0.0
UKM6	Highlands and Islands	0.0	0.0	-0.01	-0.01	-1.3	-1.4	0.0	0.0
UKF3	Lincolnshire	3.24	532.2	2.73	2.94	449.1	482.6	530.6	556.7
UKK3	Cornwall and Isles of Scilly	3.74	615.4	4.68	5.03	770.0	827.5	636.3	667.6
UKM8	West Central Scotland	0.0	0.0	0.1	0.11	17.1	18.3	0.0	0.0
UKM9	Southern Scotland	0.0	0.0	0.06	0.06	9.0	9.7	0.0	0.0
UKJ1	Berkshire, Buckinghamshire and Oxfordshire	5.1	839.4	4.93	5.3	811.5	872.1	812.6	852.6
UKM7	Eastern Scotland	0.0	0.0	0.15	0.16	25.3	27.1	0.0	0.0
UKG2	Shropshire and Staffordshire	3.23	531.8	2.82	3.03	463.5	498.1	539.0	565.5
UKH3	Essex	2.94	483.4	2.6	2.79	426.9	458.8	460.7	483.4
UKL1	West Wales and The Valleys	5.74	943.5	4.82	5.18	792.8	852.0	935.8	981.9
UKN0	Northern Ireland			0.14	0.15	23.4	25.2	0.0	0.0
UKC1	Tees Valley and Durham	0.96	157.8	1.03	1.11	169.3	181.9	156.4	164.1
UKH2	Bedfordshire and Hertfordshire	2.3	377.5	2.06	2.22	338.9	364.2	368.4	386.5

NUTS	Region	Actual (%)	Actual (MW)	Pred. Norm (%)	Scaled Pred. Norm (%)	Pred. Norm (MW)	Scaled Pred. Norm (MW)	Pred. Abs (MW)	Scaled Pred. Abs (MW)
UKE3	South Yorkshire	0.88	145.0	1.49	1.6	244.3	262.5	141.7	148.6
UKJ3	Hampshire and Isle of Wight	4.36	717.5	4.42	4.76	727.6	781.9	722.3	757.8
UKC2	Northumberland and Tyne and Wear	0.9	147.2	1.09	1.17	179.1	192.4	139.1	145.9
UKG3	West Midlands	0.85	140.0	1.02	1.09	167.1	179.6	136.2	142.9
UKD6	Cheshire	0.87	142.4	0.88	0.94	144.1	154.9	145.4	152.6
UKD7	Merseyside	0.56	92.9	0.68	0.72	111.0	119.3	87.8	92.1
UKH1	East Anglia	10.44	1716.2	9.75	10.48	1603.4	1723.1	1714.9	1799.4
UKI7	Outer London — West and North West	0.47	77.9	0.63	0.68	104.2	112.0	75.9	79.7
UKI3	Inner London — West	0.14	23.4	0.22	0.23	35.8	38.5	23.5	24.7
UKI4	Inner London — East	0.37	61.3	0.46	0.49	75.3	80.9	60.3	63.3
UKI5	Outer London — East and North East	0.46	76.2	0.51	0.55	83.7	89.9	76.5	80.2
UKD1	Cumbria	0.74	121.7	0.74	0.79	120.9	129.9	118.6	124.4
UKD3	Greater Manchester	1.23	201.8	1.38	1.48	226.5	243.5	200.0	209.8
UKD4	Lancashire	1.35	222.5	1.43	1.54	235.2	252.8	220.3	231.1
UKE4	West Yorkshire	1.07	176.7	1.41	1.51	231.1	248.4	168.1	176.4
UKE2	North Yorkshire	1.65	271.6	1.48	1.59	243.3	261.5	264.7	277.7
UKF1	Derbyshire and Nottinghamshire	4.2	691.2	4.11	4.42	676.1	726.6	716.8	752.1
UKJ4	Kent	3.08	507.0	2.89	3.11	475.9	511.5	520.9	546.5
UKI6	Outer London — South	0.37	60.6	0.48	0.51	78.3	84.2	62.1	65.2

NUTS	Region	Actual (%)	Actual (MW)	Pred. Norm (%)	Scaled Pred. Norm (%)	Pred. Norm (MW)	Scaled Pred. Norm (MW)	Pred. Abs (MW)	Scaled Pred. Abs (MW)
UKF2	Leicestershire, Rutland and Northamptonshire	3.98	654.8	3.36	3.61	552.2	593.4	667.7	700.6
UKK4	Devon	4.01	659.2	4.7	5.05	772.9	830.7	670.0	703.0
UKK1	Gloucestershire, Wiltshire and Bristol/Bath area	8.57	1409.9	7.73	8.3	1270.9	1365.8	1474.6	1547.1
UKK2	Dorset and Somerset	5.67	931.7	5.66	6.09	931.4	1001.0	950.8	997.6
UKE1	East Yorkshire and Northern Lincolnshire	1.81	298.0	1.52	1.64	250.6	269.3	307.3	322.4
UKJ2	Surrey, East and West Sussex	3.34	549.4	3.84	4.12	631.2	678.3	531.3	557.4

4.6.6 Forecast Validation

To validate the forecasting framework used for the 2030 projection (shown in Figure 4.5), we evaluate model performance over 2017–2023 by comparing two approaches: (i) a carry-forward (*naïve*) baseline that holds the 2016 regional distribution constant and scales it by the national total of the target year, and (ii) a scaled normalized model trained on 2010–2016 data and evaluated on 2017–2023, which predicts regional shares using features observed in 2016 (e.g., land cover, GHI, and temperature). Belgium (BE) is excluded from training because regional capacity data were unavailable for 2010–2016.

For the naïve approach, the regional share of national capacity is fixed at the base year:

$$\hat{C}_{t,i}^{\text{naïve}} = s_{2016,i} \text{Nat}_t^{\text{MW}},$$

where $s_{2016,i}$ is the 2016 regional share of national capacity for region i , and Nat_t^{MW} is the national total capacity (MW) in year t . For share-based evaluation, the forecast share is given by $\hat{s}_{t,i}^{\text{naïve}} = s_{2016,i}$, with $\text{Nat}_t^{\%} = 1$ representing the national total in fractional terms.

The scaled normalized model predicts regional shares $\hat{s}_{t,i}$ using features observed in 2016 and scales them by national totals to obtain MW forecasts:

$$\hat{C}_{t,i}^{\text{model}} = \hat{s}_{t,i} \text{Nat}_t^{\text{MW}}.$$

Table 4.15 reports regional error metrics for both approaches over the full 2017–2023 horizon and for 2023 only. When regional deployment patterns are stable, the naïve carry-forward baseline provides a strong reference and often achieves low errors. However, three limitations are particularly relevant in this context:

1. In some cases, a share of the national PV capacity is not assigned to specific regions, so the observed regional shares do not sum to one. The scaled normalized model ensures that regional shares collectively equal the national total, redistributing this unlocated capacity across regions. When model performance is evaluated only against the located

subset, this redistribution can inflate apparent errors even if the underlying spatial allocation is more accurate.

2. Where regional patterns evolve, the carry-forward assumption misallocates capacity, whereas the scaled normalized model can adapt by reallocating capacity based on features.
3. The carry-forward baseline cannot be applied in countries lacking baseline regional PV capacity data, whereas the normalized model remains applicable wherever features are available.

Overall, the naïve baseline achieves strong accuracy over 2017–2023, reflecting persistence in many settings. The scaled normalized model performs comparably overall but better addresses the identified limitations. For the 2023-only evaluation, errors increase for both methods, as expected, due to the longer forecast horizon. These results demonstrate that the scaled normalized model provides a robust and transferable framework for forecasting regional PV capacity, particularly where spatial distributions evolve over time.

Table 4.15: Regional error metrics for the naïve carry-forward baseline and the scaled normalized model. Models are evaluated over the full 2017–2023 horizon and for 2023 only. Reported metrics are region-level R^2 , MAE, MSE, and RMSE, presented for both megawatts (MW) and percentage shares (%).

Model	Train R^2	Test R^2	Train MAE	Test MAE	Train MSE	Test MSE	Train RMSE	Test RMSE
<i>2017–2023 horizon</i>								
naïve benchmark (MW)	N/A	0.94	N/A	83	N/A	56012	N/A	237
Scaled normalized model (MW)	0.99	0.92	36	130	3628	70475	60	265
naïve benchmark (%)	N/A	0.91	N/A	0.38	N/A	1.36	N/A	1.16
Scaled normalized model (%)	0.97	0.87	0.34	0.76	0.28	2.50	0.53	1.58
<i>2023 only</i>								
naïve benchmark (MW)	N/A	0.91	N/A	193	N/A	183410	N/A	428
Scaled normalized model (MW)	0.99	0.89	36	240	3628	204618	60	452
naïve benchmark (%)	N/A	0.86	N/A	0.65	N/A	2.53	N/A	1.59
Scaled normalized model (%)	0.97	0.80	0.34	1.02	0.28	4.29	0.53	2.07

Table 4.16: Country-level regional error metrics for 2023 comparing the naïve carry-forward baseline and the scaled normalized model across Germany (DE), Spain (ES), the United Kingdom (UK), Belgium (BE), France (FR), and Italy (IT). Metrics are region-level R^2 , MAE, MSE, RMSE, and MAPE (%). Belgium (BE) has missing naïve values because regional PV capacity data for 2010–2016 were unavailable.

Country	R^2		MAE (MW)		MSE (MW ²)		RMSE (MW)		MAPE (%)	
	Naïve	Scaled	Naïve	Scaled	Naïve	Scaled	Naïve	Scaled	Naïve	Scaled
DE	0.96	0.92	181	262	71754	147471	268	384	10	13
ES	0.74	0.76	815	798	1347289	1282809	1161	1133	206	314
UK	0.98	0.94	38	66	3422	9528	58	98	11	1.36×10^{16}
BE	–	0.38	–	390	–	195617	–	442	–	89
FR	0.96	0.95	108	119	21223	24798	146	157	18	20
IT	0.95	0.92	165	189	72600	111582	269	334	11	15

Table 4.16 presents regional error metrics by country for the year 2023. In Germany, France, and Italy, both approaches perform well, with MAPE values ranging between 10% and 20% per region. In Spain, MAPE is high for both approaches because several regions have very small 2023 capacities, making percentage errors highly sensitive to small denominators. Moreover, regional data for Spain are only available from 2015, which may partly explain the weaker performance relative to other countries. In the United Kingdom, MAPE for the scaled model is inflated because several regions with zero observed 2023 capacity are predicted to have small positive values; in contrast, the naïve baseline predicts zero for those regions (2016 share = 0) and therefore yields finite percentage errors. Nevertheless, the remaining error metrics indicate that forecast accuracy in the UK is very good overall. Belgium represents a case where the model forecasts capacity in a country that was not included in training. Although MAPE is higher (89%) than in countries within the training set, the other error metrics remain reasonable given that the model was extrapolated to unseen data.

Chapter 5

Great Britain Model

5.1 Abstract

Great Britain aims to meet growing electricity demand and achieve a fully decarbonised grid by 2035, targeting 70 GW of solar photovoltaic (PV) capacity. However, grid constraints and connection delays hinder solar integration. To address these integration challenges, various connection reform processes and policies are being developed [263]. This study supports the connection reforms with a model that estimates regional PV capacity at the NUTS 3 level, explaining 89% of the variation in capacity, with a mean absolute error of 20 MW and a national mean absolute percentage error of 5.4%. Artificial surfaces and agricultural areas are identified as key factors in deployment. The model has three primary applications: disaggregating national PV capacity into regional capacity, benchmarking regional PV deployment between different regions, and forecasting future PV capacity distribution. These applications support grid operators in generation monitoring and strategic grid planning by identifying regions where capacity is likely to be concentrated. This can address grid connection delays, plan network expansions, and resolve land-use conflicts.

5.2 Introduction

The demand for electricity in Great Britain is projected to grow significantly, potentially increasing by 65% by 2035. To address this rising demand, the UK government has set ambitious goals including achieving a fully decarbonised electricity system [264] and installing 70 GW of solar power by 2035 [265].

Significant progress has already been made, with 2023 marking the first year that renewable energy generation surpassed fossil fuel generation in Britain. However, the electricity grid faces mounting challenges as it nears its capacity limits. Investment in renewable generation has outpaced spending on transmission infrastructure over the past decade, creating constraints that restrict the grid’s ability to transport electricity. Consequently, energy is sometimes wasted when renewable sources are curtailed to prevent grid overloading [264].

A major obstacle to scaling up renewable energy, including solar PV, is grid connection delays. As of October 2024, 732 GW of projects were queued to connect to Great Britain’s transmission network, with renewables accounting for approximately 363 GW [74]. The queue grows as investors hedge their bets on where grid connection approvals might be granted, submitting multiple speculative applications for the same or similar projects in different locations. This strategy is often employed to improve the likelihood of securing a connection in a system where the timing and location of approvals are uncertain. However, this approach leads to an artificially inflated connection queue, as many of these projects are unlikely to materialize. It also creates inefficiencies for grid operators, who must process and manage a large volume of speculative applications, slowing down the approval process for viable projects [75]. These delays are further exacerbated by insufficient physical network infrastructure, such as cables, transformers, and substations, which are critical for accommodating new connections [76, 77]. This reflects the growing strain on Great Britain’s grid in the context of a rapidly decarbonizing energy system, a challenge mirrored globally as countries transition from fossil fuels to renewable energy sources [78].

To tackle these challenges, the UK has introduced the Connections Reform initiative to improve the efficiency of the grid connection process [78, 75]. Historically, this process

has been reactive, addressing individual connection requests with little consideration of the broader network needs. The reform aims to streamline the process, prioritize projects nearing completion, and enhance transparency for developers. However, these measures alone cannot resolve the underlying need for a comprehensive, long-term strategy to align grid development with future energy demand and renewable deployment [266, 267].

Strategic planning is especially important for solar PV deployment. Limited data on installations, due to General Data Protection Regulation (GDPR) and commercial data restrictions, complicate efforts to plan and invest in grid infrastructure effectively [76], and reduces the accuracy of PV generation estimates [77]. The lack of Distribution Network Operator (DNO) targets for new PV grid connections is also a barrier to reaching the UK’s target of 70 GW by 2035 [77]. Moreover, accurate regional capacity modelling is critical for reducing errors in PV generation estimates. While the national capacity error is around 5% [49], regional errors are likely higher given that aggregation reduces the error.

Developing a model to estimate regional PV capacity based on geographical factors (social, economic, land use, and climatic) could address these challenges by proactively identifying regions where grid development might be needed. The model could help with setting realistic PV connection targets for DNOs and provide more accurate estimates of regional PV capacity, thereby improving generation monitoring. A regional PV capacity model could help policymakers identify areas where deployment lags behind expectations, enabling targeted interventions to support underserved regions.

While we focus on the GB system in this paper, the issues are shared amongst countries. For example, grid connection constraints present significant challenges in various countries, including Austria [80], Bulgaria [80], Croatia [81], Chile [82], Finland [81], France [81], Germany [81], Greece [81, 80], Hungary [80], Ireland [81], Italy [81], Netherlands [37, 81], Poland [81], Spain [81, 83], Sweden [81], Turkey [81], and the United States [84].

5.2.1 Literature Review

Addressing grid connection constraints requires a deeper understanding of the geographical factors that influence solar PV deployment. Previous studies have highlighted a range of key factors that impact deployment, including social, economic, climatic, and land use variables. Population is positively correlated with solar PV deployment at the country level globally [229]. This is expected, as countries with larger populations generally require more electricity, which drives higher solar PV installation rates. However, the relationship between solar PV deployment and population becomes more complex when examined at regional and subregional scales. For example, studies have found a negative correlation between solar PV deployment and population at the regional level in both Germany [114] and the UK [115] and at the subregional level in Australia [105]. In the United States, Yu et al. [98] observed that residential solar PV deployment peaks at a population density of 1,000 people per square mile. Areas with very high population densities tend to have lower levels of small-scale residential solar PV deployment, as urban environments often lack suitable rooftops for installations. Conversely, regions with medium population densities are more likely to have higher solar PV capacities due to the prevalence of detached houses with rooftops that are ideal for solar PV systems.

Education is positively correlated with solar PV deployment at the country level globally [229], at the regional level in the UK [115], and at the subregional level in Connecticut, USA [98], and England [109]. Countries with higher education levels tend to have stronger economies, which often translates into higher electricity consumption and greater investment in solar PV installations. Furthermore, higher levels of education are associated with increased environmental awareness and pro-environmental behaviour [268, 269, 270], which can further drive solar PV adoption.

Average household size has a mixed impact on solar PV deployment. In the UK, it is positively correlated with deployment at the subregional level [107]. This positive relationship could be influenced by larger energy bills or increased daytime electricity usage. Conversely, a negative relationship is observed at the regional and subregional levels in the UK and at the

subregional level in Australia [115, 106, 105]. This negative relationship may be explained by larger households having reduced cash flow or prioritizing aesthetics over cost savings from solar PV systems [115].

Household size may also be linked to age demographics, which play a significant role in solar PV adoption. Individuals above age 40 are more likely to have greater access to cash for solar PV investments, while those aged 25–40 may face financial constraints that limit their ability to install solar PV systems. In England, a higher share of the population above age 40 is associated with greater PV installation at the subregional level [109]. Conversely, in Australia, a higher share of the population aged 25–40 has a negative impact on PV adoption at the subregional level [105].

Income shows a mixed relationship with solar PV deployment. In the UK, income is positively correlated with PV adoption at the subregional and regional levels, and similarly, a positive correlation is observed at the subregional level in the United States [106, 115, 98]. However, a negative correlation between income and solar PV adoption has been found at the subregional level in Australia [105]. This mixed relationship may be explained by differences in motivation. While individuals with higher incomes can afford to install solar PV systems, they may choose not to if they are not concerned about cost savings or if they dislike the aesthetics of solar panels.

GDP is strongly correlated with solar PV capacity at the country level globally and in China [229, 116]. However, in other studies, GDP does not show a proportional relationship to a country's solar PV capacity [117]. At the regional level in Germany, GDP is negatively correlated with PV deployment [114].

Industrial added value is positively correlated with solar PV deployment at the country level [229, 116]. This relationship may reflect the fact that countries with higher industrial output tend to have more resources and investments available for renewable energy technologies, including solar PV. Additionally, industrial sectors often benefit from renewable energy adoption both through direct use in manufacturing processes and through the promotion of green technologies, which may further drive solar PV deployment.

Electricity consumption is positively correlated with solar PV deployment at the country level globally [229]. This relationship extends to the regional and subregional levels in the UK, where higher electricity consumption is also associated with greater PV deployment [106, 109, 115, 107].

Solar radiation has been found to have a positive correlation with solar PV deployment at the regional and subregional levels [110, 98, 103, 115]. This is logical, as regions with abundant sunlight provide more favourable conditions for solar power generation, making solar PV an attractive energy option. However, other studies suggest that a country's solar radiation potential is not always proportional to solar PV deployment [117, 229]. This discrepancy may be linked to the geography of a country or region. In countries with abundant agricultural land receiving high levels of solar radiation, there is often more solar PV capacity, as large open spaces are ideal for solar farms. In contrast, countries such as Austria, where two-thirds of the land is covered by the Alpine mountains [271], and where solar radiation is concentrated in urban areas, may face challenges in deploying solar PV, as urban environments typically offer limited space for large-scale installations.

Rural areas are positively correlated with solar PV deployment [110, 109, 103]. This is likely due to the availability of space for solar installations and the higher likelihood that rural areas have houses with suitable rooftops for PV systems.

Solar PV is correlated with agricultural areas globally [119, 229, 249], as well as with the gross value added by agriculture in Germany [114]. This is expected, as large-scale PV installations are often sited on agricultural land.

These studies show that the relationship between solar PV deployment and geographical factors is complex and varies depending on the geographical region and analysis resolution (national, regional, subregional). These factors are crucial in determining where and how solar PV systems are installed. By incorporating them into a comprehensive model, we can better predict regional solar PV deployment patterns, identify areas with high potential for future installations, and uncover barriers to deployment in underserved regions.

None of the existing studies comprehensively account for all types of solar PV capacity

(residential, commercial, and utility-scale) installed in a country. A key challenge in modelling regional and subregional PV capacity is the limited availability of relevant social, economic, climatic, and land use data. This study introduces the first comprehensive regional PV capacity model that disaggregates national capacity across 168 NUTS 3 regions. The model will be used to allocate unknown capacity to specific geographic regions, as a benchmarking tool for assessing regional performance, and to serve as a forecasting tool.

5.3 Methodology

We previously analysed the relationship between social, economic, climatic, and land use factors and national solar PV deployment, showing these variables can effectively model national PV capacity [229]. To identify factors associated with regional solar PV deployment in Great Britain, we apply a similar methodology at the NUTS 3 level, examining correlations between geographical factors and PV capacity using Pearson and Spearman correlation coefficients. We also assess linear relationships using the coefficient of determination (R^2). The top correlated features are then used to train an XGBoost model to estimate regional capacity percentages for 168 GB regions from 2010 to 2023.

5.3.1 Data and Data Processing

Data for this analysis were obtained from publicly accessible sources. Climate variables were sourced from the Copernicus Climate Change Service [230]. Land use data were derived from the CORINE Land Cover 2018 dataset [231]. Economic indicators, such as regional gross value added (GVA) by industry, were acquired from the UK's Office for National Statistics [272]. Renewables Obligation Certificates (ROCs) buy-out prices and obligation levels were obtained from the Office of Gas and Electricity Markets (Ofgem) [273]. Solar PV capacity data were provided by Sheffield Solar [274].

The analysis was conducted within the framework of the NUTS 2021 (Nomenclature of Territorial Units for Statistics) classification, a hierarchical system used to delineate European Union territories for statistical purposes [234]. The study focused on NUTS 3 regions, which

accommodate populations ranging from 150,000 to 800,000, as this level offers optimal data availability in Great Britain. All data were processed according to the NUTS 3 classification and aggregated on an annual basis.

The CORINE land cover dataset uses a three-tier hierarchical classification system. Level 3 offers detailed analysis with 44 thematic classes and has a thematic accuracy of over 85%, leading to potential misclassification [235]. Higher levels (Levels 1 and 2) have fewer categories, improving accuracy and reducing misclassification risk [236, 237, 238], but provide less detail. To balance detail and accuracy, our analysis uses all three levels.

Capacity data were processed at both NUTS 3 and NUTS 0 levels, with the regional capacity distribution presented in Figure 5.2a. GVA data were directly available at the International Territorial Level (ITL), which corresponds to NUTS 3 regions [275]. Climate data, originally available at the NUTS 2 level, were assumed to be uniform across all NUTS 3 regions within the same NUTS 2 area.

Data normalization was applied according to the variable type. Climate data were normalized to national averages to emphasize significant regional climatic variations. Economic indicators, such as GVA, were normalized to national totals. Regional solar PV capacities were expressed as a proportion of the national total capacity.

5.3.2 Feature Selection

The feature selection followed the DAMA data quality framework [239], which focuses on accuracy, completeness, uniqueness, consistency, timeliness, and validity. Priority was given to features that were complete for most regions and consistent over time. Selection criteria included correlation with PV capacity and data availability. To capture both linear and non-linear associations, we averaged Spearman and Pearson correlations. Various correlation thresholds (0.2 to 0.5) were tested, with a data availability threshold of 90%.

5.3.3 Model Training

As a benchmark, we first implemented a multiple linear regression (MLR) model trained on the same features as the XGBoost model, enabling a direct comparison between linear and nonlinear approaches. The linear model was implemented using `scikit-learn`, a Python library that provides a wide range of machine-learning algorithms for supervised and unsupervised tasks [214]. The MLR model is expressed as shown in Equation (5.1).

$$y = \beta_0 + \sum_{i=1}^p \beta_i x_i + \varepsilon, \quad (5.1)$$

where y is the target variable, β_0 is the intercept, β_i are the model coefficients, x_i are the predictor variables for $i = 1, \dots, p$, and ε is the error term capturing the discrepancy between the observed and predicted values. Parameters were estimated using Ordinary Least Squares (OLS). To ensure robustness, GroupKFold cross-validation was applied with grouping by year, where each fold corresponded to leaving out one year from the training window (2010–2019). The final model was then evaluated on an independent test set (2020–2022). The year 2023 was excluded due to missing data for some variables. The MLR is reported as an unscaled model. In this context, an unscaled model refers to using the raw regional predictions directly, without adjusting them so that their sum matches the corresponding national total. Scaling requires complete regional coverage for every year, which was not available in this case, and therefore could not be applied to the MLR.

We then selected the extreme gradient boosting (XGBoost) algorithm as our primary modelling approach due to its advantages: it is nonparametric, making it well-suited for non-normally distributed data [241], and it performs particularly well with tabular datasets [242], which aligns with the structure of our data. Data were grouped by year, with 2010–2020 used for training and 2021–2023 for testing. An XGBoost model was trained, with hyperparameters fine-tuned using grid search and GroupKFold cross-validation, where each fold corresponded to holding out one year of data for validation. The trained model predicts the percentage of solar PV capacity within a NUTS 3 region and is referred to as the "unscaled

model”. To ensure regional predictions sum to the national capacity, we scaled the regional predictions so that their total matched the national value. This approach is referred to as the ”scaled model”. A further advantage of XGBoost is its ability to handle missing data, which is particularly important in this context, as scaling requires predictions for all regions in each year.

For the XGBoost model, we used Shapley Additive Explanations (SHAP) values to measure feature importance. SHAP is a game-theoretic approach designed to explain the output of any machine learning model, where the values are additive [247].

To evaluate similarity among the selected predictors, we applied the VARCLUS algorithm in JMP Pro [188, 189] to identify clusters of related variables. Where clusters were detected, we examined their composition and structure to provide additional context for interpreting model performance.

Model performance was evaluated using Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). We avoid using MAPE when evaluating regional capacity percentage because it can yield large values due to data representation in percentages. Instead, MAPE is specifically applied to evaluate national capacity. Metrics are reported for both the benchmark MLR model and the XGBoost models, enabling a clear assessment of the value added by the nonlinear approach.

5.3.4 Model Applications

The models can be applied to disaggregate national capacity into regional capacity, to allocate capacity with an unknown location to a geographical region, and to serve as a benchmarking tool. A Solar PV Deployment Index (SPVDI) was developed previously to assess national capacity in a global context [229]. In this study, we adapt the same concept to evaluate regional PV capacity within GB.

The SPVDI serves as a benchmarking tool by comparing solar PV deployment in a region relative to other regions with similar social, economic, climatic, and land use characteristics.

It is calculated as shown in Equation (5.2), where t_{start} represents the initial year and t_{end} the final year of analysis. The index sums the difference between actual and predicted capacity for region i over the specified time range. The resulting index value indicates whether a region's capacity is above or below expectations, with a positive value indicating higher-than-expected capacity and a negative value indicating lower-than-expected capacity. The SPVDI allows for performance comparisons of regions over multiple years and time periods and serves as a tool to rank regions based on their solar PV deployment.

$$\text{SPVDI}_i = \sum_{j=t_{\text{start}}}^{t_{\text{end}}} (\text{Actual Capacity}_{i,j} - \text{Predicted Capacity}_{i,j}) \quad (5.2)$$

5.4 Results and Discussion

Appendix 5.6.1 presents the analysis results for the normalized regional data, with features ranked from highest to lowest based on their average correlation. GVA by veterinary activities emerges as the most strongly correlated feature. Additionally, regions with higher percentages of agricultural areas and artificial surfaces tend to have a greater share of PV capacity.

Interestingly, climatic factors such as GHI ranks 18, while air temperature ranks 24 in terms of average correlation. The reason climatic factors rank lower than other factors may be attributed to threshold effects. Regions need to meet certain levels of solar resource potential to justify investments in PV systems. Once these thresholds are reached, other factors - such as land availability or socioeconomic characteristics - become more critical in determining the proportional share of PV deployment. Previous studies have identified a radiation threshold of 4.5 kWh/m²/day above which deployment is triggered [98].

The strong correlation between artificial surfaces and solar PV capacity can be attributed to the inclusion of energy production and distribution infrastructure within this category [231]. These features could act as proxies for grid connection points, which are crucial for integrating grid-connected solar PV systems. Regions with more access to such infrastructure are better positioned to support large-scale solar installations. Furthermore, the presence of industrial agricultural facilities, which are also classified as artificial surfaces, likely con-

tributes to the observed correlation with PV capacity. The correlation between agricultural land and solar PV capacity is strong. This is expected since commercial, industrial, and utility-scale solar PV projects are predominantly situated on agricultural land, particularly arable land [249, 119].

GVA by veterinary activities is correlated with artificial surfaces and agricultural land, which might explain why it ranks the highest in terms of correlation with PV capacity. This could also be linked to the presence of farm animals, as veterinary activities are more likely to be concentrated in regions with higher demand, such as areas with farms and extensive agricultural land.

The same set of predictors was used for both the benchmark MLR and the XGBoost models to enable a direct comparison. The selected features were: GVA by veterinary activities, arable land (21), non-irrigated arable land (211), urban fabric (11), artificial surfaces (1), discontinuous urban fabric (112), agricultural Areas (2), sport and leisure facilities (142), artificial, non-agricultural vegetated areas (14), and mine, dump and construction sites (13).

Table 5.1 shows the regional error metrics for the MLR and XGBoost unscaled and scaled models. For the MLR model, performance was modest, with the model explaining around 67–68% of the variation in regional PV capacity and achieving a MAE of 44 MW and RMSE of 68 MW on the test set. This highlights its value as a stable but relatively weak benchmark. In contrast, the XGBoost models substantially improved predictive performance, explaining about 89–91% of the variation, with a test MAE of about 20 MW and RMSE of 41 MW. Scaling did not significantly alter the regional results in the XGBoost model, although it is still applied in cases where national capacity is disaggregated to ensure consistency with reported totals.

Figure 5.1 shows the actual vs. estimated regional capacity for the XGBoost unscaled and scaled models. The unscaled model performs well, and scaling does not make any significant changes to the regional estimates.

Table 5.2 shows the national error metrics for the XGBoost unscaled and scaled models. The unscaled model has an error of 5.4% in estimating national capacity. Interestingly, this

Table 5.1: Regional error metrics for the XGBoost and MLR models, reported separately for training and test sets in both megawatts (MW) and percentages (%). For the Scaled XGBoost model, regional predictions are adjusted to ensure they sum to the national capacity. Metrics reported are the coefficient of determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

Model	Train R^2	Test R^2	Train MAE	Test MAE	Train MSE	Test MSE	Train RMSE	Test RMSE
MLR Unscaled (MW)	0.67	0.68	24.11	44.39	2208	4592	46.99	67.76
XGBoost Unscaled (MW)	0.99	0.89	5.60	20.27	93.68	1701	9.68	41.25
XGBoost Scaled (MW)	0.99	0.90	5.66	19.94	87.94	1610	9.38	40.13
MLR Unscaled (%)	0.60	0.68	0.33	0.31	0.29	0.31	0.54	0.48
XGBoost Unscaled (%)	0.98	0.90	0.09	0.13	0.02	0.07	0.13	0.26
XGBoost Scaled (%)	0.98	0.91	0.09	0.13	0.02	0.06	0.14	0.25

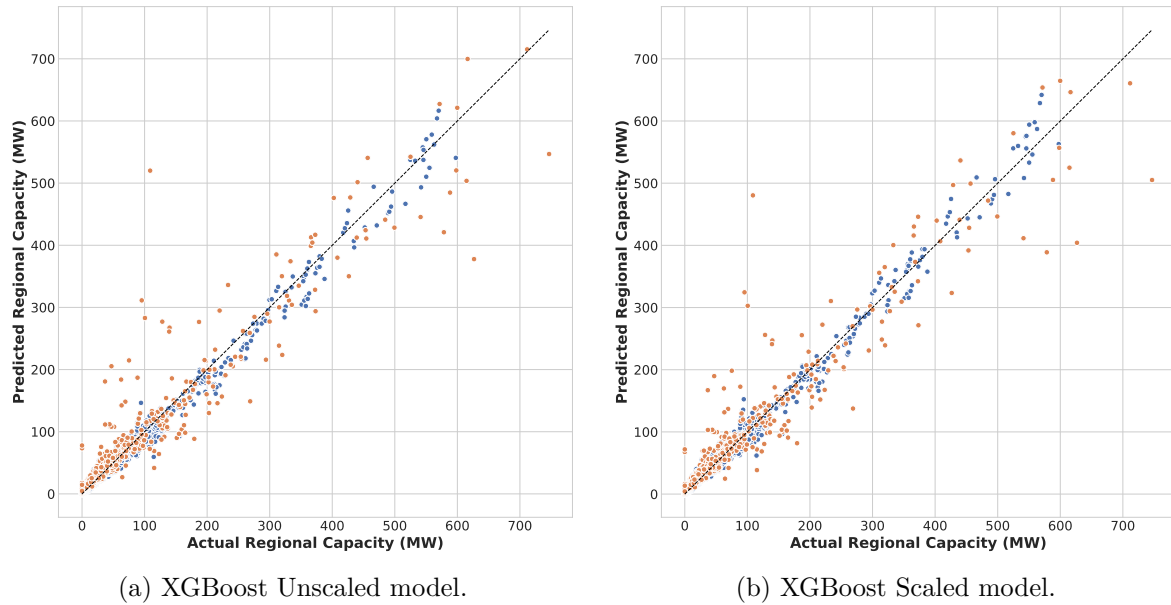


Figure 5.1: Actual versus predicted regional photovoltaic capacity using the XGBoost model for 168 NUTS 3 regions in Great Britain, covering the years 2010 to 2023. Blue points represent training data, while orange points represent test data. Panel (a) shows the unscaled model; panel (b) shows the scaled model, where predictions are adjusted to match national totals.

Table 5.2: National error metrics for the XGBoost unscaled and scaled models, evaluated over the entire dataset. These metrics compare the actual national capacity with the estimated national capacity, which is derived from summing the regional capacities. Scaling refers to adjusting regional capacities so that their sum matches the actual national capacity.

Metric	R^2	MAE (MW)	MSE (MW²)	RMSE (MW)	MAPE (%)
XGBoost Unscaled Model (entire dataset)	0.99	441	322000	568	5.4
XGBoost Scaled Model (entire dataset)	1	0	0	0	0.0

is very similar to the 5% error reported by [49].

The feature importance analysis for the XGBoost model reveals that non-irrigated arable land (211) accounts for 36% of the total SHAP values, followed by discontinuous urban fabric (112) at 21%, artificial surfaces (1) at 12%, sport and leisure facilities (142) at 7%, agricultural areas (2) at 7%, GVA by veterinary activities at 6%, urban fabric (11) at 5%, mine, dump, and construction sites (13) at 4%, non-agricultural vegetated areas (14) at 2%, and arable land (21) at 0%. Furthermore, the analysis highlights that artificial surfaces have the most significant overall impact, accounting for 51% of the total SHAP values. This is followed by agricultural land, which contributes 43%, and GVA by veterinary services, contributing 6%.

Clustering the selected features with the VARCLUS algorithm revealed two distinct groups. The first cluster includes artificial surfaces (1), discontinuous urban fabric (112), urban fabric (11), non-agricultural vegetated areas (14), sport and leisure facilities (142), and GVA by veterinary services. The second cluster includes arable land (21), non-irrigated arable land (211), agricultural areas (2), and mine, dump, and construction sites (13). The most representative variable in the first cluster is artificial surfaces (1), which explains 96% of the variation within the cluster. For the second cluster, the most representative variable is arable land (21), which explains 92% of the variation in the cluster. These findings align with the results above, which highlight artificial surfaces and agricultural land as the most significant factors influencing PV deployment at the NUTS 3 level.

5.4.1 Model Applications

Figure 5.2a illustrates the actual regional capacities for 2023. In some cases, solar PV system locations are not recorded. Figure 5.3 shows the unallocated capacity per year from 2010 to 2023. Initially, this wasn't an issue due to the low capacity, but as it grows, it becomes more difficult to monitor generation accurately. As PV capacity increases, it becomes increasingly important to know the geographical distribution of these capacities due to their impact on grid stability. For example, in 2023, these unrecorded systems accounted for 829 MW in GB. The scaled XGBoost model was used to allocate these capacities, leading to updated regional capacity estimates, as depicted in Figure 5.2b. We provide a dataset of regional PV capacities for 168 GB regions spanning from 2010 to 2023, enabling more accurate generation monitoring at the regional level. Unlike simple allocation methods that distribute PV capacity based on the percentage of deployment per region, our model accounts for the unique geographical factors that affect capacity diffusion. This approach addresses systematic issues, such as inconsistent reporting, under-reporting in less monitored regions, and challenges in capturing small-scale installations.

The unscaled XGBoost model serves as a valuable benchmarking tool for regional solar PV deployment. Figure 5.2c shows the solar PV deployment index from 2010 to 2023, highlighting the top and bottom regions based on capacity differences. Cambridgeshire CC (UKH12) stands out, exceeding expected capacity by 780 MW, while Barnsley, Doncaster, and Rotherham (UKE31) have the largest deficit, falling short by -985 MW. These findings align with those of Collier et al. [107], who modelled small-scale domestic solar PV deployment in England and Wales at the LSOA level, and observed that local authority districts within Cambridgeshire CC (UKH12) exhibited higher than expected capacities.

Figure 5.4 shows the SHAP analysis of the average contribution of the features from 2010 to 2023 in Cambridgeshire, and Barnsley, Doncaster, and Rotherham. The unscaled XGBoost model's average prediction is 0.56% of national capacity, as determined by SHAP analysis. In Cambridgeshire, the SHAP analysis of feature contributions from 2010 to 2023 reveals that non-irrigated arable land (211) is the largest contributor to the prediction, with

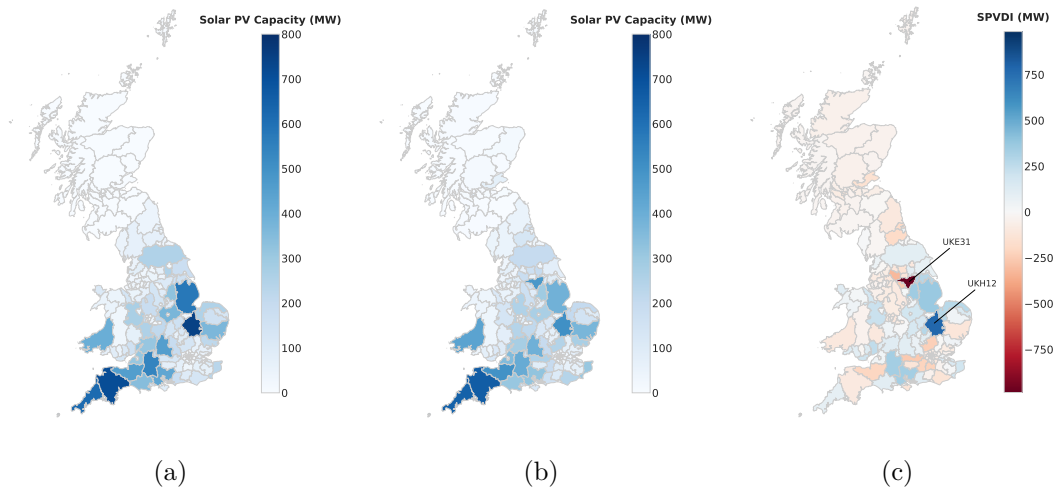


Figure 5.2: (a) Actual regional solar photovoltaic capacity across 168 NUTS 3 regions in Great Britain for the year 2023. (b) Predicted regional solar photovoltaic capacity across 168 NUTS 3 regions in Great Britain for the year 2023 using the scaled XGBoost model. (c) Solar PV Deployment Index (SPVDI) calculated using the XGBoost model predictions for the years 2010 to 2023. Positive values indicate regions where the actual capacity exceeds the expected capacity, while negative values indicate regions where the actual capacity is less than expected. The regions included are Barnsley, Doncaster and Rotherham (UKE31) which under deployed by 985 MW and Cambridgeshire CC (UKH12) which over deployed by 780 MW. *Note: Map lines delineate study areas and do not necessarily depict accepted national boundaries.*

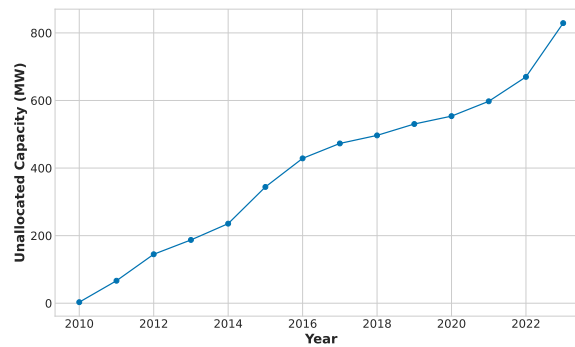


Figure 5.3: Unallocated solar photovoltaic capacity for 168 NUTS 3 regions in Great Britain from the year 2010 to 2023.

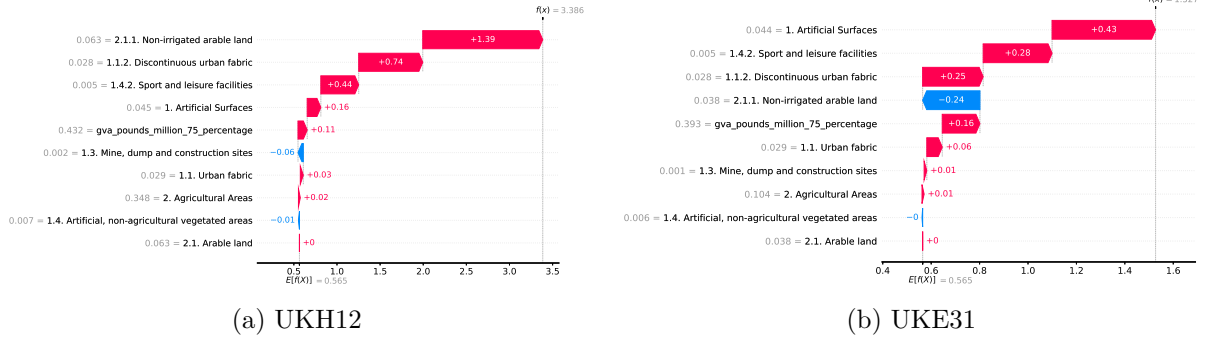


Figure 5.4: SHAP analysis for the unscaled XGBoost model, showing the mean contribution of factors to solar PV capacity percentage between 2010 and 2023 for (a) Cambridgeshire CC (UKH12) and (b) Barnsley, Doncaster, and Rotherham (UKE31). In the SHAP framework, $E[f(x)]$ represents the model’s average prediction, while $f(x)$ is the predicted value, calculated as the sum of the contributions of the features and the model’s average prediction.

a mean contribution of 1.39%. Other contributors include discontinuous urban fabric (112) at 0.74%, sport and leisure facilities (142) at 0.44%, artificial surfaces (1) at 0.16%, GVA by veterinary services at 0.11%, construction sites (13) at -0.06%, urban fabric (11) at 0.03%, agricultural areas (2) at 0.02%, non-agricultural vegetated areas (14) at -0.01%, and arable land (21) at 0%. This results in an average predicted capacity for Cambridgeshire of 3.39% of the national total, equivalent to an average capacity of 341 MW between 2010 and 2023. The high prediction for Cambridgeshire is primarily driven by non-irrigated arable land (211), but other factors, such as access to grid connection points, could also explain why this region has more capacity than expected. We hypothesize that the region’s southern location, the prevalence of agricultural areas, and its proximity to transmission lines - where fewer constraints exist compared to other regions - may play a significant role. Additionally, social effects, such as peer influence, could help explain the higher than expected capacity in this area.

Based on SHAP analysis of the mean contribution of features between 2010 and 2023 in Barnsley, Doncaster, and Rotherham, the largest contributor is artificial surfaces (1) with a contribution of 0.43%. This is followed by sport and leisure facilities (142) at 0.28%, discontinuous urban fabric (112) at 0.25%, and non-irrigated arable land (211) at -0.24%. Other contributors include GVA by veterinary services at 0.16%, urban fabric (11) at 0.06%,

construction sites (13) at 0.01%, agricultural areas (2) at 0.01%, non-agricultural vegetated areas (14) at 0%, and arable land (21) at 0%. This results in an average prediction for Barnsley, Doncaster, and Rotherham of 1.53% of national capacity, equivalent to an average capacity of 141 MW between 2010 and 2023.

The SHAP analysis indicates that the high prediction for the region is primarily driven by artificial surfaces. The predominance of artificial surfaces in this region suggests that solar PV systems are primarily residential and commercial. This observation aligns with findings by Westacott and Candelise [110], which identify domestic installations as the dominant source of capacity in the area. However, the region has 958 MW less capacity than expected, suggesting that domestic systems alone cannot explain the shortfall. Since agricultural areas are excluded based on SHAP analysis, the gap is likely attributable to commercial systems.

To explain the shortfall, there may be unique characteristics of artificial surfaces in this region that differentiate it from others. For example, variations in business ownership or type could play a role. Westacott and Candelise [110] highlights that the non-domestic rooftop market is significantly underdeveloped compared to the domestic and ground-mounted markets in the UK. Additionally, grid connection constraints for commercial installations and differences in the planning permission process may further influence deployment in this region.

Finally, Barnsley, Doncaster, and Rotherham’s gross disposable household income (GDHI) per head, measured as indices where the UK equals 100, ranged from 79.6 to 81.3 between 2010 and 2022 [272]. This consistently lower GDHI compared to the national average may help explain the lower than expected solar PV capacity, as reduced disposable income could limit investment in domestic installations. GDHI is excluded from the model because the data is typically published with a delay of about two years, meaning data for 2021 was only made available in September 2023.

The XGBoost models serve three main purposes. First, they enable the disaggregation of national PV capacity, with the scaled model being the most appropriate for this task.

This is particularly useful when there is a significant amount of unallocated capacity at the national level. This can help with more accurate generation monitoring. Second, the models facilitate the benchmarking of regional capacities within GB using the unscaled model. This approach allows for ranking regions based on their PV deployment, helping to identify areas that may require interventions, such as improving access to grid connection points or policy adjustments. Finally, the scaled model can be used for forecasting capacity. This is beneficial for strategic grid expansion planning and can help in addressing land use conflicts by pinpointing regions where capacity is likely to be concentrated.

5.5 Conclusion

This study provides a comprehensive analysis of regional solar PV capacity in Great Britain, identifying artificial surfaces and agricultural areas as the most significant factors influencing deployment at the NUTS 3 regional level. To address the need for detailed capacity estimates, we developed a model that explains 89% of the variation in PV capacity, with a mean absolute error (MAE) of 20 MW and a national mean absolute percentage error (MAPE) of 5.4%. Additionally, we provide a dataset of regional capacity estimates for NUTS 3 regions in Great Britain, covering the years 2010 to 2023.

The XGBoost models serve three primary purposes: disaggregating national PV capacity into regional capacity, benchmarking regional solar PV deployment, and forecasting future solar PV deployment. They offer practical applications for grid operators by improving generation monitoring, which requires precise knowledge of the geographical distribution of capacity. The models also support long-term grid network planning by identifying where capacity is likely to be distributed and enabling targeted grid expansions in these areas. Furthermore, they can help pinpoint regions where deployment falls short of expectations, offering insights into where interventions, such as policy adjustments or infrastructure investments, may be needed.

Future research could explore regions where solar PV capacity significantly exceeds or falls below expectations. These case studies could provide valuable insights into barriers to

deployment and identify historical incentives or policies that have successfully driven higher deployment.

Overall, this study offers valuable tools and data to enhance operational and strategic decision-making for Great Britain's solar PV sector and lays a foundation for further research into regional capacity estimation and renewable energy planning.

Data Availability

The supplementary code and data supporting this manuscript are available at <https://doi.org/10.5281/zenodo.17178404>.

5.6 Appendix

5.6.1 Normalized data analysis

Table 5.3: Normalized data analysis. Features considered for modelling the percentage of regional solar photovoltaic capacity. All features presented are normalized relative to national values and expressed as percentages. The data availability, coefficient of determination (R^2), Pearson correlation, Spearman correlation, correlation average are shown. Relevant literature that explores similar features is cited. Land cover features are preceded by their CORINE Land Cover (CLC) classification codes (e.g., 1.2.1 for industrial or commercial units).

Feature	Data Availability (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
GVA by veterinary activities	93	0.41	0.64***	0.63***	0.64	
2.1. Arable land	100	0.43	0.66***	0.56***	0.61	[249]
2.1.1. Non-irrigated arable land	100	0.43	0.66***	0.56***	0.61	[249]
1.1. Urban fabric	100	0.33	0.58***	0.62***	0.60	[229]
1. Artificial Surfaces	100	0.37	0.61***	0.59***	0.60	
1.1.2. Discontinuous urban fabric	100	0.34	0.58***	0.62***	0.60	
2. Agricultural Areas	100	0.44	0.66***	0.53***	0.59	[119, 229]
1.4.2. Sport and leisure facilities	100	0.31	0.55***	0.55***	0.55	
1.4. Artificial, non-agricultural vegetated areas	100	0.27	0.52***	0.55***	0.54	
1.3. Mine, dump and construction sites	100	0.28	0.53***	0.48***	0.51	
GVA by Production sector	93	0.19	0.44***	0.55***	0.49	
1.2.4. Airports	100	0.33	0.58***	0.38***	0.48	

Feature	Data Availability (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
3.1.1. Broad-leaved forest	100	0.21	0.45***	0.52***	0.48	
GVA by manufacturing	93	0.15	0.38***	0.55***	0.47	[229]
GVA by manufacture of electronic, optical and electrical products	93	0.17	0.41***	0.53***	0.47	
1.3.1. Mineral extraction sites	100	0.22	0.47***	0.43***	0.45	
Mean sea level pressure	90	0.14	0.37***	0.52***	0.45	
GHI	90	0.13	0.37***	0.51***	0.44	[110, 98, 103, 115, 117, 229]
GVA by agriculture, forestry and fishing; mining and quarrying	93	0.16	0.4***	0.46***	0.43	[114, 229]
3.1.3. Mixed forest	100	0.20	0.45***	0.42***	0.43	
GVA by specialised construction activities	93	0.14	0.38***	0.48***	0.43	
GVA by manufacture of wood and paper products and printing	93	0.14	0.38***	0.49***	0.43	
1.2. Industrial, commercial and transport units	100	0.20	0.44***	0.4***	0.42	
Air temperature	90	0.14	0.37***	0.47***	0.42	[98, 229]
2.3. Pastures	100	0.18	0.43***	0.41***	0.42	
GVA by activities of households	93	0.17	0.41***	0.43***	0.42	
GVA by residential care activities	93	0.15	0.38***	0.47***	0.42	
2.3.1. Pastures	100	0.18	0.43***	0.41***	0.42	
1.3.3. Construction sites	100	0.10	0.32***	0.44***	0.38	
GVA by other personal service activities	93	0.11	0.33***	0.43***	0.38	

Feature	Data Availability (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
GVA by motor trades	93	0.05	0.22***	0.53***	0.38	
GVA by other manufacturing, repair and installation	93	0.07	0.27***	0.47***	0.37	
2.4. Heterogeneous agricultural areas	100	0.16	0.4***	0.32***	0.36	
Region area	100	0.09	0.29***	0.43***	0.36	[107, 105]
GVA by manufacture of machinery and transport equipment	93	0.04	0.19***	0.51***	0.35	
GVA by owner-occupiers' imputed rental	93	0.10	0.32***	0.38***	0.35	
GVA by construction	93	0.09	0.3***	0.4***	0.35	
GVA by manufacture of petroleum, chemicals and other minerals	93	0.04	0.21***	0.5***	0.35	
GVA by manufacture of basic and fabricated metal products	93	0.04	0.19***	0.49***	0.34	
2.4.3. Land principally occupied by agriculture, with significant areas of natural vegetation	100	0.14	0.38***	0.3***	0.34	
1.2.1. Industrial or commercial units	100	0.12	0.34***	0.34***	0.34	
2.4.2. Complex cultivation patterns	100	0.09	0.3***	0.33***	0.32	
GVA by wholesale and retail trade; repair of motor vehicles	93	0.04	0.21***	0.4***	0.30	
GVA by manufacture of food, beverages and tobacco	93	0.10	0.31***	0.29***	0.30	
GVA by accommodation	93	0.05	0.23***	0.34***	0.29	
GVA by social work activities	93	0.05	0.23***	0.34***	0.29	
GVA by retail trade	93	0.04	0.2***	0.38***	0.29	
Wind speed	90	0.09	0.3***	0.29***	0.29	[98]
GVA by accommodation and food service activities	93	0.04	0.19***	0.36***	0.28	

Feature	Data Availability (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
4.1.1. Inland marshes	100	0.07	0.27***	0.28***	0.28	
GVA by civil engineering	93	0.05	0.22***	0.35***	0.28	
GVA by real estate activities	93	0.05	0.22***	0.35***	0.28	
GVA by electricity, gas, water; sewerage and waste management	93	0.06	0.24***	0.31***	0.28	
GVA by other service activities	93	0.02	0.13***	0.4***	0.27	
GVA by education	93	0.06	0.24***	0.3***	0.27	
GVA by wholesale trade	93	0.02	0.15***	0.38***	0.27	
GVA by postal and courier activities	93	0.03	0.18***	0.36***	0.27	
GVA by public administration and defence	93	0.06	0.23***	0.29***	0.26	
GVA by human health and social work activities	93	0.05	0.22***	0.28***	0.25	
GVA by construction of buildings	93	0.03	0.17***	0.33***	0.25	
2.2.2. Fruit trees and berry plantations	100	0.01	0.08***	0.4***	0.24	
GVA by food and beverage service activities	93	0.02	0.15***	0.33***	0.24	
2.2. Permanent crops	100	0.01	0.08***	0.4***	0.24	
1.3.2. Dump sites	100	0.06	0.25***	0.21***	0.23	
GVA by all industries	93	0.01	0.11***	0.36***	0.23	[229, 116]
GVA by architectural and engineering activities	93	0.02	0.14***	0.31***	0.23	
4.2.1. Salt marshes	100	0.05	0.23***	0.23***	0.23	
3.1. Forests	100	0.00	0.06**	0.39***	0.23	
2.4.4. Agro-forestry areas	100	0.08	0.28***	0.13***	0.21	

Feature	Data Availability (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
GVA by other professional, scientific and technical activities	93	0.01	0.1***	0.3***	0.20	
GVA by services sector	93	0.00	0.06**	0.32***	0.19	
GVA by rental and leasing activities	93	0.01	0.08***	0.28***	0.18	
GVA by human health activities	93	0.02	0.14***	0.2***	0.17	
4.2. Marine wetlands	100	0.06	0.24***	0.09***	0.16	
GVA by real estate activities, excluding imputed rental	93	0.00	0.04*	0.28***	0.16	
GVA by repair of computers, personal and household goods	93	0.00	0.06**	0.24***	0.15	
1.2.2. Road and rail networks and associated land	100	0.01	0.1***	0.2***	0.15	
GVA by transportation and storage	93	0.01	0.07***	0.24***	0.15	
GVA by activities of membership organisations	93	0.00	-0.03	0.32***	0.15	
4.1.2. Peat bogs	100	0.01	-0.1***	-0.21***	-0.15	
GVA by head offices and management consultancy	93	0.00	-0.01	0.3***	0.14	
GVA by services to buildings and landscape activities	93	0.01	0.08***	0.21***	0.14	
GVA by professional, scientific and technical activities	93	0.00	-0.0	0.28***	0.14	
3.3.3. Sparsely vegetated areas	100	0.01	-0.11***	-0.16***	-0.14	
Total precipitation	100	0.00	-0.04*	-0.23***	-0.14	[229]
GVA by administrative and support service activities	93	0.00	0.02	0.24***	0.13	
GVA by warehousing and transport support activities	93	0.00	0.04	0.22***	0.13	
GVA by gambling and betting; sports and recreation activities	93	0.00	0.04	0.2***	0.12	
4.2.3. Intertidal flats	100	0.03	0.18***	0.07**	0.12	
GVA by land, water and air transport	93	0.00	0.04*	0.2***	0.12	

Feature	Data Availability (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
GVA by office administration and business support activities	93	0.00	-0.0	0.23***	0.12	
3.3.4. Burnt areas	100	0.01	-0.07***	-0.16***	-0.12	
3. Forest And Seminatural Areas	100	0.00	-0.05*	0.26***	0.11	
GVA by creative, arts, entertainment and cultural activities	93	0.00	-0.03	0.25***	0.11	
5.2.1. Coastal lagoons	100	0.02	0.15***	0.07***	0.11	
5. Water Bodies	100	0.00	-0.06**	0.29***	0.11	
GVA by research and development; advertising and market research	93	0.00	0.02	0.18***	0.10	
3.3.1. Beaches, dunes, sands	100	0.00	0.06**	0.13***	0.10	
GVA by legal and accounting activities	93	0.00	-0.06**	0.26***	0.10	
GVA by telecommunications; information technology	93	0.00	0.01	0.19***	0.10	
GVA by arts, entertainment and recreation	93	0.00	-0.0	0.2***	0.10	
GVA by manufacture of textiles, wearing apparel and leather	93	0.00	0.06**	0.15***	0.10	
3.1.2. Coniferous forest	100	0.01	-0.08***	0.28***	0.10	
3.3.2. Bare rocks	100	0.01	-0.09***	-0.1***	-0.10	
5.1.2. Water bodies	100	0.01	-0.07***	0.25***	0.09	
GVA by employment activities; tourism and security services	93	0.00	-0.03	0.2***	0.09	
5.1. Inland waters	100	0.00	-0.07***	0.25***	0.09	
5.2.2. Estuaries	100	0.03	0.17***	-0.01	0.08	
GVA by information and communication	93	0.00	-0.02	0.18***	0.08	
GVA by financial service activities	93	0.01	-0.07***	0.24***	0.08	

Feature	Data Availability (%)	R-squared	Pearson Correla- tion	Spearman Correla- tion	Correlation Average	Relevant Literature
GVA by financial and insurance activities	93	0.00	-0.07**	0.24***	0.08	
GVA by publishing; film and TV production and broadcasting	93	0.00	-0.07**	0.2***	0.07	
GVA by insurance, pension funding and auxiliary financial activities	93	0.00	-0.06**	0.21***	0.07	
3.3. Open spaces with little or no vegetation	100	0.01	-0.1***	-0.02	-0.06	
4.1. Inland wetlands	100	0.01	-0.1***	-0.02	-0.06	
1.2.3. Port areas	100	0.00	-0.0	0.1***	0.05	
4. Wetlands	100	0.01	-0.09***	-0.01	-0.05	
1.4.1. Green urban areas	100	0.01	-0.11***	0.06**	-0.03	
3.2.4. Transitional woodland-scrub	100	0.01	-0.11***	0.15***	0.02	
3.2.2. Moors and heathland	100	0.01	-0.1***	0.06**	-0.02	
5.1.1. Water courses	100	0.01	0.09***	-0.06**	0.01	
3.2. Scrub and/or herbaceous associations	100	0.01	-0.09***	0.11***	0.01	
1.1.1. Continuous urban fabric	100	0.00	-0.04	0.03	-0.01	
3.2.1. Natural grassland	100	0.00	-0.06**	0.03	-0.01	
5.2.3. Sea and ocean	100	0.00	-0.03	0.0	-0.01	
5.2. Marine waters	100	0.00	-0.0	0.0	0.00	

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Chapter 6

Concluding Discussion

This thesis presents a comprehensive investigation into the factors influencing solar PV deployment across multiple geographic scales. The development of three distinct models—the Global Model (national scale), the European (EU) Model (regional/NUTS 2 scale), and the Great Britain (GB) Model (subregional/NUTS 3 scale)—has enabled a nuanced understanding of the spatial dynamics of solar PV deployment. Each model supports one or more critical use cases: disaggregation, benchmarking, and forecasting.

In the global model, which examines national level solar PV deployment across countries, the most important explanatory variables include the previous year’s cumulative PV capacity, population, agricultural land area, tertiary education levels, and electricity net consumption. Among the 36 features analysed, economic factors emerged as the most influential, followed by social factors—particularly education. Land-use characteristics played a notable but less dominant role, while climatic factors, such as solar radiation, were not found to be significant.

It is intuitive that economic factors rank highest in explaining national solar PV deployment. Countries with stronger economies are better positioned to invest in renewable energy infrastructure and implement large-scale policies that support solar PV adoption. Historically, the high upfront costs of PV systems meant that only wealthier nations could afford to support deployment through public investment or incentives. Liu et al. [116] identified GDP, final consumer expenditure, and industrial added value as strong predictors of solar

PV capacity, reinforcing this link between economic strength and deployment. However, the relationship is not universal—Celik and Özgür [117] reported cases where GDP was not significantly correlated with PV capacity, suggesting that institutional or policy contexts also play a role.

Social factors, particularly the level of tertiary education, also contribute to explaining PV deployment. Education may serve as a proxy for social awareness, administrative capacity, and the ability to manage complex regulatory and financial mechanisms associated with renewable energy transitions. Land use also influences deployment, as agricultural land is frequently used for siting PV installations [119]. Nonetheless, while solar radiation is often assumed to be a driver, Celik and Özgür [117] found that solar radiation itself is not a predictor of national deployment—likely due to overriding economic and institutional constraints. Public incentives can drive adoption regardless of solar resource availability [118], underscoring the central role of policy in shaping deployment outcomes.

In the European model, which investigates solar PV deployment at the regional level, artificial surfaces and agricultural areas emerge as the most important predictors of installed capacity. Unlike at the national scale, where economic and social factors dominate, land-use characteristics play a more prominent role in shaping regional deployment patterns. Artificial surfaces likely act as a proxy for access to grid infrastructure and urban development, and are often associated with higher densities of rooftop solar PV, particularly in residential and commercial zones. Agricultural land, by contrast, is typically associated with utility-scale installations [119] due to its flat terrain, open space, and generally higher levels of solar irradiation.

The relationship between land use and PV deployment is supported by multiple studies. Westacott and Candélise [110] found that rural regions with high irradiation are more likely to host PV installations. Similarly, Thormeyer, Sasse, and Trutnevyte [111] highlight the importance of the urban–rural divide, showing that municipalities with a focus on agriculture and forestry typically see higher levels of solar deployment, while more urbanised municipalities tend to lag. Mayer et al. [114] also reported a strong positive correlation between

agricultural gross value added and PV adoption, suggesting that solar PV is often integrated into regions with active agricultural economies.

Climatic factors, including solar irradiation, show slightly stronger correlations at the regional level than they do nationally, but still remain weaker predictors overall. While resource availability is necessary, it is not sufficient to drive deployment without economic and infrastructural support. For example, Rigo et al. [112] found that electricity tariffs were the strongest predictor of municipal PV adoption, followed by solar irradiation and GDP. Additional regional predictors included education, company presence, demographic density, and vehicle fleet size—highlighting the relevance of socioeconomic structures in shaping adoption patterns. Moreover, McEachern and Hanson [113] showed that remote, economically disadvantaged areas with strong solar resources often turn to PV as a substitute for grid electricity.

These patterns collectively reinforce that regional deployment is driven by a complex interplay of land availability, economic conditions, infrastructure access, and policy context, with land-use variables taking a particularly central role in shaping spatial patterns across Europe.

In the GB model, which explores solar PV deployment at the subregional level, artificial surfaces and agricultural areas once again emerge as key spatial predictors. However, what distinguishes the subregional context is the prominence of the gross value added (GVA) by veterinary activities, which shows the strongest correlation with PV capacity. While this may at first seem surprising, veterinary services are closely tied to the wider agricultural economy, particularly in regions with significant livestock production. This suggests that it is not simply the physical presence of agricultural land that matters for deployment, but rather the strength of the broader agricultural economy and its associated services. Regions with a more developed agricultural sector—signalled here by higher GVA in veterinary activities—may be better positioned to invest in solar technologies or may attract more utility-scale PV projects due to their stronger financial base and access to capital.

This interpretation is supported by the findings of Mayer et al. [114], who developed a

deep learning framework to detect PV systems from aerial imagery in Germany and conducted a socio-economic analysis of PV adoption. The study reported a strong positive correlation (0.75) between agricultural gross value added and PV system frequency across the 53 counties of North-Rhine Westphalia, Germany's most populous state. This relationship suggests that areas with more economically productive agricultural sectors are more likely to host solar PV systems, aligning with the observed importance of agriculture-related economic indicators—including veterinary activities—in the GB subregional model. Similarly, Balta-Ozkan, Yildirim, and Connor [115] found that PV adoption at the subregional level was positively associated with income per capita, electricity sales, and the share of detached houses—all indicators of economic and infrastructural capacity. These findings reinforce the idea that land value and economic strength are more relevant drivers of deployment than land cover alone.

Climatic variables in the GB model, such as solar irradiation, show a weak to moderate correlation with deployment—stronger than at the national scale, but still less influential than land-use and economic factors. While solar irradiance is a driver for deployment, high irradiance is not sufficient on its own. As Balta-Ozkan, Yildirim, and Connor [115] also found, solar irradiation positively correlates with PV capacity at the subregional level, but its effect is conditional on the enabling economic and social context.

Altogether, the GB model highlights how subregional PV deployment is shaped by the intersection of land-use suitability, economic productivity, and moderate climatic potential. Artificial surfaces likely proxy for both grid accessibility and rooftop PV potential in residential or commercial areas, while economically productive agricultural regions offer space, demand, and capital for utility-scale development. These results underscore the importance of integrating land value and regional economic indicators into models aiming to forecast or explain local renewable energy deployment.

Across all three models, a consistent spatial pattern emerges: broader economic and social conditions are the dominant drivers of solar PV deployment at the national level, while land-use characteristics and the economic productivity of land become increasingly important at

regional and subregional scales. At the national level, deployment is primarily influenced by economic strength and education levels, which reflect a country’s institutional capacity, financial resources, and administrative ability to implement supportive policies.

In contrast, regional and subregional deployment is shaped more directly by the physical and economic attributes of space. Artificial surfaces and agricultural areas consistently emerge as strong predictors of capacity, acting as proxies for grid access, rooftop suitability, and available land for utility-scale installations. This shift in drivers illustrates how the constraints and enablers of deployment vary by scale: while national trends are governed by macroeconomic conditions and policy frameworks, local deployment depends on spatially distributed infrastructure, land availability, and institutional readiness at finer geographic levels.

Although climatic factors such as solar irradiation are often assumed to play a major role, the models show they are generally outweighed by structural and socioeconomic conditions—particularly in developed markets where solar technologies are already viable.

All three models demonstrated robust performance and effectively captured the dynamics of solar PV capacity across multiple spatial scales. The Global Model, applied at the national level to estimate annual capacity additions, achieved a global error of 9.7%. This error was calculated as the mean absolute error (MAE) divided by the mean cumulative capacity across countries. The European Model, which estimates cumulative capacity at the regional scale (NUTS 2), was evaluated using the mean absolute percentage error (MAPE). When applied across all European countries—including those without available regional PV data—the normalized model reported a national error of 19.5%. The absolute model initially exhibited a very high error in this same context. However, after applying scaling to align regional estimates with national totals, the error was effectively reduced to zero. For countries where regional capacity data are available—namely the United Kingdom, Italy, Spain, Belgium, Germany, and France—the absolute model achieved a MAPE of 6.4%, while the normalized model demonstrated superior accuracy with a MAPE of 2.5%. The high MAPE values observed when evaluating across all countries are primarily driven by countries with near-zero

PV capacity, which inflates percentage-based error metrics. Finally, the GB Model, which estimates cumulative capacity at the subregional (NUTS 3) level, achieved a national MAPE of 5.4%. These results collectively affirm the accuracy and scalability of the modelling framework for disaggregating, benchmarking, and forecasting solar PV deployment across spatial hierarchies.

The disaggregation of national PV capacity is most effectively addressed using the scaled EU models. In particular, the scaled absolute EU model accurately allocates unknown national capacities to specific geographic regions, ensuring alignment with national totals. The scaled normalized EU model is especially valuable for estimating regional capacities in countries without official subnational data, offering internally consistent and geographically meaningful estimates. At the subregional scale, the GB Model provides a robust framework for disaggregating capacity within Great Britain, effectively capturing fine-grained spatial variation in deployment.

Three Solar PV Deployment Indices were developed to benchmark deployment performance at the national, regional, and subregional levels. The Global Model provides a Solar PV Deployment Index that enables benchmarking of national performance against comparable countries. This index was used to assess policy impacts by comparing expected versus actual deployment in countries such as Italy, the United Kingdom, Spain, and Mexico. Countries were also ranked based on their deployment performance, providing insights into relative national progress. At the regional level, the absolute EU model facilitates cross-country comparisons, while the normalized model supports within-country benchmarking, identifying regions with higher or lower than expected deployment. The absolute model highlighted Brandenburg (Germany) and Castilla-La Mancha (Spain) as the top performing regions, both exceeding expectations relative to their geographic peers. The normalized model further revealed regional disparities in countries such as Belgium and Spain, identifying areas that significantly over- or underperformed relative to national trends. At the subregional level, the GB model was used to benchmark deployment across Great Britain. Here, Cambridgeshire stood out, while Barnsley, Doncaster, and Rotherham exhibited the

largest deficits relative to model predictions.

A key contribution of this thesis is the development of forecasting tools at national, regional, and subregional scales through the Global, EU, and GB models. These models enable spatially explicit forecasts of solar PV deployment, supporting more strategic and geographically informed planning. The Global Model can be used to project national capacity growth in alignment with a country's geographic and socioeconomic characteristics, offering a foundation for setting realistic national targets and informing international comparisons. These national forecasts can then be disaggregated using the EU and GB models, which provide detailed projections of regional and subregional deployment. The scaled models, in particular, are well-suited for this task as they maintain alignment with national totals while capturing local variation. At the subregional level, the GB model offers high-resolution forecasts that are particularly valuable for infrastructure planning and grid connection management - critical issues in Great Britain, where a significant backlog of renewable energy projects is awaiting connection. Importantly, these forecasting models also enable the identification of regions where land-use conflicts are likely to emerge, especially in areas with intense competition between agricultural and energy sectors. By highlighting these potential conflict zones, the models provide an evidence base for targeted interventions, such as the promotion of agrivoltaic systems, which balance energy generation with agricultural productivity. In this way, the models developed in this thesis offer a robust framework for guiding sustainable and equitable solar PV expansion over the coming decades.

6.1 Future Research

This study provides a foundation for understanding and forecasting solar PV deployment using spatially explicit models, but several avenues for future research remain. First, future work could build on the geographical modelling approach developed here by incorporating computer vision models—such as those using satellite or aerial imagery—to validate predicted PV capacities at different time points. Vision-based methods can also serve to improve the accuracy of geographical models by identifying previously undocumented installations,

tracking temporal changes in deployment, and refining land-use classifications. Such hybrid approaches, which fuse image recognition with socioeconomic and climatic predictors, could provide a more comprehensive picture of deployment trends across multiple spatial scales.

Second, the models developed in this thesis could be used to evaluate the effectiveness of specific policy interventions. By comparing observed deployment with model-predicted capacity, researchers can identify regions that over- or underperform relative to their geographic and socioeconomic conditions. This can support causal investigations into which policy instruments are most effective under different spatial and institutional contexts. Moreover, it can offer insight into how geographic factors—such as land availability, urban density, or infrastructure access—amplify or limit the impact of policies, ultimately contributing to more targeted and spatially appropriate policy design. The integration of political and governance-related indicators (e.g., local election results, public acceptance indices, or permitting frameworks) could further contextualize why regions with similar technical and economic profiles diverge in deployment outcomes.

Third, methodological extensions could strengthen the robustness and interpretability of forecasts. Uncertainty quantification beyond single-value errors (e.g., MAE or RMSE) is an important avenue: future work could explore prediction intervals, quantile regression, or probabilistic forecasting frameworks that generate full predictive distributions rather than point estimates. Such methods would allow decision-makers to assess not only the most likely deployment pathway but also the range of plausible outcomes. Similarly, given the hierarchical structure of the Global, EU, and GB models, forecast reconciliation methods that enforce coherence across spatial levels—or compositional approaches such as Dirichlet models that explicitly handle proportion data—could improve the internal consistency of predictions. Exploring alternative machine learning models, including probabilistic or generative frameworks, may also yield performance gains while providing additional interpretability and uncertainty metrics.

Finally, future studies could explore in greater depth the reasons why certain regions deviate significantly from expected capacity levels. Investigating whether such deviations

are driven by economic barriers (e.g., lack of investment), infrastructural constraints (e.g., grid limitations or permitting delays), or geographic factors (e.g., unsuitable terrain or land use conflicts) will help clarify the persistent gaps between technical potential and realized deployment. Combining model outputs with stakeholder interviews, policy timelines, or infrastructure maps could enrich this line of research. Moreover, future work could also explore the integration of additional data sources—such as high-resolution satellite surveys, grid curtailment statistics, or political and governance indicators—directly into modelling frameworks. Such fusion of diverse data streams would not only enhance predictive accuracy but also provide a more comprehensive understanding of the multidimensional barriers and enablers of solar PV expansion.

Bibliography

- [1] Philip Wolfe. “What Is Photovoltaics?” en. In: *The Solar Generation*. John Wiley & Sons, Ltd, 2018, pp. 9–24. ISBN: 978-1-119-42561-8. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119425618.ch2> (visited on 03/13/2025).
- [2] Becquerel M. E. “Memoire sur les Effets Electriques Produits sous l’Influence des Rayons Solaires”. In: *Comptes Rendus Hebdomadaires des Seances de L’Academie des Sciences* 9 (1839), pp. 561–567. URL: <https://cir.nii.ac.jp/crid/1570572699975814144> (visited on 03/11/2025).
- [3] “Effect of Light on Selenium During the Passage of An Electric Current *”. en. In: *Nature* 7.173 (Feb. 1873). Publisher: Nature Publishing Group, pp. 303–303. ISSN: 1476-4687. DOI: 10.1038/007303e0. URL: <https://www.nature.com/articles/007303e0> (visited on 03/13/2025).
- [4] William Grylls Adams and R. E. Day. “IX. The action of light on selenium”. In: *Philosophical Transactions of the Royal Society of London* 167 (1877). Publisher: Royal Society, pp. 313–349. DOI: 10.1098/rstl.1877.0009. URL: <https://royalsocietypublishing.org/doi/10.1098/rstl.1877.0009> (visited on 03/13/2025).
- [5] C. E. Fritts. “On a new form of selenium cell, and some electrical discoveries made by its use”. en. In: *American Journal of Science* s3-26.156 (Dec. 1883). Publisher: American Journal of Science, pp. 465–472. DOI: 10.2475/ajs.s3-26.156.465. URL: <https://ajsonline.org/article/65415-on-a-new-form-of-selenium-cell-and-some-electrical-discoveries-made-by-its-use> (visited on 03/13/2025).
- [6] Lewis M. Fraas. “History of Solar Cell Development”. en. In: *Low-Cost Solar Electric Power*. Ed. by Lewis M. Fraas. Cham: Springer International Publishing, 2014, pp. 1–12. ISBN: 978-3-319-07530-3. DOI: 10.1007/978-3-319-07530-3_1. URL: https://doi.org/10.1007/978-3-319-07530-3_1 (visited on 03/13/2025).
- [7] A. Einstein. “Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt”. en. In: *Annalen der Physik* 322.6 (1905), pp. 132–148. ISSN: 1521-3889. DOI: 10.1002/andp.19053220607. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19053220607> (visited on 03/13/2025).
- [8] *Nobel Prize in Physics 1921*. en-US. 2025. URL: <https://www.nobelprize.org/prizes/physics/1921/summary/> (visited on 03/13/2025).

- [9] D. M. Chapin, C. S. Fuller, and G. L. Pearson. “A New Silicon p-n Junction Photocell for Converting Solar Radiation into Electrical Power”. In: *Journal of Applied Physics* 25.5 (May 1954), pp. 676–677. ISSN: 0021-8979. DOI: 10.1063/1.1721711. URL: <https://doi.org/10.1063/1.1721711> (visited on 03/11/2025).
- [10] Dennis J. Flood. “SPACE PHOTOVOLTAICS – History, PROGRESS AND PROMISE”. In: *Modern Physics Letters B* 15.17n19 (Aug. 2001). Publisher: World Scientific Publishing Co., pp. 561–570. ISSN: 0217-9849. DOI: 10.1142/S0217984901002038. URL: <https://www.worldscientific.com/doi/abs/10.1142/S0217984901002038> (visited on 03/13/2025).
- [11] Samantha Hoang. *The Environmental History of Solar Photovoltaic Cells*. 2017. URL: <https://repository.wellesley.edu/object/ir1017> (visited on 03/13/2025).
- [12] R. B. Hall et al. “Thin-Film Polycrystalline (CdZn)S/Cu₂S Solar Cells of 10% Conversion Efficiency”. en. In: *Photovoltaic Solar Energy Conference*. Ed. by W. Palz. Dordrecht: Springer Netherlands, 1981, pp. 1094–1096. ISBN: 978-94-009-8423-3. DOI: 10.1007/978-94-009-8423-3_184.
- [13] A. Catalano et al. “Attainment of 10% conversion efficiency in amorphous silicon solar cells”. In: Sept. 1982. URL: <https://www.semanticscholar.org/paper/Attainment-of-10-conversion-efficiency-in-amorphous-Catalano-D'aiello/33940406766535d9c008ebb7e5af7b3c3b08bccb> (visited on 03/13/2025).
- [14] Luca Casalino. “Solar photovoltaic technology from an innovation perspective: a study using patents as a proxy of innovation”. en. PhD thesis. NHH Norwegian School of Economics, 2011.
- [15] Johann Wackerbauer and Jana Lippelt. “Photovoltaics: Boom of the Rising Sun”. en. In: *CESifo Forum 1/2012 (Spring)* 13 (2012). URL: <https://www.cesifo.org/en/publications/2012/article-journal/photovoltaics-boom-rising-sun> (visited on 03/26/2025).
- [16] Isaac Wilhelm and Sven Teske. *Solar photovoltaic electricity empowering the world*. Tech. rep. EPIA, 2011. URL: <https://www.greenpeace.org/static/planet4-netherlands-stateless/2018/06/Final-SolarGeneration-VI-full-report-lr.pdf> (visited on 03/26/2025).
- [17] International Renewable Energy Agency. *Renewable energy statistics 2024*. en. Tech. rep. Abu Dhabi, July 2024. URL: <https://www.irena.org/Publications/2024/Jul/Renewable-energy-statistics-2024> (visited on 03/11/2025).
- [18] V. V. Tyagi et al. “Progress in solar PV technology: Research and achievement”. In: *Renewable and Sustainable Energy Reviews* 20 (Apr. 2013), pp. 443–461. ISSN: 1364-0321. DOI: 10.1016/j.rser.2012.09.028. URL: <https://www.sciencedirect.com/science/article/pii/S1364032112005291> (visited on 03/26/2025).
- [19] IRENA (International Renewable Energy Agency) et al. *Solar (photovoltaic) panel prices vs. cumulative capacity*. en. 2024. URL: <https://ourworldindata.org/grapher/solar-pv-prices-vs-cumulative-capacity> (visited on 03/13/2025).
- [20] International Renewable Energy Agency. *Renewable Power Generation Costs in 2023*. en. Tech. rep. Abu Dhabi, Sept. 2024. URL: <https://www.irena.org/Publications/2024/Sep/Renewable-Power-Generation-Costs-in-2023> (visited on 03/11/2025).

- [21] Larry Partain et al. “‘Swanson’s Law’ plan to mitigate global climate change”. In: *2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC)*. June 2016, pp. 3335–3340. DOI: 10.1109/PVSC.2016.7750284. URL: <https://ieeexplore.ieee.org/document/7750284> (visited on 03/27/2025).
- [22] Bo Bai, Zheng Wang, and Jing Chen. “Shaping the solar future: An analysis of policy evolution, prospects and implications in China’s photovoltaic industry”. In: *Energy Strategy Reviews* 54 (July 2024), p. 101474. ISSN: 2211-467X. DOI: 10.1016/j.esr.2024.101474. URL: <https://www.sciencedirect.com/science/article/pii/S2211467X24001810> (visited on 03/11/2025).
- [23] Zihan Chen, Bo Song, and Jiachen Yan. “Comparative Analysis of the Photovoltaic Industry in China and the US: Policy, Innovation, and Market Competitiveness”. en. In: *Journal of Education, Humanities and Social Sciences* 42 (Dec. 2024), pp. 839–846. ISSN: 2771-2907. DOI: 10.54097/bqgdq374. URL: <https://drpress.org/ojs/index.php/EHSS/article/view/28309> (visited on 03/12/2025).
- [24] Sanzana Tabassum et al. “Solar Energy in the United States: Development, Challenges and Future Prospects”. en. In: *Energies* 14.23 (Jan. 2021), p. 8142. ISSN: 1996-1073. DOI: 10.3390/en14238142. URL: <https://www.mdpi.com/1996-1073/14/23/8142> (visited on 03/12/2025).
- [25] Eric O’Shaughnessy et al. “Policy-driven solar innovation and deployment remains critical for US grid decarbonization”. In: *Joule* 6.9 (Sept. 2022), pp. 1965–1968. ISSN: 2542-4351. DOI: 10.1016/j.joule.2022.07.012. URL: <https://www.sciencedirect.com/science/article/pii/S2542435122003592> (visited on 03/12/2025).
- [26] Paula Mints. “The history and future of incentives and the photovoltaic industry and how demand is driven”. en. In: *Progress in Photovoltaics: Research and Applications* 20.6 (2012), pp. 711–716. ISSN: 1099-159X. DOI: 10.1002/pip.1214. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/pip.1214> (visited on 03/18/2025).
- [27] T Erge, V. U Hoffmann, and K Kiefer. “The German experience with grid-connected PV-systems”. In: *Solar Energy. Grid-Connected PV* 70.6 (Jan. 2001), pp. 479–487. ISSN: 0038-092X. DOI: 10.1016/S0038-092X(00)00143-2. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X00001432> (visited on 03/18/2025).
- [28] Joern Hoppmann, Joern Huenteler, and Bastien Girod. “Compulsive policy-making—The evolution of the German feed-in tariff system for solar photovoltaic power”. In: *Research Policy* 43.8 (Oct. 2014), pp. 1422–1441. ISSN: 0048-7333. DOI: 10.1016/j.respol.2014.01.014. URL: <https://www.sciencedirect.com/science/article/pii/S0048733314000249> (visited on 03/18/2025).
- [29] Karolina Jankowska. “The German Policy Support Mechanism for Photovoltaics: The Road to Grid Parity”. en. In: *The Political Economy of Renewable Energy and Energy Security: Common Challenges and National Responses in Japan, China and Northern Europe*. Ed. by Espen Moe and Paul Midford. London: Palgrave Macmillan UK, 2014, pp. 258–275. ISBN: 978-1-137-33887-7. DOI: 10.1057/9781137338877_13. URL: https://doi.org/10.1057/9781137338877_13 (visited on 03/17/2025).

- [30] Toby D. Couture et al. *A Policymaker's Guide to Feed-in Tariff Policy Design*. en. Tech. rep. NREL, July 2010. URL: <https://ppp.worldbank.org/public-private-partnership/library/policymaker%E2%80%99s-guide-feed-tariff-policy-design>.
- [31] Véronique Vasseur and René Kemp. “The role of policy in the evolution of technological innovation systems for photovoltaic power in Germany and the Netherlands”. In: *International Journal of Technology, Policy and Management* 11.3-4 (Jan. 2011). Publisher: Inderscience Publishers, pp. 307–327. ISSN: 1468-4322. DOI: 10.1504/IJTPM.2011.042089. URL: <https://www.inderscienceonline.com/doi/abs/10.1504/IJTPM.2011.042089> (visited on 03/18/2025).
- [32] Statistics Netherlands. *Subsidy scheme for solar electricity systems*. en-GB. webpagina. Last Modified: 2008-05-15T15:00:02+02:00. URL: <https://www.cbs.nl/en-gb/news/2008/20/solar-electricity-not-very-popular-in-the-netherlands/subsidy-scheme-for-solar-electricity-systems> (visited on 03/19/2025).
- [33] IEA/IRENA Renewable Energy Policies and Measures Database. *MEP: Environmental Quality of Electricity Production (Milieukwaliteit van de Elektriciteitsproductie)*. en-GB. Mar. 2013. URL: <https://www.iea.org/policies/3831-mep-environmental-quality-of-electricity-production-milieukwaliteit-van-de-elektriciteitsproductie> (visited on 03/19/2025).
- [34] IEA Photovoltaic Power Systems Programme. *IEA PVPS Annual Report 2008*. Tech. rep. International Energy Agency (IEA), 2009. URL: https://iea-pvps.org/wp-content/uploads/2020/01/ar_2008.pdf (visited on 03/20/2025).
- [35] IEA Photovoltaic Power Systems Programme. *IEA PVPS Annual Report 2009*. Tech. rep. International Energy Agency (IEA), 2010. URL: <https://iea-pvps.org/wp-content/uploads/2020/01/AR-2009-Replacement-File-FINAL.pdf> (visited on 03/20/2025).
- [36] IEA Photovoltaic Power Systems Programme. *IEA PVPS Annual Report 2010*. Tech. rep. International Energy Agency (IEA), 2011. URL: https://iea-pvps.org/wp-content/uploads/2020/01/ar_2010.pdf (visited on 03/20/2025).
- [37] Zsuzsanna Pató. *Gridlock in the Netherlands*. en-US. Tech. rep. Regulatory Assistance Project (RAP), Feb. 2024. URL: <https://www.raponline.org/knowledge-center/gridlock-in-netherlands/> (visited on 01/21/2025).
- [38] Marc Londo et al. “Alternatives for current net metering policy for solar PV in the Netherlands: A comparison of impacts on business case and purchasing behaviour of private homeowners, and on governmental costs”. In: *Renewable Energy* 147 (Mar. 2020), pp. 903–915. ISSN: 0960-1481. DOI: 10.1016/j.renene.2019.09.062. URL: <https://www.sciencedirect.com/science/article/pii/S0960148119313928> (visited on 03/18/2025).
- [39] International Renewable Energy Agency (IRENA). *Renewable capacity statistics 2024*. en. Mar. 2024. URL: <https://www.irena.org/Publications/2024/Mar/Renewable-capacity-statistics-2024> (visited on 03/23/2025).

- [40] International Renewable Energy Agency (IRENA). *IRENA REsource Data Methodology*. Tech. rep. May 2015. URL: <https://dashboard.irena.org/download/methodology.pdf> (visited on 03/23/2025).
- [41] IEA. *Renewables 2024 Dataset*. en-GB. Oct. 2024. URL: <https://www.iea.org/data-and-statistics/data-product/renewables-2024-dataset> (visited on 03/23/2025).
- [42] IEA. *IEA Market Report Series Renewables 2024 Documentation*. Tech. rep. International Energy Agency, 2024. URL: <https://iea.blob.core.windows.net/assets/ac03ab85-8058-4622-8afb-bc2137485dc2/Documentation-IEAMarketReportSeriesRenewables2024.pdf> (visited on 03/23/2025).
- [43] Hongzhi Mao et al. “Advances and prospects on estimating solar photovoltaic installation capacity and potential based on satellite and aerial images”. In: *Renewable and Sustainable Energy Reviews* 179 (June 2023), p. 113276. ISSN: 1364-0321. DOI: 10.1016/j.rser.2023.113276. URL: <https://www.sciencedirect.com/science/article/pii/S1364032123001326> (visited on 02/21/2025).
- [44] Sheffield Solar. *Sheffield Solar API Platform: Capacity Data*. URL: <https://api.solar.sheffield.ac.uk/capacity/> (visited on 07/24/2024).
- [45] *Microgeneration Certification Scheme (MCS)*. en-GB. URL: <https://mcscertified.com/> (visited on 03/23/2025).
- [46] Office of Gas and Electricity Markets(Ofgem). *Feed-in Tariff Installation Report 30 September 2024*. en. Oct. 2024. URL: <https://www.ofgem.gov.uk/publications/feed-tariff-installation-report-30-september-2024> (visited on 03/23/2025).
- [47] Department for Energy Security and Net Zero. *Renewable Energy Planning Database: quarterly extract*. en. Mar. 2025. URL: <https://www.gov.uk/government/publications/renewable-energy-planning-database-monthly-extract> (visited on 03/23/2025).
- [48] Solar Media. *Utility-Scale Solar: UK Completed Assets Database*. en-US. URL: <https://marketresearch.solarmedia.co.uk/reports/utility-scale-solar-uk-completed-assets-database/> (visited on 03/23/2025).
- [49] O. T. Huxley et al. “The uncertainties involved in measuring national solar photovoltaic electricity generation”. In: *Renewable and Sustainable Energy Reviews* 156 (Mar. 2022), p. 112000. ISSN: 1364-0321. DOI: 10.1016/j.rser.2021.112000. URL: <https://www.sciencedirect.com/science/article/pii/S1364032121012636> (visited on 07/30/2024).
- [50] Terna. *Statistical Data of Energy Generation*. en. URL: <https://dati.terna.it/en/generation/statistical-data> (visited on 03/23/2025).
- [51] Red Eléctrica. *Installed capacity*. URL: <https://www.ree.es/en/datos/generation/installed-capacity> (visited on 04/16/2024).
- [52] Elia Transmission Belgium SA. *Photovoltaic power production estimation and forecast on Belgian grid (Historical)*. en-GB. URL: <https://opendata.elia.be/explore/dataset/ods032/information/?flg=en-gb> (visited on 07/24/2024).

- [53] marktstammdatenregister. *Current unit overview / MaStR*. URL: <https://www.marktstammdatenregister.de/MaStR/Einheit/Einheiten/OeffentlicheEinheitenuebersicht> (visited on 07/24/2024).
- [54] Florian Kotthoff et al. “Monitoring Germany’s Core Energy System Dataset: A Data Quality Analysis of the Marktstammdatenregister”. In: *SIGENERGY Energy Inform. Rev.* 4.4 (Feb. 2025), pp. 88–99. DOI: 10.1145/3717413.3717421. URL: <https://doi.org/10.1145/3717413.3717421> (visited on 05/29/2025).
- [55] Open Data Réseaux Énergies (ODRÉ). *Registre national des installations de production et de stockage d’électricité (au 31/12/2024)*. en-US. URL: <https://odre.opendatasoft.com/explore/dataset/registre-national-installation-production-stockage-electricite-agrege-311224/> (visited on 04/16/2024).
- [56] The National Institute of Statistics and Economic Studies (INSEE). *Definition - IRIS*. URL: <https://www.insee.fr/en/metadonnees/definition/c1523> (visited on 03/23/2025).
- [57] U.S. Energy Information Administration (EIA). *Annual Electric Power Industry Report, Form EIA-860 detailed data with previous form data (EIA-860A/860B)*. URL: <https://www.eia.gov/electricity/data/eia860/index.php> (visited on 03/27/2025).
- [58] U.S. Energy Information Administration (EIA). *Form EIA-861M (formerly EIA-826) detailed data*. URL: <https://www.eia.gov/electricity/data/eia861m/index.php> (visited on 03/27/2025).
- [59] U. S. Energy Information Administration (EIA). *EPM Technical Notes*. Tech. rep. URL: <https://www.eia.gov/electricity/monthly/pdf/technotes.pdf> (visited on 03/27/2025).
- [60] Dan Stowell et al. “A harmonised, high-coverage, open dataset of solar photovoltaic installations in the UK”. en. In: *Scientific Data* 7.1 (Nov. 2020). Publisher: Nature Publishing Group, p. 394. ISSN: 2052-4463. DOI: 10.1038/s41597-020-00739-0. URL: <https://www.nature.com/articles/s41597-020-00739-0> (visited on 12/24/2024).
- [61] Hamid Shaker, Hamidreza Zareipour, and David Wood. “A Data-Driven Approach for Estimating the Power Generation of Invisible Solar Sites”. In: *IEEE Transactions on Smart Grid* 7.5 (Sept. 2016). Conference Name: IEEE Transactions on Smart Grid, pp. 2466–2476. ISSN: 1949-3061. DOI: 10.1109/TSG.2015.2502140. URL: <https://ieeexplore.ieee.org/abstract/document/7347457> (visited on 09/30/2024).
- [62] Iain Staffell and Stefan Pfenninger. “The increasing impact of weather on electricity supply and demand”. In: *Energy* 145 (Feb. 2018), pp. 65–78. ISSN: 0360-5442. DOI: 10.1016/j.energy.2017.12.051. URL: <https://www.sciencedirect.com/science/article/pii/S0360544217320844> (visited on 07/30/2024).
- [63] Sebastian Schierenbeck et al. “Ein distanzbasiertes Hochrechnungsverfahren für die Einspeisung aus Photovoltaik”. In: *ET. Energiewirtschaftliche Tagesfragen* 60 (Jan. 2010), pp. 60–64.

- [64] Elke Lorenz et al. “Regional PV power prediction for improved grid integration”. en. In: *Progress in Photovoltaics: Research and Applications* 19.7 (2011), pp. 757–771. ISSN: 1099-159X. DOI: 10.1002/pip.1033. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/pip.1033> (visited on 09/30/2024).
- [65] Hamid Shaker, Hamidreza Zareipour, and David Wood. “Estimating Power Generation of Invisible Solar Sites Using Publicly Available Data”. In: *IEEE Transactions on Smart Grid* 7.5 (Sept. 2016). Conference Name: IEEE Transactions on Smart Grid, pp. 2456–2465. ISSN: 1949-3061. DOI: 10.1109/TSG.2016.2533164. URL: <https://ieeexplore.ieee.org/abstract/document/7426401> (visited on 09/30/2024).
- [66] Y. M. Saint-Drenan et al. “Analysis of the uncertainty in the estimates of regional PV power generation evaluated with the upscaling method”. In: *Solar Energy* 135 (Oct. 2016), pp. 536–550. ISSN: 0038-092X. DOI: 10.1016/j.solener.2016.05.052. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X16301694> (visited on 10/03/2024).
- [67] Elke Lorenz, Detlev Heinemann, and Christian Kurz. “Local and regional photovoltaic power prediction for large scale grid integration: Assessment of a new algorithm for snow detection”. en. In: *Progress in Photovoltaics: Research and Applications* 20.6 (2012), pp. 760–769. ISSN: 1099-159X. DOI: 10.1002/pip.1224. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/pip.1224> (visited on 09/30/2024).
- [68] Hamid Shaker, Daniel Manfre, and Hamidreza Zareipour. “Forecasting the aggregated output of a large fleet of small behind-the-meter solar photovoltaic sites”. In: *Renewable Energy* 147 (Mar. 2020), pp. 1861–1869. ISSN: 0960-1481. DOI: 10.1016/j.renene.2019.09.102. URL: <https://www.sciencedirect.com/science/article/pii/S0960148119314405> (visited on 10/03/2024).
- [69] Jamie M. Bright et al. “Improved satellite-derived PV power nowcasting using real-time power data from reference PV systems”. In: *Solar Energy*. Advances in Solar Resource Assessment and Forecasting 168 (July 2018), pp. 118–139. ISSN: 0038-092X. DOI: 10.1016/j.solener.2017.10.091. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X17309714> (visited on 10/03/2024).
- [70] Sven Killinger et al. “Upscaling PV Power Considering Module Orientations”. In: *IEEE Journal of Photovoltaics* 7.3 (May 2017). Conference Name: IEEE Journal of Photovoltaics, pp. 941–944. ISSN: 2156-3403. DOI: 10.1109/JPHOTOV.2017.2684908. URL: <https://ieeexplore.ieee.org/abstract/document/7890470/authors#authors> (visited on 10/04/2024).
- [71] Sven Killinger et al. “A tuning routine to correct systematic influences in reference PV systems’ power outputs”. In: *Solar Energy* 157 (Nov. 2017), pp. 1082–1094. ISSN: 0038-092X. DOI: 10.1016/j.solener.2017.09.001. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X17307697> (visited on 10/04/2024).
- [72] Y. M. Saint-Drenan, G. H. Good, and M. Braun. “A probabilistic approach to the estimation of regional photovoltaic power production”. In: *Solar Energy* 147 (May 2017), pp. 257–276. ISSN: 0038-092X. DOI: 10.1016/j.solener.2017.03.007. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X17301676> (visited on 10/03/2024).

- [73] Yves-Marie Saint-Drenan et al. “An approach for the estimation of the aggregated photovoltaic power generated in several European countries from meteorological data”. English. In: *Advances in Science and Research*. Vol. 15. ISSN: 1992-0628. Copernicus GmbH, May 2018, pp. 51–62. DOI: 10.5194/asr-15-51-2018. URL: <https://asr.copernicus.org/articles/15/51/2018/> (visited on 10/03/2024).
- [74] Energy Networks Association. *Ofgem Connections Delivery Board - October 2024 minutes*. en. Nov. 2024. URL: <https://www.energynetworks.org/publications/ofgem-connections-delivery-board-october-2024-minutes> (visited on 01/15/2025).
- [75] UK Government. *Clean Power 2030: Action Plan: A new era of clean electricity*. en. Tech. rep. Dec. 2024.
- [76] Environmental Audit Committee. *Technological innovations and climate change: onshore solar power*. May 2023. URL: <https://committees.parliament.uk/publications/39836/documents/193860/default/> (visited on 12/23/2024).
- [77] Exawatt, Open Climate Fix, and The University of Sheffield. *Written Evidence Submitted to the Environmental Audit Committee Inquiry on Technological Innovations and Climate Change: Onshore Solar Energy*. Dec. 2022. URL: <https://committees.parliament.uk/writtenevidence/114007/pdf/> (visited on 12/23/2024).
- [78] Electricity System Operator (ESO). *Connections Reform Summary of final recommendations*. Tech. rep. Dec. 2023. URL: <https://www.neso.energy/document/298491/download> (visited on 01/15/2025).
- [79] Rob Jetten. *New measures for grid congestion*. Oct. 2023. URL: <https://open.overheid.nl/documenten/fc7477f5-775b-4851-bdb3-30f89f1ad15a/file> (visited on 03/25/2025).
- [80] Božena Gajdzik et al. “Barriers to Renewable Energy Source (RES) Installations as Determinants of Energy Consumption in EU Countries”. en. In: *Energies* 16.21 (Jan. 2023), p. 7364. ISSN: 1996-1073. DOI: 10.3390/en16217364. URL: <https://www.mdpi.com/1996-1073/16/21/7364> (visited on 01/21/2025).
- [81] Christina Prifti et al. *Grid access challenges for wind farms in Europe*. Tech. rep. WindEurope, June 2024. URL: <https://windeurope.org/intelligence-platform/product/grid-access-challenges-for-wind-farms-in-europe/> (visited on 01/21/2025).
- [82] Shahriyar Nasirov, Carlos Silva, and Claudio A. Agostini. “Investors’ Perspectives on Barriers to the Deployment of Renewable Energy Sources in Chile”. en. In: *Energies* 8.5 (May 2015), pp. 3794–3814. ISSN: 1996-1073. DOI: 10.3390/en8053794. URL: <https://www.mdpi.com/1996-1073/8/5/3794> (visited on 01/21/2025).
- [83] REN21. *Renewables 2023 Global Status Report collection, Renewables in Energy Supply*. en. Tech. rep. 2023. URL: <https://www.ren21.net/gsr-2023/> (visited on 02/07/2025).
- [84] Will Gorman et al. “Grid connection barriers to renewable energy deployment in the United States”. In: *Joule* (Dec. 2024), p. 101791. ISSN: 2542-4351. DOI: 10.1016/j.joule.2024.11.008. URL: <https://www.sciencedirect.com/science/article/pii/S2542435124005038> (visited on 01/21/2025).

- [85] Pablo del Río and Pere Mir-Artigues. “A Cautionary Tale: Spain’s Solar PV Investment Bubble”. en. In: (Mar. 2014). URL: <https://policycommons.net/artifacts/614858/a-cautionary-tale/1595240/> (visited on 03/24/2025).
- [86] Javier López Prol. “Regulation, profitability and diffusion of photovoltaic grid-connected systems: A comparative analysis of Germany and Spain”. In: *Renewable and Sustainable Energy Reviews* 91 (Aug. 2018), pp. 1170–1181. ISSN: 1364-0321. DOI: 10.1016/j.rser.2018.04.030. URL: <https://www.sciencedirect.com/science/article/pii/S1364032118302594> (visited on 03/24/2025).
- [87] Haruki Yamaya et al. “PV market in Japan and impacts of grid constriction”. In: *2015 IEEE 42nd Photovoltaic Specialist Conference (PVSC)*. June 2015, pp. 1–6. DOI: 10.1109/PVSC.2015.7356204. URL: <https://ieeexplore.ieee.org/document/7356204> (visited on 03/25/2025).
- [88] Eugenia Giannini et al. “Penetration of Photovoltaics in Greece”. en. In: *Energies* 8.7 (June 2015), pp. 6497–6508. ISSN: 1996-1073. DOI: 10.3390/en8076497. URL: <https://www.mdpi.com/1996-1073/8/7/6497> (visited on 04/22/2025).
- [89] Georgios Christoforidis et al. “Assessing policies for photovoltaic net-metering in Greece”. In: Jan. 2014. DOI: 10.1049/cp.2014.1657.
- [90] Prateek Aggarwal et al. *Mapping India’s Energy Policy 2022*. en. Tech. rep. Canada: International Institute for Sustainable Development, May 2022. URL: <https://www.iisd.org/publications/mapping-india-energy-policy-2022> (visited on 03/24/2025).
- [91] Aditya Lolla et al. *India’s Race to 175 GW*. en-US. Tech. rep. EMBER, Apr. 2022. URL: <https://ember-energy.org/latest-insights/india-race-to-175-gw-renewables> (visited on 03/24/2025).
- [92] K. Calvert and W. Mabee. “More solar farms or more bioenergy crops? Mapping and assessing potential land-use conflicts among renewable energy technologies in eastern Ontario, Canada”. In: *Applied Geography* 56 (Jan. 2015), pp. 209–221. ISSN: 0143-6228. DOI: 10.1016/j.apgeog.2014.11.028. URL: <https://www.sciencedirect.com/science/article/pii/S0143622814002835> (visited on 09/17/2024).
- [93] Sanderine Nonhebel. “Renewable energy and food supply: will there be enough land?”. In: *Renewable and Sustainable Energy Reviews* 9.2 (Apr. 2005), pp. 191–201. ISSN: 1364-0321. DOI: 10.1016/j.rser.2004.02.003. URL: <https://www.sciencedirect.com/science/article/pii/S1364032104000498> (visited on 09/17/2024).
- [94] Ernesto Marcheggiani, Hubert Gulinck, and Andrea Galli. “Detection of Fast Landscape Changes: The Case of Solar Modules on Agricultural Land”. en. In: *Computational Science and Its Applications – ICCSA 2013*. Ed. by Beniamino Murgante et al. Berlin, Heidelberg: Springer, 2013, pp. 315–327. ISBN: 978-3-642-39649-6. DOI: 10.1007/978-3-642-39649-6_23.
- [95] Elnaz H. Adeh et al. “Solar PV Power Potential is Greatest Over Croplands”. en. In: *Scientific Reports* 9.1 (Aug. 2019). Publisher: Nature Publishing Group, p. 11442. ISSN: 2045-2322. DOI: 10.1038/s41598-019-47803-3. URL: <https://www.nature.com/articles/s41598-019-47803-3> (visited on 07/21/2024).

- [96] Katja Müller and Mareike Pampus. “The solar rush: invisible land grabbing in East Germany”. In: *International Journal of Sustainable Energy* 42.1 (Dec. 2023). Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/14786451.2023.2260009>, pp. 1264–1277. ISSN: 1478-6451. DOI: 10.1080/14786451.2023.2260009. URL: <https://doi.org/10.1080/14786451.2023.2260009> (visited on 09/22/2024).
- [97] Leonhard Späth. “Large-scale photovoltaics? Yes please, but not like this! Insights on different perspectives underlying the trade-off between land use and renewable electricity development”. In: *Energy Policy* 122 (Nov. 2018), pp. 429–437. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2018.07.029. URL: <https://www.sciencedirect.com/science/article/pii/S0301421518304762> (visited on 09/23/2024).
- [98] Jiafan Yu et al. “DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States”. English. In: *Joule* 2.12 (Dec. 2018). Publisher: Elsevier, pp. 2605–2617. ISSN: 2542-4785, 2542-4351. DOI: 10.1016/j.joule.2018.11.021. URL: [https://www.cell.com/joule/abstract/S2542-4351\(18\)30570-1](https://www.cell.com/joule/abstract/S2542-4351(18)30570-1) (visited on 01/22/2025).
- [99] Wei Hu et al. “What you get is not always what you see—pitfalls in solar array assessment using overhead imagery”. In: *Applied Energy* 327 (Dec. 2022), p. 120143. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2022.120143. URL: <https://www.sciencedirect.com/science/article/pii/S0306261922014003> (visited on 03/25/2025).
- [100] Copernicus Data Space Ecosystem. *Sentinel data*. en. URL: <https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data> (visited on 03/27/2025).
- [101] Inayatullah Jan, Waheed Ullah, and Muhammad Ashfaq. “Social acceptability of solar photovoltaic system in Pakistan: Key determinants and policy implications”. In: *Journal of Cleaner Production* 274 (Nov. 2020), p. 123140. ISSN: 0959-6526. DOI: 10.1016/j.jclepro.2020.123140. URL: <https://www.sciencedirect.com/science/article/pii/S0959652620331851> (visited on 03/27/2025).
- [102] Joshua Letchford, Kiran Lakkaraju, and Yevgeniy Vorobeychik. “Individual Household Modeling of Photovoltaic Adoption”. en. In: 2014.
- [103] Michaël Aklin, Chao-yo Cheng, and Johannes Urpelainen. “Geography, community, household: Adoption of distributed solar power across India”. In: *Energy for Sustainable Development* 42 (Feb. 2018), pp. 54–63. ISSN: 0973-0826. DOI: 10.1016/j.esd.2017.09.010. URL: <https://www.sciencedirect.com/science/article/pii/S0973082617301266> (visited on 01/22/2025).
- [104] Scott A. Robinson and Varun Rai. “Determinants of spatio-temporal patterns of energy technology adoption: An agent-based modeling approach”. In: *Applied Energy* 151 (Aug. 2015), pp. 273–284. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2015.04.071. URL: <https://www.sciencedirect.com/science/article/pii/S0306261915005346> (visited on 01/22/2025).

- [105] Paul Marty Jordan Fuentes, Kaveh Khalilpour, and Alexey Voinov. “Solar energy surge: The socio-economic determinants of the photovoltaic systems growth in Australia”. In: *Energy Research & Social Science* 116 (Oct. 2024), p. 103695. ISSN: 2214-6296. DOI: 10.1016/j.erss.2024.103695. URL: <https://www.sciencedirect.com/science/article/pii/S221462962400286X> (visited on 01/22/2025).
- [106] Ali Alderete Peralta, Nazmiye Balta-Ozkan, and Philip Longhurst. “Spatio-temporal modelling of solar photovoltaic adoption: An integrated neural networks and agent-based modelling approach”. In: *Applied Energy* 305 (Jan. 2022), p. 117949. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2021.117949. URL: <https://www.sciencedirect.com/science/article/pii/S0306261921012599> (visited on 01/22/2025).
- [107] Samuel H. C. Collier et al. “Distributed local energy: Assessing the determinants of domestic-scale solar photovoltaic uptake at the local level across England and Wales”. In: *Renewable and Sustainable Energy Reviews* 171 (Jan. 2023), p. 113036. ISSN: 1364-0321. DOI: 10.1016/j.rser.2022.113036. URL: <https://www.sciencedirect.com/science/article/pii/S1364032122009170> (visited on 01/23/2025).
- [108] Marcello Graziano and Kenneth Gillingham. “Spatial patterns of solar photovoltaic system adoption: The influence of neighbors and the built environment ‡”. In: *Journal of Economic Geography* 15.4 (July 2015), pp. 815–839. ISSN: 1468-2702. DOI: 10.1093/jeg/lbu036. URL: <https://doi.org/10.1093/jeg/lbu036> (visited on 01/22/2025).
- [109] Laura Williams and Mita Kerai. *Identifying trends in the deployment of domestic solar PV under the Feed-in Tariff scheme*. Tech. rep. Department of Energy & Climate Change (DECC), June 2012. URL: <https://www.gov.uk/government/publications/identifying-trends-in-the-deployment-of-domestic-solar-pv-under-the-feed-in-tariff-scheme> (visited on 01/22/2025).
- [110] Paul Westacott and Chiara Candelise. “A Novel Geographical Information Systems Framework to Characterize Photovoltaic Deployment in the UK: Initial Evidence”. en. In: *Energies* 9.1 (Jan. 2016), p. 26. ISSN: 1996-1073. DOI: 10.3390/en9010026. URL: <https://www.mdpi.com/1996-1073/9/1/26> (visited on 12/31/2024).
- [111] Christoph Thormeyer, Jan-Philipp Sasse, and Evelina Trutnevyte. “Spatially-explicit models should consider real-world diffusion of renewable electricity: Solar PV example in Switzerland”. In: *Renewable Energy* 145 (Jan. 2020), pp. 363–374. ISSN: 0960-1481. DOI: 10.1016/j.renene.2019.06.017. URL: <https://www.sciencedirect.com/science/article/pii/S0960148119308390> (visited on 01/31/2025).
- [112] Paula Donaduzzi Rigo et al. “How explain on-grid PV systems diffusion? Review and application in Brazil”. In: *Renewable Energy* 230 (Sept. 2024), p. 120862. ISSN: 0960-1481. DOI: 10.1016/j.renene.2024.120862. URL: <https://www.sciencedirect.com/science/article/pii/S0960148124009303> (visited on 02/05/2025).
- [113] Menzie McEachern and Susan Hanson. “Socio-geographic perception in the diffusion of innovation: Solar energy technology in Sri Lanka”. In: *Energy Policy* 36.7 (July 2008), pp. 2578–2590. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2008.03.020. URL: <https://www.sciencedirect.com/science/article/pii/S0301421508001389> (visited on 01/22/2025).

- [114] Kevin Mayer et al. “DeepSolar for Germany: A deep learning framework for PV system mapping from aerial imagery”. In: *2020 International Conference on Smart Energy Systems and Technologies (SEST)*. Sept. 2020, pp. 1–6. DOI: 10.1109/SEST48500.2020.9203258. URL: <https://ieeexplore.ieee.org/abstract/document/9203258> (visited on 01/22/2025).
- [115] Nazmiye Balta-Ozkan, Julide Yildirim, and Peter M. Connor. “Regional distribution of photovoltaic deployment in the UK and its determinants: A spatial econometric approach”. In: *Energy Economics* 51 (Sept. 2015), pp. 417–429. ISSN: 0140-9883. DOI: 10.1016/j.eneco.2015.08.003. URL: <https://www.sciencedirect.com/science/article/pii/S0140988315002248> (visited on 01/22/2025).
- [116] Bingchun Liu et al. “Forecasting of China’s solar PV industry installed capacity and analyzing of employment effect: based on GRA-BiLSTM model”. en. In: *Environmental Science and Pollution Research* 29.3 (Jan. 2022), pp. 4557–4573. ISSN: 1614-7499. DOI: 10.1007/s11356-021-15957-1. URL: <https://doi.org/10.1007/s11356-021-15957-1> (visited on 01/22/2025).
- [117] Ali Naci Celik and Evren Özgür. “Review of Turkey’s photovoltaic energy status: Legal structure, existing installed power and comparative analysis”. In: *Renewable and Sustainable Energy Reviews* 134 (Dec. 2020), p. 110344. ISSN: 1364-0321. DOI: 10.1016/j.rser.2020.110344. URL: <https://www.sciencedirect.com/science/article/pii/S1364032120306328> (visited on 01/22/2025).
- [118] Anita M. Bunea et al. “Diffusion of Solar PV Energy in the UK: A Comparison of Sectoral Patterns”. en. In: *Forecasting* 4.2 (June 2022), pp. 456–476. ISSN: 2571-9394. DOI: 10.3390/forecast4020026. URL: <https://www.mdpi.com/2571-9394/4/2/26> (visited on 12/20/2024).
- [119] L. Kruitwagen et al. “A global inventory of photovoltaic solar energy generating units”. en. In: *Nature* 598.7882 (Oct. 2021). Publisher: Nature Publishing Group, pp. 604–610. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03957-7. URL: <https://www.nature.com/articles/s41586-021-03957-7> (visited on 07/17/2024).
- [120] Xin Li et al. “How does residential solar PV system diffusion occur in Australia?-A logistic growth curve modelling approach”. In: *Sustainable Energy Technologies and Assessments* 56 (Mar. 2023), p. 103060. ISSN: 2213-1388. DOI: 10.1016/j.seta.2023.103060. URL: <https://www.sciencedirect.com/science/article/pii/S2213138823000528> (visited on 01/31/2025).
- [121] Carolyn Davidson et al. “Modeling photovoltaic diffusion: an analysis of geospatial datasets”. en. In: *Environmental Research Letters* 9.7 (July 2014). Publisher: IOP Publishing, p. 074009. ISSN: 1748-9326. DOI: 10.1088/1748-9326/9/7/074009. URL: <https://dx.doi.org/10.1088/1748-9326/9/7/074009> (visited on 02/04/2025).
- [122] Jonas Müller and Evelina Trutnevyte. “Spatial projections of solar PV installations at subnational level: Accuracy testing of regression models”. In: *Applied Energy* 265 (May 2020), p. 114747. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2020.114747. URL: <https://www.sciencedirect.com/science/article/pii/S0306261920302592> (visited on 02/04/2025).

- [123] Chaonan Ji et al. “Solar photovoltaic module detection using laboratory and airborne imaging spectroscopy data”. In: *Remote Sensing of Environment* 266 (Dec. 2021), p. 112692. ISSN: 0034-4257. DOI: 10.1016/j.rse.2021.112692. URL: <https://www.sciencedirect.com/science/article/pii/S0034425721004120> (visited on 04/08/2025).
- [124] Moussa Sofiane Karoui et al. “Partial Linear NMF-Based Unmixing Methods for Detection and Area Estimation of Photovoltaic Panels in Urban Hyperspectral Remote Sensing Data”. en. In: *Remote Sensing* 11.18 (Jan. 2019), p. 2164. ISSN: 2072-4292. DOI: 10.3390/rs11182164. URL: <https://www.mdpi.com/2072-4292/11/18/2164> (visited on 04/08/2025).
- [125] Jordan M. Malof et al. “Automatic detection of solar photovoltaic arrays in high resolution aerial imagery”. In: *Applied Energy* 183 (Dec. 2016), pp. 229–240. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2016.08.191. URL: <https://www.sciencedirect.com/science/article/pii/S0306261916313009> (visited on 04/08/2025).
- [126] Jordan M. Malof et al. “Automatic solar photovoltaic panel detection in satellite imagery”. In: *2015 International Conference on Renewable Energy Research and Applications (ICRERA)*. Nov. 2015, pp. 1428–1431. DOI: 10.1109/ICRERA.2015.7418643. URL: <https://ieeexplore.ieee.org/document/7418643> (visited on 04/01/2025).
- [127] Xunhe Zhang et al. “Texture Is Important in Improving the Accuracy of Mapping Photovoltaic Power Plants: A Case Study of Ningxia Autonomous Region, China”. en. In: *Remote Sensing* 13.19 (Jan. 2021), p. 3909. ISSN: 2072-4292. DOI: 10.3390/rs13193909. URL: <https://www.mdpi.com/2072-4292/13/19/3909> (visited on 04/08/2025).
- [128] Veerle Plakman, Job Rosier, and Jasper van Vliet. “Solar park detection from publicly available satellite imagery”. EN. In: *GIScience & Remote Sensing* (Dec. 2022). Publisher: Taylor & Francis. ISSN: 1548-1603. URL: <https://www.tandfonline.com/doi/abs/10.1080/15481603.2022.2036056> (visited on 04/08/2025).
- [129] Marek Vasku. “An exploration of automatic detection of large-scale solar plants: application of machine learning-based image classification in Google Earth Engine”. en. PhD thesis. Copenhagen, Denmark: Aalborg University, 2019.
- [130] Min Wang et al. “Photovoltaic panel extraction from very high-resolution aerial imagery using region–line primitive association analysis and template matching”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 141 (July 2018), pp. 100–111. ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2018.04.010. URL: <https://www.sciencedirect.com/science/article/pii/S0924271618301205> (visited on 04/08/2025).
- [131] Amir Mohammad Moradi Sizkouhi et al. “Automatic Boundary Extraction of Large-Scale Photovoltaic Plants Using a Fully Convolutional Network on Aerial Imagery”. In: *IEEE Journal of Photovoltaics* 10.4 (July 2020), pp. 1061–1067. ISSN: 2156-3403. DOI: 10.1109/JPHOTOV.2020.2992339. URL: <https://ieeexplore.ieee.org/document/9095250> (visited on 04/10/2025).

- [132] Omar Elharrouss et al. “Backbones-review: Feature extractor networks for deep learning and deep reinforcement learning approaches in computer vision”. In: *Computer Science Review* 53 (Aug. 2024), p. 100645. ISSN: 1574-0137. DOI: 10.1016/j.cosrev.2024.100645. URL: <https://www.sciencedirect.com/science/article/pii/S1574013724000297> (visited on 04/10/2025).
- [133] Maximilian Schulz, Bilel Boughattas, and Frank Wendel. “DetEEktor: Mask R-CNN based neural network for energy plant identification on aerial photographs”. In: *Energy and AI* 5 (Sept. 2021), p. 100069. ISSN: 2666-5468. DOI: 10.1016/j.egyai.2021.100069. URL: <https://www.sciencedirect.com/science/article/pii/S2666546821000239> (visited on 04/10/2025).
- [134] SiMing Liang et al. “Mask R-CNN based segmentation method for satellite imagery of photovoltaics generation systems”. In: *2020 39th Chinese Control Conference (CCC)*. ISSN: 1934-1768. July 2020, pp. 5343–5348. DOI: 10.23919/CCC50068.2020.9189474. URL: <https://ieeexplore.ieee.org/document/9189474> (visited on 04/10/2025).
- [135] Tomohiro Ishii et al. “Detection by classification of buildings in multispectral satellite imagery”. In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. Dec. 2016, pp. 3344–3349. DOI: 10.1109/ICPR.2016.7900150. URL: <https://ieeexplore.ieee.org/document/7900150> (visited on 04/10/2025).
- [136] Ayobami S. Edun et al. “Unsupervised azimuth estimation of solar arrays in low-resolution satellite imagery through semantic segmentation and Hough transform”. In: *Applied Energy* 298 (Sept. 2021), p. 117273. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2021.117273. URL: <https://www.sciencedirect.com/science/article/pii/S0306261921006905> (visited on 04/10/2025).
- [137] Joseph Camilo et al. *Application of a semantic segmentation convolutional neural network for accurate automatic detection and mapping of solar photovoltaic arrays in aerial imagery*. Jan. 2018. DOI: 10.48550/arXiv.1801.04018. URL: <http://arxiv.org/abs/1801.04018> (visited on 04/10/2025).
- [138] Andrés Pérez-González, Álvaro Jaramillo-Duque, and Juan Bernardo Cano-Quintero. “Automatic Boundary Extraction for Photovoltaic Plants Using the Deep Learning U-Net Model”. en. In: *Applied Sciences* 11.14 (Jan. 2021), p. 6524. ISSN: 2076-3417. DOI: 10.3390/app11146524. URL: <https://www.mdpi.com/2076-3417/11/14/6524> (visited on 04/10/2025).
- [139] Amirmohammad Moradi Sizkouhi, Mohammadreza Aghaei, and Sayyed Majid Esmailfar. *Aerial Imagery of PV Plants for boundary detection*. en. Apr. 2020. URL: <https://ieee-dataport.org/documents/aerial-imagery-pv-plants-boundary-detection> (visited on 04/10/2025).
- [140] Abraham Noah Wu and Filip Biljecki. “Roofpedia: Automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability”. In: *Landscape and Urban Planning* 214 (Oct. 2021), p. 104167. ISSN: 0169-2046. DOI: 10.1016/j.landurbplan.2021.104167. URL: <https://www.sciencedirect.com/science/article/pii/S0169204621001304> (visited on 04/10/2025).

- [141] Roberto Castello et al. “Deep learning in the built environment: automatic detection of rooftop solar panels using Convolutional Neural Networks”. en. In: *Journal of Physics: Conference Series* 1343.1 (Nov. 2019). Publisher: IOP Publishing, p. 012034. ISSN: 1742-6596. DOI: 10.1088/1742-6596/1343/1/012034. URL: <https://dx.doi.org/10.1088/1742-6596/1343/1/012034> (visited on 04/10/2025).
- [142] Matthias Zech and Joseph Ranalli. “Predicting PV Areas in Aerial Images with Deep Learning”. In: *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*. ISSN: 0160-8371. June 2020, pp. 0767–0774. DOI: 10.1109/PVSC45281.2020.9300636. URL: <https://ieeexplore.ieee.org/document/9300636> (visited on 04/11/2025).
- [143] Yongshi Jie et al. “Combined Multi-Layer Feature Fusion and Edge Detection Method for Distributed Photovoltaic Power Station Identification”. en. In: *Energies* 13.24 (Jan. 2020), p. 6742. ISSN: 1996-1073. DOI: 10.3390/en13246742. URL: <https://www.mdpi.com/1996-1073/13/24/6742> (visited on 04/11/2025).
- [144] IEA. *Deploying Renewables 2011: Best and Future Policy Practice*. en-GB. Tech. rep. Paris: IEA, Nov. 2011. URL: <https://www.iea.org/reports/deploying-renewables-2011-best-and-future-policy-practice> (visited on 04/17/2025).
- [145] D. Jacobs and B. K. Sovacool. “Feed-In Tariffs and Other Support Mechanisms for Solar PV Promotion”. In: *Comprehensive Renewable Energy*. Ed. by Ali Sayigh. Oxford: Elsevier, Jan. 2012, pp. 73–109. ISBN: 978-0-08-087873-7. DOI: 10.1016/B978-0-08-087872-0.00104-9. URL: <https://www.sciencedirect.com/science/article/pii/B9780080878720001049> (visited on 04/11/2025).
- [146] António Cardoso Marques, José Alberto Fuinhas, and Daniela Pereira Macedo. “The impact of feed-in and capacity policies on electricity generation from renewable energy sources in Spain”. In: *Utilities Policy* 56 (Feb. 2019), pp. 159–168. ISSN: 0957-1787. DOI: 10.1016/j.jup.2019.01.001. URL: <https://www.sciencedirect.com/science/article/pii/S0957178718300079> (visited on 04/11/2025).
- [147] Boqiang Lin and Yongjing Xie. “How feed-in-tariff subsidies affect renewable energy investments in China? New evidence from firm-level data”. In: *Energy* 294 (May 2024), p. 130853. ISSN: 0360-5442. DOI: 10.1016/j.energy.2024.130853. URL: <https://www.sciencedirect.com/science/article/pii/S036054422400625X> (visited on 02/27/2025).
- [148] Mutaka Alolo, Alcino Azevedo, and Izidin El Kalak. “The effect of the feed-in-system policy on renewable energy investments: Evidence from the EU countries”. In: *Energy Economics* 92 (Oct. 2020), p. 104998. ISSN: 0140-9883. DOI: 10.1016/j.eneco.2020.104998. URL: <https://www.sciencedirect.com/science/article/pii/S0140988320303388> (visited on 04/11/2025).
- [149] Elbert Dijkgraaf, Tom P. van Dorp, and Emiel Maasland. “On the Effectiveness of Feed-in Tariffs in the Development of Solar Photovoltaics”. EN. In: *The Energy Journal* 39.1 (Jan. 2018). Publisher: SAGE Publications, pp. 81–100. ISSN: 0195-6574. DOI: 10.5547/01956574.39.1.edij. URL: <https://doi.org/10.5547/01956574.39.1.edij> (visited on 04/11/2025).

- [150] Daniele Poponi, Riccardo Basosi, and Lado Kurdgelashvili. “Subsidisation cost analysis of renewable energy deployment: A case study on the Italian feed-in tariff programme for photovoltaics”. In: *Energy Policy* 154 (July 2021), p. 112297. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2021.112297. URL: <https://www.sciencedirect.com/science/article/pii/S030142152100166X> (visited on 04/11/2025).
- [151] C. Batlle, I. J. Pérez-Arriaga, and P. Zambrano-Barragán. “Regulatory design for RES-E support mechanisms: Learning curves, market structure, and burden-sharing”. In: *Energy Policy*. Modeling Transport (Energy) Demand and Policies 41 (Feb. 2012), pp. 212–220. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2011.10.039. URL: <https://www.sciencedirect.com/science/article/pii/S0301421511008238> (visited on 06/05/2025).
- [152] Yuliya Karneyeva and Rolf Wüstenhagen. “Solar feed-in tariffs in a post-grid parity world: The role of risk, investor diversity and business models”. In: *Energy Policy* 106 (July 2017), pp. 445–456. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2017.04.005. URL: <https://www.sciencedirect.com/science/article/pii/S0301421517302306> (visited on 04/19/2025).
- [153] IRENA. *Renewable Energy Auctions: Status and Trends Beyond Price*. en. Tech. rep. Abu Dhabi, Dec. 2019. URL: <https://www.irena.org/publications/2019/Dec/Renewable-energy-auctions-Status-and-trends-beyond-price> (visited on 04/19/2025).
- [154] Ángel Ordóñez et al. “Net-metering and net-billing in photovoltaic self-consumption: The cases of Ecuador and Spain”. In: *Sustainable Energy Technologies and Assessments* 53 (Oct. 2022), p. 102434. ISSN: 2213-1388. DOI: 10.1016/j.seta.2022.102434. URL: <https://www.sciencedirect.com/science/article/pii/S2213138822004866> (visited on 04/22/2025).
- [155] E. Doris and V. A. Krasko. *Strategic Sequencing for State Distributed PV Policies: A Quantitative Analysis of Policy Impacts and Interactions*. English. Tech. rep. NREL/TP-7A30-56428. National Renewable Energy Lab. (NREL), Golden, CO (United States), Oct. 2012. DOI: 10.2172/1054826. URL: <https://www.osti.gov/biblio/1054826> (visited on 04/21/2025).
- [156] D. Steward et al. *Effectiveness of State-Level Policies on Solar Market Development in Different State Contexts*. English. Tech. rep. NREL/TP-7A40-61029. National Renewable Energy Lab (NREL), Feb. 2014. DOI: 10.2172/1126288. URL: <https://www.osti.gov/biblio/1126288> (visited on 04/21/2025).
- [157] Haitao Yin and Nicholas Powers. “Do state renewable portfolio standards promote in-state renewable generation?”. In: *Energy Policy* 38.2 (Feb. 2010), pp. 1140–1149. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2009.10.067. URL: <https://www.sciencedirect.com/science/article/pii/S0301421509008283> (visited on 04/21/2025).
- [158] Gilbert Michaud. “Net Energy Metering and Community Shared Solar Deployment in the U.S.: Policy Perspectives, Barriers, and Opportunities”. en. PhD thesis. Richmond, Virginia: Virginia Commonwealth University, 2016. URL: <https://www.proquest>.

- com/openview/1999bc759ed6e0495346aaba61f2e05f/1?cbl=18750&pq-origsite=gscholar (visited on 04/21/2025).
- [159] Agustin J. Ros and Sai Shetty Sai. “Residential rooftop solar demand in the U.S. and the impact of net energy metering and electricity prices”. In: *Energy Economics* 118 (Feb. 2023), p. 106491. ISSN: 0140-9883. DOI: 10.1016/j.eneco.2022.106491. URL: <https://www.sciencedirect.com/science/article/pii/S014098832200620X> (visited on 04/21/2025).
 - [160] Serena Y. Kim et al. “Spatial distribution of solar PV deployment: an application of the region-based convolutional neural network”. en. In: *EPJ Data Science* 12.1 (Dec. 2023), pp. 1–34. ISSN: 2193-1127. DOI: 10.1140/epjds/s13688-023-00399-1. URL: <https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-023-00399-1> (visited on 03/06/2025).
 - [161] Ofgem. *Renewables Obligation closure*. en. URL: <https://www.ofgem.gov.uk/environmental-and-social-schemes/renewables-obligation-ro/renewables-obligation-ro-ro-closure> (visited on 12/21/2022).
 - [162] UK Government. *Renewables Obligation Call for Evidence on introducing Fixed Price Certificates into the UK-wide Renewables Obligation schemes*. en. Tech. rep. July 2023.
 - [163] Nicole Watson and Paul Bolton. *Contracts for Difference Scheme*. en-GB. Tech. rep. Oct. 2024. URL: <https://commonslibrary.parliament.uk/research-briefings/cbp-9871/> (visited on 06/05/2025).
 - [164] Christine Lasco Crago and Ilya Chernyakhovskiy. “Are policy incentives for solar power effective? Evidence from residential installations in the Northeast”. In: *Journal of Environmental Economics and Management* 81 (Jan. 2017), pp. 132–151. ISSN: 0095-0696. DOI: 10.1016/j.jeem.2016.09.008. URL: <https://www.sciencedirect.com/science/article/pii/S0095069616302996> (visited on 04/21/2025).
 - [165] Gireesh Shrimali and Steffen Jenner. “The impact of state policy on deployment and cost of solar photovoltaic technology in the U.S.: A sector-specific empirical analysis”. In: *Renewable Energy* 60 (Dec. 2013), pp. 679–690. ISSN: 0960-1481. DOI: 10.1016/j.renene.2013.06.023. URL: <https://www.sciencedirect.com/science/article/pii/S0960148113003169> (visited on 04/21/2025).
 - [166] REN21. *Renewables 2022 Global Status Report*. en. Tech. rep. Paris: REN21, 2022. URL: <https://www.ren21.net/gsr-2022>.
 - [167] IRENA. *Renewable Power Generation Costs in 2020*. en. Tech. rep. June 2021. URL: <https://www.irena.org/publications/2021/Jun/Renewable-Power-Costs-in-2020> (visited on 04/17/2025).
 - [168] Government of China. *China’s Achievements, New Goals and New Measures for Nationally Determined Contributions*. Tech. rep. Oct. 2021. URL: <https://www.fao.org/faolex/results/details/en/c/LEX-FA0C207444/> (visited on 04/17/2025).
 - [169] Agency for Natural Resources and Energy (ANRE)}. *Outline of Strategic Energy Plan*. Tech. rep. 2021. URL: https://www.enecho.meti.go.jp/en/category/others/basic_plan/pdf/6th_outline.pdf (visited on 04/17/2025).

- [170] Ministry of Economic Development (MISE), Ministry of the Environment and Protection of Natural Resources and the Sea, and Ministry of Infrastructure and Transport. *Integrated National Energy and Climate Plan*. Dec. 2019. URL: https://energy.ec.europa.eu/system/files/2020-02/it_final_necp_main_en_0.pdf (visited on 04/17/2025).
- [171] Federal Ministry For Economic Affairs and Climate Action (BMWK). *Overview of the Easter Package*. Tech. rep. Berlin, Apr. 2022. URL: https://www.bmwk.de/Redaktion/EN/Downloads/Energy/0406_ueberblickspapier_osterpaket_en.pdf?__blob=publicationFile&v=5 (visited on 04/17/2025).
- [172] Government of Spain. *Draft of The Integrated National Energy and Climate Plan 2021-2030*. Tech. rep. Feb. 2019. URL: https://energy.ec.europa.eu/system/files/2019-06/ec_courtesy_translation_es_necp_0.pdf (visited on 04/17/2025).
- [173] REN21. *Renewables 2021 Global Status Report*. Tech. rep. Paris: REN21, 2021. URL: https://www.ren21.net/wp-content/uploads/2019/05/GSR2021_Full_Report.pdf (visited on 04/17/2025).
- [174] Department of Public Enterprises, South Africa. *Roadmap for Eskom in a Reformed Electricity Supply Industry*. Tech. rep. 2019. URL: https://www.gov.za/sites/default/files/gcis_document/201910/roadmap-eskom.pdf (visited on 04/17/2025).
- [175] REN21. *Renewables 2020 Global Status Report*. en. Tech. rep. Paris, 2020. URL: <https://www.ren21.net/gsr-2020> (visited on 04/17/2025).
- [176] US Census Bureau. *Glossary*. en. Section: Government. 2022. URL: <https://www.census.gov/programs-surveys/geography/about/glossary.html> (visited on 12/26/2022).
- [177] IEA. *World Energy Model Documentation*. Tech. rep. IEA, Oct. 2021. URL: https://iea.blob.core.windows.net/assets/932ea201-0972-4231-8d81-356300e9fc43/WEM_Documentation_WEO2021.pdf (visited on 04/18/2025).
- [178] Nancy M. Haegel et al. “Terawatt-scale photovoltaics: Trajectories and challenges”. In: *Science* 356.6334 (Apr. 2017). Publisher: American Association for the Advancement of Science, pp. 141–143. DOI: 10.1126/science.aal1288. URL: <https://www.science.org/doi/10.1126/science.aal1288> (visited on 04/18/2025).
- [179] IRENA. *Global energy transformation: A roadmap to 2050 (2019 edition)*. en. Tech. rep. IRENA, Apr. 2019. URL: <https://www.irena.org/publications/2019/Apr/Global-energy-transformation-A-roadmap-to-2050-2019Edition> (visited on 04/18/2025).
- [180] Rashmi Ravishankar et al. “Capacity Estimation of Solar Farms Using Deep Learning on High-Resolution Satellite Imagery”. en. In: *Remote Sensing* 15.1 (Jan. 2023), p. 210. ISSN: 2072-4292. DOI: 10.3390/rs15010210. URL: <https://www.mdpi.com/2072-4292/15/1/210> (visited on 04/18/2025).

- [181] Simiao Ren et al. “Utilizing Geospatial Data for Assessing Energy Security: Mapping Small Solar Home Systems Using Unmanned Aerial Vehicles and Deep Learning”. en. In: *ISPRS International Journal of Geo-Information* 11.4 (Apr. 2022), p. 222. ISSN: 2220-9964. DOI: 10.3390/ijgi11040222. URL: <https://www.mdpi.com/2220-9964/11/4/222> (visited on 04/18/2025).
- [182] Ali Imran et al. “Electric-Dipole Gated Two Terminal Phototransistor for Charge-Coupled Device”. en. In: *Advanced Optical Materials* 11.22 (2023), p. 2300910. ISSN: 2195-1071. DOI: 10.1002/adom.202300910. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/adom.202300910> (visited on 04/18/2025).
- [183] Marcel Suri et al. *Global Photovoltaic Power Potential by Country*. Tech. rep. Washington, D.C: World Bank, 2020. URL: <http://documents.worldbank.org/curated/en/466331592817725242>.
- [184] IEA and CMCC. *Weather, Climate and Energy Tracker*. en-GB. 2022. URL: <https://www.iea.org/data-and-statistics/data-tools/weather-for-energy-tracker> (visited on 08/07/2022).
- [185] World Bank. *World Bank Open Data*. en. URL: <https://data.worldbank.org> (visited on 04/18/2025).
- [186] IRENA. *Statistics Data*. en. URL: <https://www.irena.org/Data> (visited on 01/16/2022).
- [187] EIA. *International - U.S. Energy Information Administration (EIA) Electricity Data*. URL: <https://www.eia.gov/international/data/world> (visited on 04/06/2022).
- [188] JMP Statistical Discovery. *Statistical Details for the Cluster Variables Platform*. Mar. 2025. URL: <https://www.jmp.com/support/help/en/18.2/index.shtml#page/jmp/statistical-details-for-the-cluster-variables-platform.shtml#> (visited on 08/07/2025).
- [189] SAS Institute Inc. *The VARCLUS Procedure*. en. Tech. rep. Cary, NC, 2024. URL: https://go.documentation.sas.com/api/collections/pgmsascdc/v_049/docsets/statug/content/varclus.pdf.
- [190] UNESCO Institute for Statistics. *Data for the Sustainable Development Goals*. eng. 2021. URL: <https://uis.unesco.org/> (visited on 03/22/2022).
- [191] World Bank. *Labor force*. en. 2022. URL: <https://data.worldbank.org> (visited on 03/22/2022).
- [192] International Labour Organization (ILOSTAT). *ILOSTAT database*. en-US. 2022. URL: <https://ilostat ilo.org/data/> (visited on 04/11/2022).
- [193] World Bank. *Total Population*. en. 2022. URL: <https://data.worldbank.org> (visited on 03/21/2022).
- [194] Savina Gygli et al. “The KOF Globalisation Index – revisited”. en. In: *The Review of International Organizations* 14.3 (Sept. 2019), pp. 543–574. ISSN: 1559-744X. DOI: 10.1007/s11558-019-09344-2. URL: <https://doi.org/10.1007/s11558-019-09344-2> (visited on 04/20/2025).

- [195] Axel Dreher. “Does globalization affect growth? Evidence from a new index of globalization”. In: *Applied Economics* 38.10 (June 2006), pp. 1091–1110. ISSN: 0003-6846. DOI: 10.1080/00036840500392078. URL: <https://doi.org/10.1080/00036840500392078> (visited on 04/20/2025).
- [196] World Bank. *Agriculture, forestry, and fishing, value added (current US\$)*. en. 2022. URL: <https://data.worldbank.org> (visited on 03/30/2022).
- [197] OECD. *OECD National Accounts Data*. 2022. URL: <https://data-explorer.oecd.org/> (visited on 04/20/2025).
- [198] World Bank. *Manufacturing, value added (current US\$)*. en. 2022. URL: <https://data.worldbank.org> (visited on 03/30/2022).
- [199] World Bank. *Industry (including construction), value added (current US\$)*. en. 2022. URL: <https://data.worldbank.org> (visited on 04/20/2025).
- [200] World Bank. *GDP (current US\$)*. en. 2022. URL: <https://data.worldbank.org> (visited on 03/22/2022).
- [201] OECD. *Nominal gross domestic product (GDP)*. en. 2022. URL: <https://www.oecd.org/en/data/indicators/nominal-gross-domestic-product-gdp.html> (visited on 04/20/2025).
- [202] World Bank. *GNI (current US\$)*. en. 2022. URL: <https://data.worldbank.org> (visited on 03/22/2022).
- [203] OECD. *Gross national income*. en. 2022. URL: <https://www.oecd.org/en/data/indicators/gross-national-income.html> (visited on 03/22/2022).
- [204] World Bank. *Ease of doing business rank*. en. 2022. URL: <https://data.worldbank.org> (visited on 03/23/2022).
- [205] IRENA. *Public Investment Trends in Renewables Dataset*. en. 2021. URL: <https://www.irena.org/Data/View-data-by-topic/Finance-and-Investment/Renewable-Energy-Finance-Flows> (visited on 02/11/2022).
- [206] François Lafond et al. “How well do experience curves predict technological progress? A method for making distributional forecasts”. In: *Technological Forecasting and Social Change* 128 (Mar. 2018), pp. 104–117. ISSN: 0040-1625. DOI: 10.1016/j.techfore.2017.11.001. URL: <https://www.sciencedirect.com/science/article/pii/S0040162517303736> (visited on 04/20/2025).
- [207] World Bank et al. *World Bank Global Electrification Database*. 2022. URL: <https://trackingsdg7.esmap.org/> (visited on 03/22/2022).
- [208] World Bank. *Private Participation in Infrastructure (PPI)*. 2022. URL: <https://ppi.worldbank.org/en/ppi> (visited on 03/11/2022).
- [209] World Bank. *Gini index*. en. 2022. URL: <https://data.worldbank.org> (visited on 03/23/2022).
- [210] Center for International Earth Science and Information Network - CIESIN - Columbia University. *Low Elevation Coastal Zone (LECZ) Urban-Rural Population and Land Area Estimates, Version 2*. 2013. URL: <https://sedac.ciesin.columbia.edu/data/set/leczy-urban-rural-population-land-area-estimates-v2>.

- [211] World Bank. *Agricultural land (sq. km)*. en. 2021. URL: <https://data.worldbank.org> (visited on 03/22/2022).
- [212] World Bank. *Land area (sq. km)*. en. 2022. URL: <https://data.worldbank.org> (visited on 03/21/2022).
- [213] World Bank. *Forest area (sq. km)*. en. 2022. URL: <https://data.worldbank.org> (visited on 03/22/2022).
- [214] Fabian Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *J. Mach. Learn. Res.* 12.null (Nov. 2011), pp. 2825–2830. ISSN: 1532-4435.
- [215] Christopher M Gotwalt. *JMP Neural Network Methodology*. Tech. rep. SAS Institute. URL: <https://community.jmp.com/t5/Technical-Reports/JMP-Neural-Network-Methodology/ta-p/637085>.
- [216] Trevor Hastie, Jerome Friedman, and Robert Tibshirani. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY: Springer, 2001. ISBN: 978-1-4899-0519-2 978-0-387-21606-5. URL: <http://link.springer.com/10.1007/978-0-387-21606-5> (visited on 04/24/2025).
- [217] ETA Florence Renewable Energies. *Photovoltaic Roofs Programme*. URL: <https://new.etaflorence.it/projects/photovoltaic-roofs-programme/> (visited on 01/02/2024).
- [218] Italian Government. *Photovoltaic Roofs Program*. it. 2001. URL: <https://www.gazzettaufficiale.it/eli/id/2001/03/29/001A3359/sg> (visited on 12/11/2023).
- [219] Aldo Orioli and Alessandra Di Gangi. “Six-years-long effects of the Italian policies for photovoltaics on the grid parity of grid-connected photovoltaic systems installed in urban contexts”. In: *Energy* 130 (July 2017), pp. 55–75. ISSN: 0360-5442. DOI: 10.1016/j.energy.2017.04.069. URL: <https://www.sciencedirect.com/science/article/pii/S0360544217306357> (visited on 04/20/2025).
- [220] ofgem. *Feed-in Tariffs (FIT) | Ofgem*. en. URL: <https://www.ofgem.gov.uk/environmental-and-social-schemes/feed-tariffs-fit> (visited on 01/07/2024).
- [221] L. Dusonchet and E. Telaretti. “Comparative economic analysis of support policies for solar PV in the most representative EU countries”. In: *Renewable and Sustainable Energy Reviews* 42 (Feb. 2015), pp. 986–998. ISSN: 1364-0321. DOI: 10.1016/j.rser.2014.10.054. URL: <https://www.sciencedirect.com/science/article/pii/S136403211400879X> (visited on 04/20/2025).
- [222] AURES II. *Auctions for the support of renewable energy in Mexico*. Tech. rep. 2019. URL: http://aures2project.eu/wp-content/uploads/2019/12/AURES_II_case_study_Mexico.pdf (visited on 04/20/2025).
- [223] Pablo del Río. *Auctions for Renewable Support in Mexico: Instruments and lessons learnt*. 2017. URL: http://aures2project.eu/wp-content/uploads/2021/07/mexico_final.pdf (visited on 04/20/2025).
- [224] Emiliano Bellini. *Mexico eliminates 15% customs duties on solar module imports*. en-US. June 2018. URL: <https://www.pv-magazine.com/2018/06/18/mexico-eliminates-15-customs-duties-on-solar-module-imports/> (visited on 01/08/2024).

- [225] John S. Duffield. “The politics of renewable power in Spain”. en. In: *European Journal of Government and Economics* 9.1 (May 2020), pp. 5–25. ISSN: 2254-7088. DOI: 10.17979/ejge.2020.9.1.5231. URL: <https://revistas.udc.es/index.php/ejge/article/view/ejge.2020.9.1.5231> (visited on 04/20/2025).
- [226] Enrique Rosales-Asensio et al. “An expert judgement approach to determine measures to remove institutional barriers and economic non-market failures that restrict photovoltaic self-consumption deployment in Spain”. In: *Solar Energy* 180 (Mar. 2019), pp. 307–323. ISSN: 0038-092X. DOI: 10.1016/j.solener.2019.01.031. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X19300404> (visited on 04/20/2025).
- [227] CLIFFORD M. HURVICH and CHIH-LING TSAI. “Regression and time series model selection in small samples”. In: *Biometrika* 76.2 (June 1989), pp. 297–307. ISSN: 0006-3444. DOI: 10.1093/biomet/76.2.297. URL: <https://doi.org/10.1093/biomet/76.2.297> (visited on 09/15/2025).
- [228] International Energy Agency. “Renewable Energy Market Update - June 2023”. en. In: (2023).
- [229] Hussah Alghanem and Alastair Buckley. “Global Benchmarking and Modelling of Installed Solar Photovoltaic Capacity by Country”. en. In: *Energies* 17.8 (Jan. 2024), p. 1812. ISSN: 1996-1073. DOI: 10.3390/en17081812. URL: <https://www.mdpi.com/1996-1073/17/8/1812> (visited on 09/23/2024).
- [230] Copernicus Climate Change Service. *Climate and energy indicators for Europe from 1979 to present derived from reanalysis*. en. 2020. URL: <https://cds.climate.copernicus.eu/datasets/sis-energy-derived-reanalysis?tab=overview> (visited on 09/18/2023).
- [231] European Environment Agency. *CORINE Land Cover*. en. URL: <https://land.copernicus.eu/en/products/corine-land-cover> (visited on 09/12/2023).
- [232] Eurostat. *Home - Eurostat*. URL: <https://ec.europa.eu/eurostat> (visited on 07/24/2024).
- [233] IRENA (International Renewable Energy Agency). *Statistics Time Series*. en. URL: <https://www.irena.org/Data/View-data-by-topic/Capacity-and-Generation/Statistics-Time-Series> (visited on 04/16/2024).
- [234] *Common classification of territorial units for statistics (NUTS) | Fact Sheets on the European Union | European Parliament*. en. Feb. 2024. URL: <https://www.europarl.europa.eu/factsheets/en/sheet/99/common-classification-of-territorial-units-for-statistics-nuts-> (visited on 07/30/2024).
- [235] Adrien Moiret-Guigand et al. *Validation report – CORINE Land Cover 2018 and CORINE Land Cover Change 2012-2018*. Tech. rep. GMES Initial Operations/Copernicus Land monitoring services, Feb. 2021. URL: <https://land.copernicus.eu/en/technical-library/clc-2018-and-clc-change-2012-2018-validation-report> (visited on 07/30/2024).

- [236] Linda Aune-Lundberg and Geir-Harald Strand. “The content and accuracy of the CORINE Land Cover dataset for Norway”. In: *International Journal of Applied Earth Observation and Geoinformation* 96 (Apr. 2021), p. 102266. ISSN: 1569-8432. DOI: 10.1016/j.jag.2020.102266. URL: <https://www.sciencedirect.com/science/article/pii/S0303243420309090> (visited on 07/30/2024).
- [237] Ana Pérez-Hoyos, Francisco Javier García-Haro, and Nuria Valcárcel. “Incorporating Sub-Dominant Classes in the Accuracy Assessment of Large-Area Land Cover Products: Application to GlobCover, MODISLC, GLC2000 and CORINE in Spain”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7.1 (Jan. 2014). Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, pp. 187–205. ISSN: 2151-1535. DOI: 10.1109/JSTARS.2013.2258659. URL: <https://ieeexplore.ieee.org/document/6517278> (visited on 07/30/2024).
- [238] Rasim Latifovic and Ian Olthof. “Accuracy assessment using sub-pixel fractional error matrices of global land cover products derived from satellite data”. In: *Remote Sensing of Environment* 90.2 (Mar. 2004), pp. 153–165. ISSN: 0034-4257. DOI: 10.1016/j.rse.2003.11.016. URL: <https://www.sciencedirect.com/science/article/pii/S0034425703003729> (visited on 07/30/2024).
- [239] Government Data Quality Hub. *Meet the data quality dimensions*. en. June 2021. URL: <https://www.gov.uk/government/news/meet-the-data-quality-dimensions> (visited on 07/25/2024).
- [240] Alessandro Rovetta. “Raiders of the Lost Correlation: A Guide on Using Pearson and Spearman Coefficients to Detect Hidden Correlations in Medical Sciences”. In: *Cureus* 12.11 (Nov. 2020), e11794. ISSN: 2168-8184. DOI: 10.7759/cureus.11794. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7779167/> (visited on 05/02/2025).
- [241] Fariha Imam, Petr Musilek, and Marek Z. Reformat. “Parametric and Nonparametric Machine Learning Techniques for Increasing Power System Reliability: A Review”. en. In: *Information* 15.1 (Jan. 2024), p. 37. ISSN: 2078-2489. DOI: 10.3390/info15010037. URL: <https://www.mdpi.com/2078-2489/15/1/37> (visited on 04/23/2025).
- [242] Ravid Shwartz-Ziv and Amitai Armon. “Tabular data: Deep learning is not all you need”. In: *Information Fusion* 81 (May 2022), pp. 84–90. ISSN: 1566-2535. DOI: 10.1016/j.inffus.2021.11.011. URL: <https://www.sciencedirect.com/science/article/pii/S1566253521002360> (visited on 04/23/2025).
- [243] Robert Gilchrist and Denise Drinkwater. “The use of the Tweedie distribution in statistical modelling”. en. In: *COMPSTAT*. Ed. by Jelke G. Bethlehem and Peter G. M. van der Heijden. Heidelberg: Physica-Verlag HD, 2000, pp. 313–318. ISBN: 978-3-642-57678-2. DOI: 10.1007/978-3-642-57678-2_39.
- [244] Christoph F. Kurz. “Tweedie distributions for fitting semicontinuous health care utilization cost data”. en. In: *BMC Medical Research Methodology* 17.1 (Dec. 2017), p. 171. ISSN: 1471-2288. DOI: 10.1186/s12874-017-0445-y. URL: <https://doi.org/10.1186/s12874-017-0445-y> (visited on 08/29/2024).

- [245] Peter K. Dunn. “Occurrence and quantity of precipitation can be modelled simultaneously”. en. In: *International Journal of Climatology* 24.10 (2004), pp. 1231–1239. ISSN: 1097-0088. DOI: 10.1002/joc.1063. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/joc.1063> (visited on 09/03/2024).
- [246] Gordon K. Smyth and Bent Jørgensen. “Fitting Tweedie’s Compound Poisson Model to Insurance Claims Data: Dispersion Modelling”. In: *ASTIN Bulletin* 32.1 (2002). Publisher: Cambridge University Press, pp. 143–157. URL: https://econpapers.repec.org/article/cupastinb/v_3a32_3ay_3a2002_3ai_3a01_3ap_3a143-157_5f01.htm (visited on 09/03/2024).
- [247] Scott M. Lundberg and Su-In Lee. “A unified approach to interpreting model predictions”. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS’17. Red Hook, NY, USA: Curran Associates Inc., Dec. 2017, pp. 4768–4777. ISBN: 978-1-5108-6096-4. (Visited on 09/16/2024).
- [248] Christoph Molnar. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 3rd ed. 2025. ISBN: 978-3-911578-03-5. URL: <https://christophm.github.io/interpretable-ml-book>.
- [249] Dirk-Jan van de Ven et al. “The potential land requirements and related land use change emissions of solar energy”. en. In: *Scientific Reports* 11.1 (Feb. 2021). Publisher: Nature Publishing Group, p. 2907. ISSN: 2045-2322. DOI: 10.1038/s41598-021-82042-5. URL: <https://www.nature.com/articles/s41598-021-82042-5> (visited on 06/27/2024).
- [250] Pengyi Zhang et al. “Revisiting the land use conflicts between forests and solar farms through energy efficiency”. In: *Journal of Cleaner Production* 434 (Jan. 2024), p. 139958. ISSN: 0959-6526. DOI: 10.1016/j.jclepro.2023.139958. URL: <https://www.sciencedirect.com/science/article/pii/S0959652623041161> (visited on 07/21/2024).
- [251] Boqiang Lin and Suppawit Kaewkhunok. “The role of socio-Culture in the solar power adoption: The inability to reach government policies of marginalized groups”. In: *Renewable and Sustainable Energy Reviews* 144 (July 2021), p. 111035. ISSN: 1364-0321. DOI: 10.1016/j.rser.2021.111035. URL: <https://www.sciencedirect.com/science/article/pii/S1364032121003257> (visited on 06/26/2024).
- [252] Shah Rukh Shakeel et al. “Solar PV adoption at household level: Insights based on a systematic literature review”. In: *Energy Strategy Reviews* 50 (Nov. 2023), p. 101178. ISSN: 2211-467X. DOI: 10.1016/j.esr.2023.101178. URL: <https://www.sciencedirect.com/science/article/pii/S2211467X23001281> (visited on 04/23/2025).
- [253] Marie Briguglio and Glenn Formosa. “When households go solar: Determinants of uptake of a Photovoltaic Scheme and policy insights”. In: *Energy Policy* 108 (Sept. 2017), pp. 154–162. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2017.05.039. URL: <https://www.sciencedirect.com/science/article/pii/S0301421517303233> (visited on 04/23/2025).

- [254] Syamal K. Sen and Ravi P. Agarwal. “3 - History of zero including its representation and role”. In: *Zero*. Ed. by Syamal K. Sen and Ravi P. Agarwal. Academic Press, Jan. 2016, pp. 29–75. ISBN: 978-0-08-100774-7. DOI: 10.1016/B978-0-08-100774-7.00003-X. URL: <https://www.sciencedirect.com/science/article/pii/B978008100774700003X> (visited on 05/05/2025).
- [255] SolarPower Europe. *Floating PV Best Practice Guidelines Version 1.0*. en. Tech. rep. 2023. URL: https://api.solarpowereurope.org/uploads/3323_SPE_Floating_PV_report_02_mr_74f6db82ca.pdf?updated_at=2023-12-07T08:41:39.869Z (visited on 07/22/2024).
- [256] Government of Extremadura. *Plan Extremeño Integrado de Energía y Clima (PEIEC) 2021-2030*. es-ES. 2021. URL: <https://www.juntaex.es/w/actuacion-plan-extremeno-integrado-de-energia-y-clima-peiec-2021-2030> (visited on 09/09/2024).
- [257] Aurora Energy Research. *Grid management challenges costing Spanish energy consumers*. en-GB. Apr. 2023. URL: <https://auroraer.com/media/grid-management-challenges-costing-spanish-energy-consumers/> (visited on 10/09/2024).
- [258] Olivier De Groote, Guido Pepermans, and Frank Verboven. “Heterogeneity in the adoption of photovoltaic systems in Flanders”. In: *Energy Economics* 59 (Sept. 2016), pp. 45–57. ISSN: 0140-9883. DOI: 10.1016/j.eneco.2016.07.008. URL: <https://www.sciencedirect.com/science/article/pii/S0140988316301803> (visited on 09/15/2024).
- [259] Interreg Europe. *ASTER - Access to Sustainability for Tenants through Energy Effective Retrofit*. en. May 2024. URL: <https://www.interregeurope.eu/good-practices/aster-access-to-sustainability-for-tenants-through-energy-effective-retrofit-0> (visited on 09/13/2024).
- [260] ASTER. *Cooperative company - ASTER*. en. URL: <https://aster.vlaanderen/nl/wat-doet-aster-cv/cooperatieve-vennootschap> (visited on 09/13/2024).
- [261] Federal Ministry for Economic Affairs and Energy. *Stepping up the PV roll-out*. en. June 2023. URL: <https://www.bmwk-energiewende.de/EWD/Redaktion/EN/Newsletter/2023/05/Meldung/direkt-answers.html> (visited on 06/05/2025).
- [262] German Federal Statistical Agency. *Wem gehört die Landwirtschaft? Bedeutung von Unternehmensgruppen erstmals untersucht*. en. 2021. URL: https://www.destatis.de/DE/Presse/Pressemitteilungen/2021/07/PD21_N047_41.html (visited on 09/22/2024).
- [263] National Energy System Operator (NESO). *Connections Reform*. en-gb. URL: <https://www.neso.energy/industry-information/connections/connections-reform> (visited on 01/30/2025).
- [264] National Energy System Operator (NESO). *Beyond 2030: A national blueprint for a decarbonised electricity system in Great Britain*. Tech. rep. Mar. 2024. URL: <https://www.neso.energy/document/315516/download> (visited on 12/31/2024).
- [265] Felicia Rankl. *Planning for solar farms*. Tech. rep. May 2024. URL: <https://researchbriefings.files.parliament.uk/documents/CBP-7434/CBP-7434.pdf> (visited on 12/24/2024).

- [266] Department for Energy Security & Net Zero (DESNZ) and Office of Gas and Electricity Markets (Ofgem). *Open letter from DESNZ and Ofgem: Aligning grid connections with strategic plans (5 November 2024)*. en. Nov. 2024. URL: <https://www.gov.uk/government/publications/aligning-grid-connections-with-strategic-plans/open-letter-from-desnz-and-ofgem-aligning-grid-connections-with-strategic-plans-5-november-2024> (visited on 01/15/2025).
- [267] Environmental Audit Committee. *Enabling sustainable electrification of the economy*. Tech. rep. May 2024. URL: <https://committees.parliament.uk/publications/45077/documents/223429/default/> (visited on 01/15/2025).
- [268] Andrew Meyer. “Does education increase pro-environmental behavior? Evidence from Europe”. In: *Ecological Economics* 116 (Aug. 2015), pp. 108–121. ISSN: 0921-8009. DOI: 10.1016/j.ecolecon.2015.04.018. URL: <https://www.sciencedirect.com/science/article/pii/S0921800915001998> (visited on 01/27/2025).
- [269] Qi Wang et al. “Green returns to education: Does education affect pro-environmental attitudes and behaviors in China?” In: *PLoS ONE* 17.2 (Feb. 2022), e0263383. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0263383. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8812898/> (visited on 01/27/2025).
- [270] Ferhat Özbay and Ibrahim Duyar. “Exploring the role of education on environmental quality and renewable energy: Do education levels really matter?” In: *Current Research in Environmental Sustainability* 4 (Jan. 2022), p. 100185. ISSN: 2666-0490. DOI: 10.1016/j.crsust.2022.100185. URL: <https://www.sciencedirect.com/science/article/pii/S2666049022000639> (visited on 01/27/2025).
- [271] Kanonier, Arthur et al. *Spatial Planning in Austria with References to Spatial Development and Regional Policy*. Austrian Conference on Spatial Planning (ÖROK) Publication Series. Vienna: Office of the Austrian Conference on Spatial Planning (ÖROK), 2018. ISBN: 978-3-9504146-3-9.
- [272] Office for National Statistics. *Regional gross value added (balanced) by industry: all ITL regions*. Apr. 2024. URL: <https://www.ons.gov.uk/economy/grossvalueaddedgva/datasets/nominalandrealregionalgrossvalueaddedbalancedbyindustry> (visited on 11/28/2024).
- [273] Ofgem. *Renewables Obligation (RO) buy-out price, mutualisation threshold and mutualisation ceilings for 2024 to 2025*. en. Apr. 2024. URL: <https://www.ofgem.gov.uk/publications/renewables-obligation-ro-buy-out-price-mutualisation-threshold-and-mutualisation-ceilings-2024-2025> (visited on 12/05/2024).
- [274] Sheffield Solar. *Sheffield Solar API Platform: Capacity Data*. Oct. 2024. URL: <https://api.solar.sheffield.ac.uk/capacity/> (visited on 12/11/2024).
- [275] Office for National Statistics. *International, regional and city statistics*. URL: <https://www.ons.gov.uk/aboutus/whatwedo/programmesandprojects/europeancitystatistics> (visited on 01/20/2025).