

Modeling Residential Adoption of Solar Photovoltaic Systems



Carolina Goldstein, José Miguel Espinosa, and Regina Bispo 

Abstract The world is on an urgent transition to renewable energies. Photovoltaic (PV) solar energy is the most viable green energy source to be produced at the domestic level, allowing every individual to contribute. Understanding the factors that influence the adoption of domestic solar energy, how it changes throughout the country and how spatial dependent factors contribute to the promotion of this technology is of the utmost importance to stimulate adoption. As to this day, to the best of my knowledge, these are not yet known. This study aims to contribute to channeling efforts to where adoption is more likely, ultimately accelerating Portugal's energy transition. Hence, the goal of this study is to build a spatial model that estimates for each spatial unit in Portugal the probability of individuals adopting domestic PV systems. The study uses data related to past solar PV installations as well as socioeconomic and demographic data from public sources. An exploratory spatial analysis including the study of spatial correlation across municipalities confirmed the importance of spatial considerations. Three dependent variables were considered sequentially: installations (binary), number of panels installed (discrete), and installed power (continuous). To model the latter, it being the main focus of the study, eight models were compared: linear regression (OLS), spatial lag (SAR), spatial error (SEM), Kelejian-Prucha (GSM), spatial lag of the explanatory variables (SLX), spatial Durbin (SDM), spatial Durbin error (SDEM), and Manski models. It was concluded that socioeconomic factors do spill over to neighbor locations and in

C. Goldstein (✉)

MSc in Analytics and Big Data Engineering, Department of Computer Science and Department of Mathematics, NOVA School of Science and Technology, NOVA University of Lisbon, Campus de Caparica, 2829-516 Caparica, Portugal
e-mail: c.goldstein@campus.fct.unl.pt

J. M. Espinosa

Tech Garage, IT & Digital Department. Galp Energia, SGPS, S.A., 1600-209 Lisboa, Portugal

R. Bispo

NOVAMATH Center for Mathematics and Applications, Department of Mathematics, NOVA School of Science and Technology, NOVA University of Lisbon, Campus de Caparica, 2829-516 Caparica, Portugal
e-mail: r.bispo@fct.unl.pt

that way influence solar PV adoption, but also that unobserved characteristics result in similar decisions in nearby municipalities. The SDEM was found to be best to fit the data and a final map representing the likelihood of adoption across the different municipalities in Portugal was produced according to its estimations.

Keywords PV system adoption · Social effects · Spatial modeling · Technology diffusion

1 Introduction

Rethinking the use of energy stemming from fossil sources and transitioning to renewable energies is increasingly becoming a necessity. Photovoltaic (PV) energy, attained through the installation of solar panels, is the most viable of being produced at the level of the individual consumer for domestic use. There are companies developing highly advanced technologies to identify the energetic potential of homes and to install these solar panels. However, inquiring about potential customers without knowing their predisposition ends up wasting many resources.

The overall goal of this project is the construction of a spatial model that estimates for each spatial unit, with the finest possible granularity, the probability of adopting domestic solar PV systems. In doing so, companies will be able to better channel their selling efforts to locations where adherence is more likely, ultimately accelerating Portugal's transition to renewable energies.

More specifically, using data related to past solar panel installations, the first goal is to describe the geographical distribution of the current installations across Portugal. Furthermore, using socioeconomic and demographic data from public sources, the goal is to cross this information and characterize each region, in order to understand the factors that may explain the decision of installing solar panels. The ultimate goal is to build a map representing the adoption likelihood for each spatial unit.

This study is structured as follows: in the first section the topic is put into context and the goals of the project are defined. The motivations for this work are also presented in this chapter, as well as a review of the literature on similar problems. In Sect. 2 the available data that are to be used are presented, along with their description, characterization, and preprocessing. This section also presents the statistical methods to be used, both to perform an initial exploratory analysis and to build different model specifications, while explaining the logic that resulted in the presented decisions. In Sect. 3, the results of the exploratory analysis and the different regression models are shown and described. In Sect. 4, the results are discussed and conclusions are presented.

1.1 *Decision-Making in PV Technology Adoption*

Schelly [1] explores the decision-making process of individuals regarding energy technology adoption through interviews with domestic PV panel owners and indicates three models to explain adoption: environmental motivations, economic rationality, and social spillover. Richter [2] studied the diffusion of solar PV technology in the United Kingdom through a panel model with time-varying fixed effects and found that higher educated neighborhoods installed more PV systems than neighborhoods with, on average, lower educated populations. The author also found a correlation between the number of systems installed in an area and the number installed three months later. Hence, Richter [2] concludes that higher educated neighborhoods may be more inclined to promote the spread of technology within their neighborhoods. Bollinger and Gillingham [3], who studied the diffusion of solar PV panels in California with a similar panel model, also found significant evidence that the decision to install PV systems may be influenced by the neighbors' previous decision to install. Graziano and Gillingham [4] examined the diffusion of this technology in Connecticut in a similar way and found that demographic and socioeconomic variables significantly influenced PV adoption and that higher numbers of previously installed systems also significantly increased the number of later adoptions nearby. Schelly and Letzelter [5] examined the decision factors that influence the adoption of residential solar electric power systems in upstate New York through questionnaire data and found that environmental motivations are slightly more important than economics. As Richter [2] points out, spatial econometric methods could allow the study of social effects across borders, recognizing the study of spillover only within the neighborhood as a limitation of her model. Baginski and Weber [6] use spatial econometric models to study the spread of PV systems over space and the factors that drive the regional uptake in Germany to conclude that spatial dependence is a relevant factor for explaining regional clusters of PV adoption and that spatial spillover is not mainly driven by social imitation but by unobserved regional characteristics. High values for solar radiation, the share of detached houses, electricity demand, and inverse population density of a region favor the PV uptake. Predicting that also in the case of this study, the demographic and socioeconomic factors as well as built environment associated with each region will be key to mapping the country's regions and identifying which are more likely to be receptive to domestic solar panels, these variables were extracted from publicly available sources to test how they fit the data. The approach of Baginski and Weber [6] will be closely followed, adding a predictive component using the results found to build a map representing the likelihood of adoption, making it more directly usable by decision-makers in the field.

Most studies that try to explain the factors influencing PV system adoption use the number of PV systems as the target variable to be explained [2–4, 7, e.g.]. Some, as in the case of Rode and Weber [8], use a variation of this discrete variable, like the number of PV installations per building and number of PV installations per owner-occupied household, transforming the target variable into a ratio and therefore essentially continuous. Others, like Baginski and Weber [6] and Schaffer and Brun

[9], use the PV installed capacity (in kw), which is a continuous variable. Naturally, the methods used in each approach will also differ accordingly. In the case of this study, there are three variables that could be used as target variables, namely *Total Price*, *Installed Power* and *Number of Panels*, since all represent the size of PV system installations. Hence, and since the installed capacity has been used by other authors, this study will entail both the analysis of *Installed Power* and *Number of Panels*. Thus analyzing both a continuous and a discrete variable.

Baginski and Weber [6] focus on the spatial diffusion of PV systems using spatial econometric models and considering both exogenous and endogenous spatial interactions. To follow this recommendation, this study will first perform a spatial exploratory analysis. Many spatial analysis authors refer to Tobler's first law of geography, which states that areas closer together are more similar than those further apart ("*the first law of geography: everything is related to everything else, but near things are more related than distant things.*" [10]). For that reason, most spatial analysis start by exploring spatial correlation, which implies the correlation among the same variable from different locations.

Spatial dependence is commonly made operational by some measure of spatial autocorrelation, which depend on the specification or estimation of a set of weights describing spatial relationships. To describe possible spatial relationships between locations, one must first define what accounts for neighbors of said locations. Some typical examples of criteria that could be used to define neighbors were described by Anselin [11], namely first-order contiguity and critical distance thresholds. Part of assigning neighbors involves applying a measure of weighting to indicate the extent to which the information from an area's neighbors impacts on the observed estimate for that area. This is commonly summarized in a spatial weights matrix.

2 Material and Methods

2.1 Data Characterization and Preprocessing

There are two important data sets to consider for the construction of the models. The main data set contains details from 441 domestic solar panel installations done in Portugal between the end of June and November of 2020, provided by a company that specializes in such installations. The second data set involves demographic and socioeconomic variables extracted from *Instituto Nacional de Estatística (INE: www.ine.pt)*. These variables were downloaded as isolated data sets and then aggregated by geographical location. The selection of the variables was based on the factors found in the literature to influence the decision to install solar panels. These were then subjected to a correlation analysis to select the final list.

The available variables regarding the solar PV installations, their type, and their meaning are described in Table 1.

Table 1 Description of the variables regarding solar PV installations

Variable	Type	Description
Date	Categorical	Date of the installation
Postal code	Categorical	Postal code of the house of installation
Number of panels	Numerical	Number of panels installed
Installed power	Numerical	Power installed, in kwh

Data preprocessing tasks included both data exclusion and variable creation. The data exclusion task involved removing some values that do not make sense and are likely to be database mistakes. The geographical aspect of the data is very important to pursue the goals of this paper. Hence, since the only variable containing geographic information in the dataset is *Postal Code*, this information was expanded to include *Locality*, *Municipality*, *District*, *NUTSII*, and *NUTSIII*, creating these 5 new variables. *NUTS* refers to the Nomenclature of Territorial Units for Statistics, and it is a standard for referencing subdivisions of European countries. *NUTSI* represents major socioeconomic regions, which corresponds to three regions in Portugal. *NUTSII* refers to basic regions for the application of regional policies and is made up of seven regions in Portugal, five if the islands are excluded. *NUTSIII* represents smaller regions and corresponds to 25 regions in Portugal, 23 of these in continental Portugal.

The list of selected explanatory variables and their description can be seen in Table 2. A summary of descriptive statistics can be seen in Table 4 (Table 3).

From the initial set of 45 variables that have data at the municipality level, these 13 were selected based primarily on correlation analysis.

2.2 Data Modeling

Spatial Weights Matrix Spatial weights represent geographic relationships between the different units in a spatially referenced dataset, usually in the form of a spatial weights matrix. This is defined as a $n \times n$ positive matrix W with elements w_{ij} at location pairs i, j ($i \neq j; i, j = 1, \dots, n$) for n locations. An element w_{ij} is the weight for each pair of locations, which is assigned by some rules that define the spatial relations between the locations.

There are several ways to define this matrix, commonly based on contiguity. A pair of spatial units is said to be contiguous if they share a common border. Rook contiguity constructs a weight object from a collection of polygons that share at least one edge. Queen contiguity is a more inclusive notion of contiguity, since it requires a pair of polygons to share one or more vertices.

Table 2 Description of explanatory variables

Variable	Description
Population	Number of people who live in each municipality
Purchase power	Purchase power per person who lives in each municipality
Gross income Gini	Gini coefficient calculated per taxable persons
Gross income	Gross income per person who lives in each municipality
Subsidies	Number of people who receive government subsidies per person who lives in each municipality
Rental agreements	Number of rental agreements per person who lives in each municipality
Energy consumption	Domestic consumption of electrical energy per consumer (kWh/consumer), where consumed energy might have been produced by hydroelectric, nuclear or thermal conventional centrals or also wave, mares, wind or solar energy
Votes in the most voted	Percentage of votes in the most voted in elections for the Assembly of the Republic
Abstention	Percentage of abstention in the elections for the Assembly of the Republic
Temperature	Average temperature in the last available year
Art exhibitions	Number of art gallery exhibitions per person who lives in each municipality
Family housing	Number of classic family housing
Habitation buildings	Number of habitation buildings

Table 3 Results for the spatial dependence tests in the OLS model

Test	Statistic	P-value
Moran (residuals)	0.0856	0.002021
LMerr	5.5948	0.01801
LMlag	1.4397	0.2302
RLMerr	6.1968	0.0128
RLMlag	2.0417	0.153
SARMA	7.6365	0.02197

Since there is very little information available about what type of relation would make a municipality in Portugal influence one more than another and to make the least number of assumptions, a k -nearest neighbors matrix will be chosen and k decided based on the most common number of neighbors a municipality has in Portugal (discovered through rook and queen contiguity).

Spatial Correlation An important part of spatial analysis is the particular analysis of spatial correlation. Popular options for area-level data that will be considered include Moran's I, Geary's C, Gettis and Ord's G and the Localized Indicators of

Table 4 Summary description of explanatory variables

Variable	Count	Mean	RSD ⁴ (%)	Min	Median	Max
Population	278	35247.69	57526.42	1634	14608	509515
Purchase power	278	80.51	18.68	55.32	77.21	219.63
Gross income Gini	272	26.89	2.34	21.10	26.70	37.20
Gross income	278	7608.82	1726.30	4352.00	7384.50	19574.00
Subsidies	278	0.03	0.01	0.01	0.03	0.13
Rental agreements	182	0.006	0.003	0.002	0.006	0.016
Energy consumption	278	2082.81	712.62	1026.00	2082.15	10393.00
Votes in the most voted	278	40.64	4.58	31.20	40.10	61.50
Abstention	278	47.35	5.79	30.40	46.50	66.20
Temperature	278	15.43	1.37	11.70	15.40	18.00
Art exhibitions	278	0.0002	0.0002	0.0000	0.0001	0.0018
Family housing	278	0.74	0.26	0.39	0.70	2.95
Habitation buildings	278	0.60	0.27	0.07	0.58	1.61

Spatial Association (LISA). It is important to be attentive of the distinction between global and localized correlation. Some methods study global clustering (like Moran’s I), which assesses spatial correlation throughout the entire study region. Localized correlation is also called local clustering or hot-spot analysis and includes methods such as LISA.

To model adoption of solar PV systems across Portugal, this study will start by estimating a binary dependent variable, namely whether there are PV installations in a certain region or not. Taking this a step further, a discrete dependent variable will be modeled, specifically the number of panels installed in each region. Finally we will focus more attentively on modeling a continuous dependent variable, namely the installed power.

For all the following models in this chapter, this notation is used:

- n is the number of observations;
- K is the number of explanatory variables;
- \mathbf{Y} is a $n \times 1$ vector of observations on the dependent variable;
- \mathbf{X} is a $n \times K$ matrix of observations on the explanatory variables with an associated vector of regression coefficients β ($K \times 1$);
- \mathbf{W} is the spatial weights matrix ($n \times n$);

- WY denotes the endogenous interactions among the dependent variable associated with the spatial autoregressive parameter ρ , which measures the effect of spatial lag on this variable;
- WX denotes the exogenous interactions among the independent variables associated with a vector of regression coefficients γ ($K \times 1$);
- Wu denotes interactions among the residuals of spatial units, associated with the spatial autocorrelation parameter λ ;
- ε represents an independently identically normally distributed error term vector with zero mean and constant variance ($\varepsilon \sim N_n(0, \sigma^2 I_n)$);
- α represents the models' intercept.

Binary Dependent Variable To model the binary dependent variable *Installation*, the probit model will be used, as it is typically privileged in econometrics.

This is a particular type of binary regression model that ultimately allows to classify the observations based on their predicted probabilities. Considering the generalized linear model framework, the probit model uses a probit link function and will be estimated using maximum likelihood. This is represented by the following equation

$$P(\mathbf{Y} = \mathbf{1} \mid X) = \Phi(X\beta) \tag{1}$$

where \mathbf{Y} is a vector of the binary outcome, $\mathbf{1}$ is a vector of ones, Φ is the cumulative distribution function of the standard normal distribution, and β is the vector of parameters β estimated by maximum likelihood.

To introduce a spatial component in this analysis, a Bayesian Estimation of Spatial Probit Models will be used. The Bayesian estimation of the spatial autoregressive probit model (SAR Probit model) is described by

$$\mathbf{Y} = \rho W\mathbf{Y} + X\beta + \varepsilon \tag{2}$$

with notation as previously described. Note that ρ is the scalar parameter that describes the strength of spatial dependence in the sample of observations.

The prior distributions are $\beta \sim N(c, T)$ and $\rho \sim Beta(a_1, a_2)$, where c is the mean value of β , T is the variance of β , while a_1 and a_2 are shape parameters.

In general the coefficients of any probit regression cannot be interpreted directly. The marginal effects of the regressors should be considered partial derivatives. Additionally, in the case of the SAR Probit model, the direct, indirect, and total effects are to be considered. A change in one explanatory variable x_{ki} for location i ($i = 1, \dots, n$) will not only affect the observations y_i directly (this is considered the direct impact), but this change can also affect the observations in locations nearby y_j (which is the indirect impact). Let $S_k(W)$ be the matrix ($n \times n$) of impacts from location i to location j for explanatory variable x_k , defined as

$$S_k(W) = \frac{dE[\mathbf{Y} \mid x_k]}{dx_k} = \phi((I_n - \rho W)^{-1} I_n \bar{x}_k \beta_k) * (I_n - \rho W)^{-1} I_n \beta_k \tag{3}$$

where \bar{x}_k denotes the mean value of variable x_k and β_k the parameter estimate for this variable.

Then the direct impact of a change in x_{ki} on y_i can be described as $S_k(W)_{ii}$ and the indirect impact from observation x_{kj} on y_i as $S_k(W)_{ij}$ ($i \neq j$). Hence, the average direct impact of k can be calculated as the average of the diagonal elements. The average total impact is the mean of the row sum, and the average indirect impacts can be calculated as the difference between average total impacts and average direct impacts.

Discrete Dependent Variable To model the Number of Panels installed across the municipalities, linear regressions were considered and estimated by OLS. The following models will be used

$$Y = X\beta + \varepsilon \tag{4}$$

$$Y = X\beta + WX\gamma + \varepsilon. \tag{5}$$

Equation 4 describes the linear regression estimated by OLS (referred to as OLS Model in Sect. 3), while Eq. 5 describes the SLX Model, which includes the spatially lagged explanatory variables, weighted by the spatial weights matrix.

Continuous Target Variable This study will take a general-to-specific approach, as suggested by Baginski and Weber [6], thus starting with a simple non-spatial linear regression and successively adding different spatial interaction effects. Still using OLS estimation, the model will be expanded with the spatial lag of the explanatory variables (SLX). The model will then be expanded with a spatially lagged dependent variable, thus estimating the spatial lag or spatial autoregressive model (SAR). The spatial error model (SEM) is also specified, incorporating spatial autoregressive process in the error term. Then, estimating a spatial durbin model (SDM) can be appropriate, where the SAR model is expanded with spatially lagged explanatory variables, as it seems reasonable to think that spatially correlated variables are probably omitted. For the same reason, the spatial durbin error model (SDEM) will also be estimated and compared. Finally, because the underlying spatial process is often unclear, all three spatial effects will be combined in the most general model, the Manski model. Here it is important to take into consideration that one of the components has to be removed for the spatial coefficients to be properly interpreted and distinguished [12].

It has been shown in LeSage and Pace [13] that a valid way to interpret the β coefficients in spatial econometric models is partial derivative interpretations of the impacts. The direct impact is the change in one location associated with the explanatory variable that affects that same region. The indirect effect is the potential effect that this explanatory variable has on all other regions it affects. The sum of both is the total effect. These impact measures are valid for models including a spatially lagged variable, thus in OLS and SEM the indirect effects are zero.

The first OLS estimation is the same as described for the discrete dependent variable in Eq. 4. Six different statistics for spatial dependence will be run to test for residual spatial dependence of the OLS regression:

- Moran's I test for residual spatial autocorrelation;
- simple LM test for error dependence (LMerr);
- simple LM test for a missing spatially lagged dependent variable (LMlag);
- variants of these robust to the presence of the other:
 - test for error dependence in the possible presence of a missing lagged dependent variable (RLMerr);
 - test for a missing spatially lagged dependent variable in the possible presence of error dependence (RLMlag);
- portmanteau test (SARMA, in fact LMerr + RLMlag).

The most straightforward way to include spatial dependence in a regression is by considering not only an explanatory variable, but also its spatial lag. This implies estimating the SLX model, described by Eq. 5.

The spatial lag model (SAR) introduces a spatial lag of the dependent variable, as seen in the following equation

$$\mathbf{Y} = \rho W\mathbf{Y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (6)$$

This model violates the exogeneity assumption, crucial for OLS to work and therefore a maximum likelihood estimation will be used.

The spatial error model (SEM) includes a spatial lag in the error term of the equation, resulting in the following term

$$\begin{aligned} \mathbf{Y} &= X\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \lambda W\mathbf{u} + \boldsymbol{\varepsilon}. \end{aligned} \quad (7)$$

As this specification violates the assumptions about the error term in a classical OLS model, a maximum likelihood will be used.

The spatial Durbin error model (SDEM) includes a spatial lag in the error term of the equation and the spatial lag of explanatory variables, resulting in

$$\begin{aligned} \mathbf{Y} &= X\boldsymbol{\beta} + WX\boldsymbol{\gamma} + \mathbf{u} \\ \mathbf{u} &= \lambda W\mathbf{u} + \boldsymbol{\varepsilon}. \end{aligned} \quad (8)$$

The spatial Durbin model (SDM) includes the spatial lag of the dependent variable and of the explanatory variables, resulting in

$$\mathbf{Y} = \rho W\mathbf{Y} + X\boldsymbol{\beta} + WX\boldsymbol{\gamma} + \boldsymbol{\varepsilon}. \quad (9)$$

The Kelejian-Prucha model (GSM) includes the spatial lag of the dependent variable and in the error term of the equation, resulting in

$$\begin{aligned} \mathbf{Y} &= \rho W\mathbf{Y} + X\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \lambda W\mathbf{u} + \boldsymbol{\varepsilon}. \end{aligned} \quad (10)$$

The Manski model is the most general model and includes the spatial lag of the dependent variable, in the error term of the equation and of the explanatory variables, resulting in

$$\begin{aligned} \mathbf{Y} &= \rho W\mathbf{Y} + X\boldsymbol{\beta} + WX\boldsymbol{\gamma} + \mathbf{u} \\ \mathbf{u} &= \lambda W\mathbf{u} + \boldsymbol{\varepsilon}. \end{aligned} \tag{11}$$

Model Selection In this study, AIC will be the main criteria used to select the best model. To do model diagnosis, residual plots will be produced for all the regressions. Some important aspects when analyzing the regression results estimated by OLS are the t and F statistics. To test for heteroskedasticity, the Breusch-Pagan test will be used. To compare different models, the Nagelkerke pseudo R-squared will be used.

Ultimately, this study sets out to build a spatial model that estimates, for each spatial unit in Portugal, the probability of adopting domestic solar PV systems. Hence, a map will be produced where each region has a value, in a scale that ranges from 0 to 1, representing the probability of a solar installation being adopted. This may be achieved by dividing the predicted value of Installed Power for each region by the total predicted value for the country.

3 Results

3.1 Exploratory Analysis

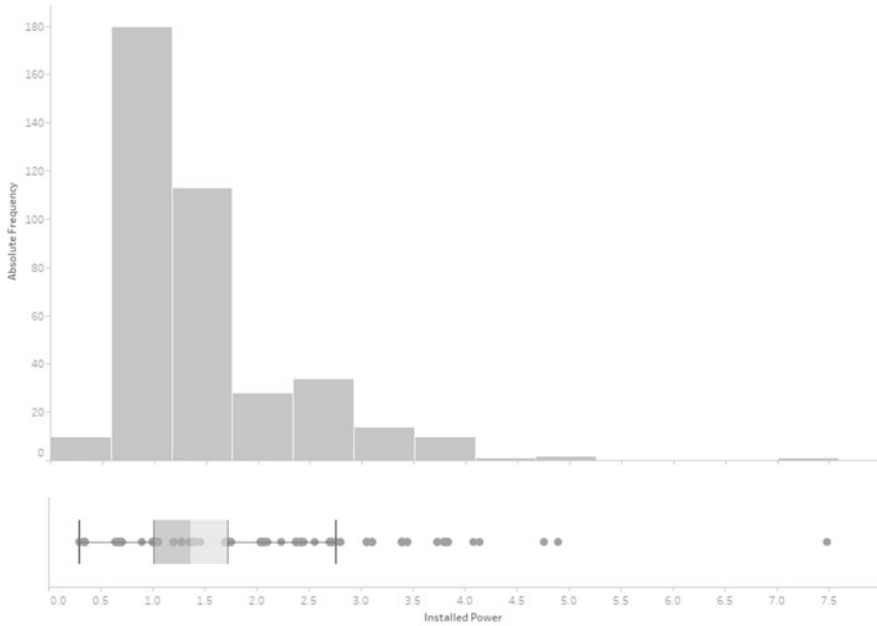
Dependent Variables Firstly, a general exploratory analysis is important to understand the distribution of the dependent variables. Bar plots and histograms were built to achieve this, as well as simple tables with descriptive statistics and maps to visualize their geographical distribution.

As mentioned before, there are three variables that will be considered to be dependent throughout this study. Installation is a binary variable that is 1 when a municipality has at least one installation. Number of Panels is a discrete variable that represents the total sum of panels. Installed Power is a continuous variable that contains the sum of installed power (in kwh). Since the patterns found in the variables Number of Panels and Installed Power are very similar, the graphs for Number of Panels are presented only in the Appendix. Indeed the similarity can be seen in the scatter plot of Fig. 7.

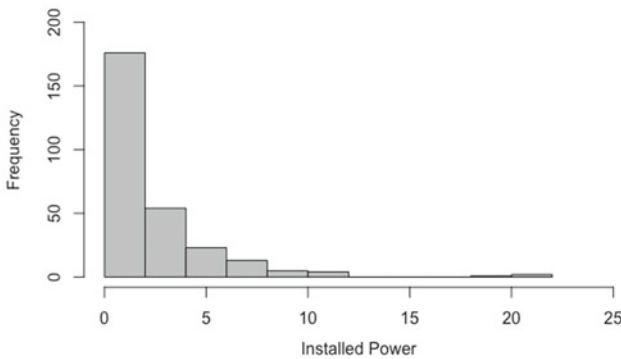
Most installations have 3 panels installed and half the installations have 4 or less panels, but the number of panels can vary between 1 and 22. The installed power ranges from 0.3 to 7.5, with half of the installations having around 1.4kwh or less. Both variables present a positive skewness at installation level, as can be seen in Figs. 1a and 8.

At municipality level, around 63% of regions have solar PV installations, hence the total number of panels and installed power show a large positive skewness and zero-inflation, as shown in Figs. 1b and 9 as well as Table 5.

Figure 2 shows that in general, the municipalities that do not have solar PV installations are mainly in the interior part of Portugal. The regions that do show some installations vary a lot in size, described by the Number of Panels and capacity, represented by the Installed Power. The biggest installations can be found in coastal



(a) Installed Power per installation



(b) Installed Power per municipality

Fig. 1 Distribution of installed power

Table 5 Summary description of numerical dependent variables

Variable	Count	Mean	RSD ⁵ (%)	Min	Median	Max
Number of panels	278	6.892	144.8758	0	3	65
Installed power (in kw)	278	2.057	148.9829	0.0	1.020	21.16

⁵Relative Standard Deviation or Coefficient of Variation

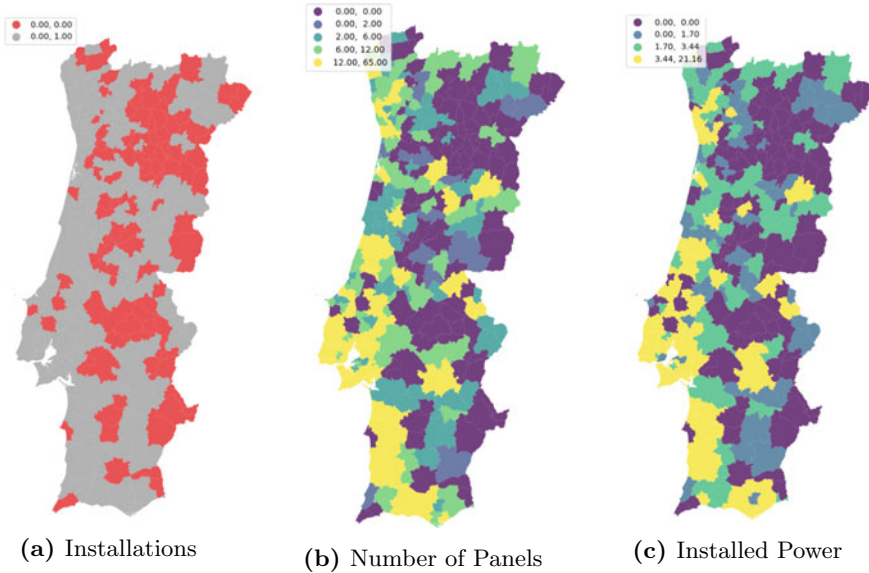


Fig. 2 Choropleth maps of dependent variables

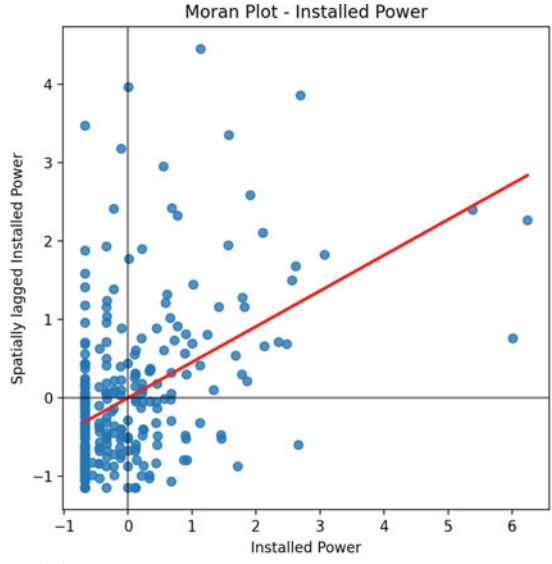
Portugal, especially in the center and south regions, but also some in the north around the city of Porto (Figs. 3, 4, 5, 6, 7, 8 and 9).

Spatial Weights Matrix In this subsection, the choice of the weights matrix is presented. In this case, the queen and rook weights matrix attribute to all locations the same neighbors. Since the grid is not regular, there is no “edge” case and so both matrices are being the same. Figure 10a shows the contiguity relationships represented by the centroids of each municipality and edges connecting them. Figure 10b shows that the minimum number of neighbors in this case is 1, while one municipality has 10 rook neighbors. The most common number of neighbors is 5.

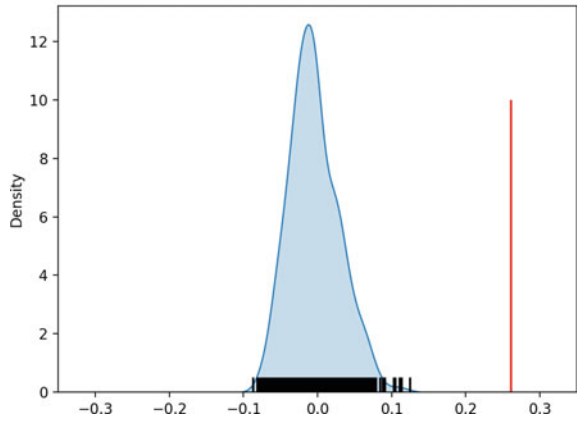
Instead of having to assume that contiguity will affect more than distance or vice-versa, a simple approach is applied by using k-nearest neighbors weights matrix and choosing $k = 5$, which is the mode of the number of neighbors.

Spatial Correlation The Moran Plot shows the relation between a variable and its spatial lag. To help with the interpretation, a linear fit, which represents the best linear

Fig. 3 Moran's I



(a) Distribution of the Installed Power against its lag



(b) Distribution of simulated Moran's I statistics for Installed Power and vertical line showing the estimated value (in red)

fit to the scatter plot is included. The slope of this line is the value of the Moran's I statistic. Figures 3a and 11 show the plots for the dependent variables. The plots display positive relationships between both variables, which is associated with the presence of positive spatial autocorrelation, meaning that similar values tend to be located close to each other.

To test whether this is statistically significant, a simulation was run with 999 permutations and the distribution of these values is shown in Figs. 3b and 12. It

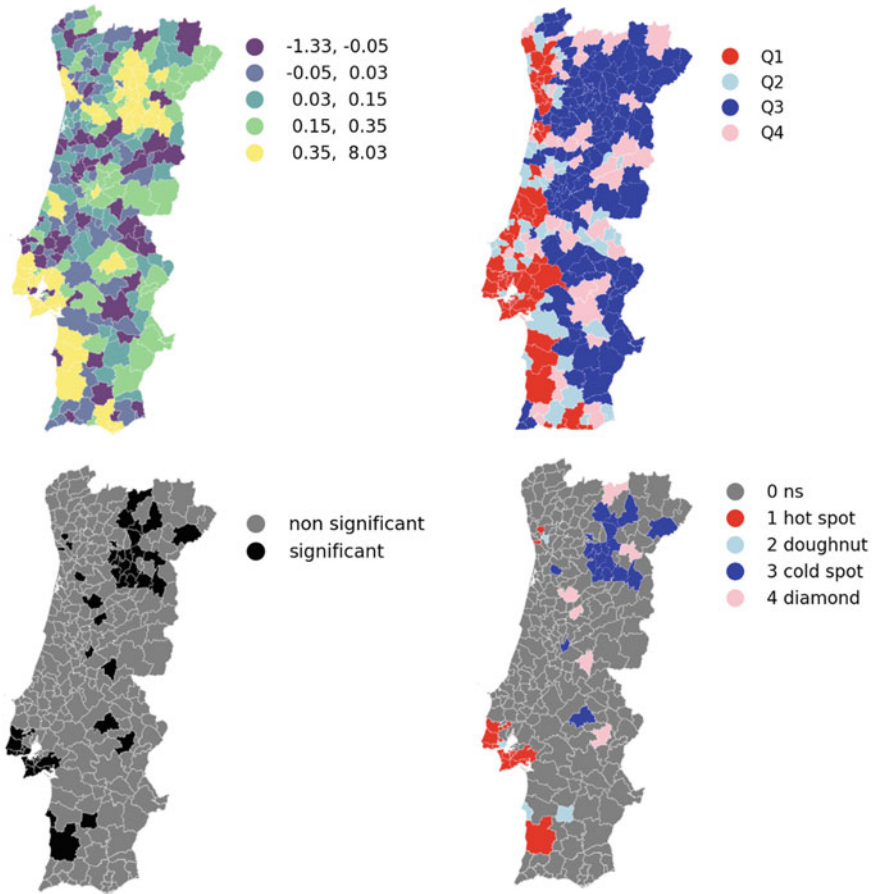


Fig. 4 LISA statistics for Installed Power across municipalities

corresponds to a kernel density estimation plot and a rug showing all of the simulated points, as well as a vertical line denoting the observed value of the statistic. It shows that it is not likely that the pattern came from a spatially random process, allowing for the conclusion that there is indeed spatial autocorrelation in the dataset.

Geary's C statistic is in line with Moran's I, as a value lower than 1 indicates that neighboring observations are similar. Geary's C simulated p-value is also 0.001. Gettis and Ord's G requires a binary spatial weights matrix with ones for all points defined as being within a certain distance of any given location, so that a different weights matrix was used to calculate this statistic. To ensure that every municipality has at least one neighbor, the minimum distance band was calculated. This needs to be at least around 31km for this data. Using this d results in the value of 0.0597 for the G statistic and the pseudo p-value of 0.001, which also suggests a clear departure from the hypothesis of no concentration. These values are summarized in Table 6.

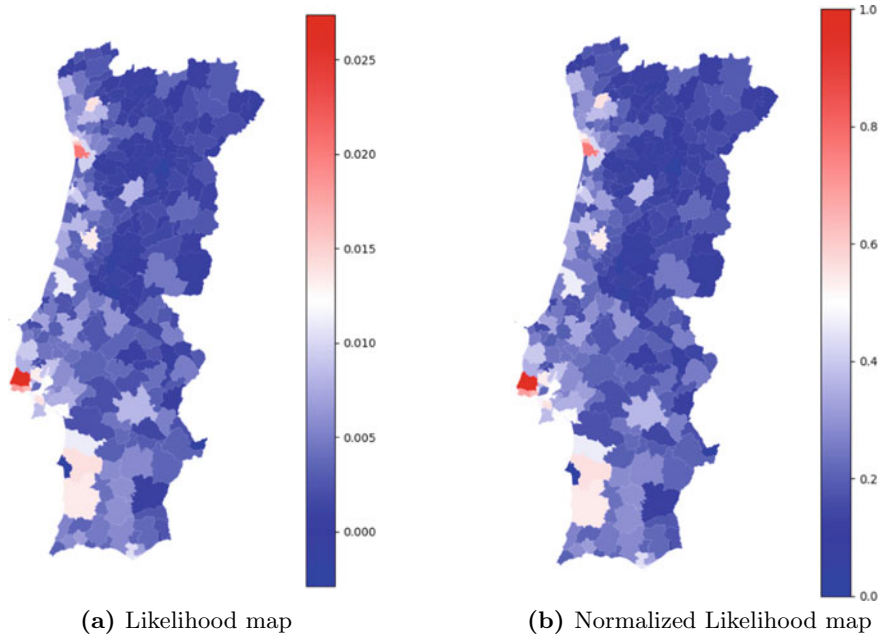


Fig. 5 Likelihood maps

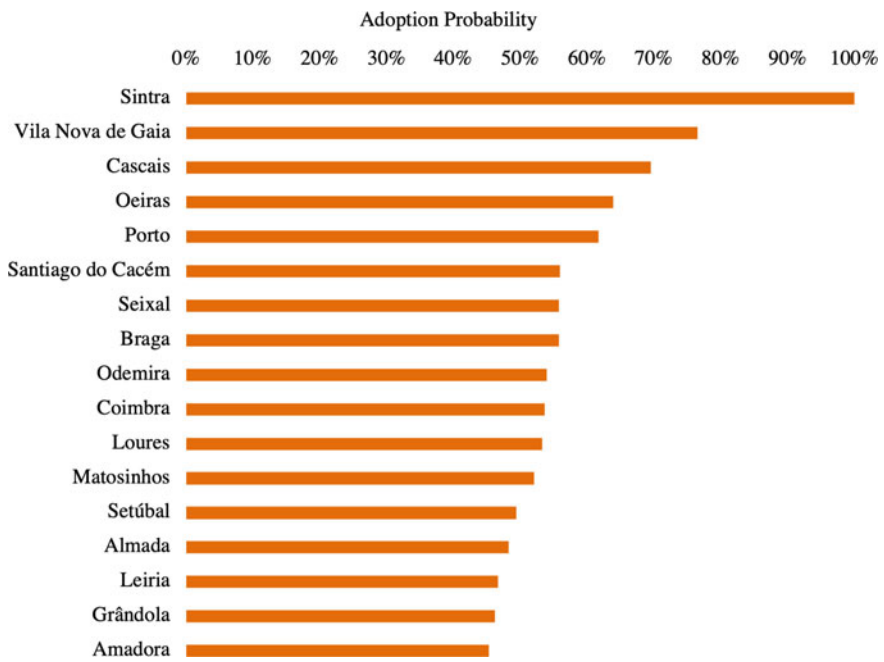


Fig. 6 Adoption probability per municipality

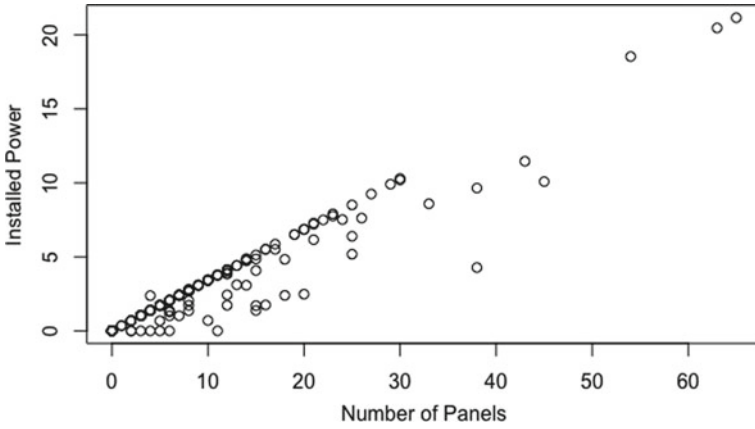


Fig. 7 Number of panels against installed power

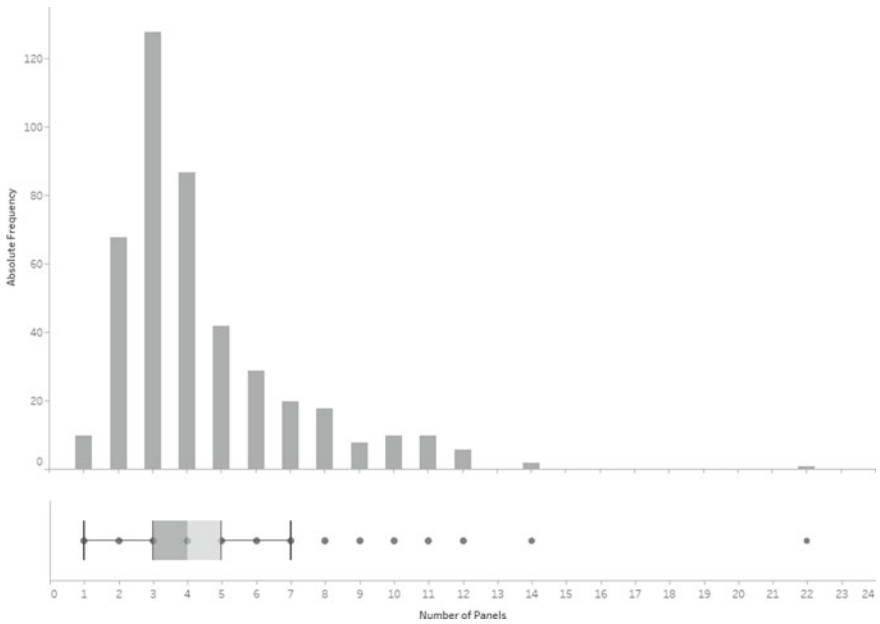


Fig. 8 Distribution of the number of panels per installation

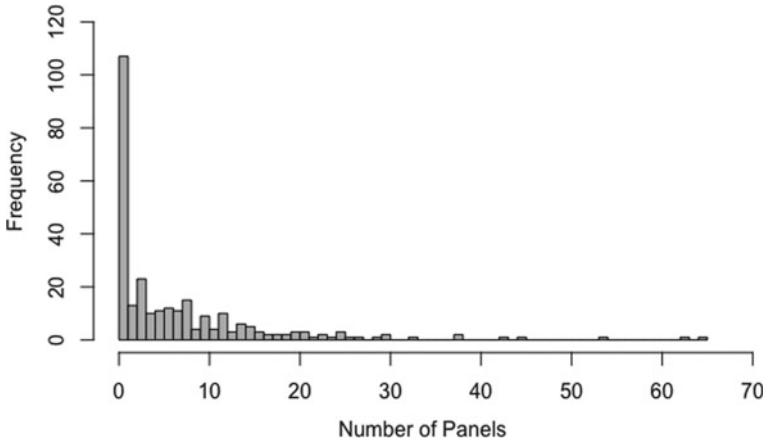
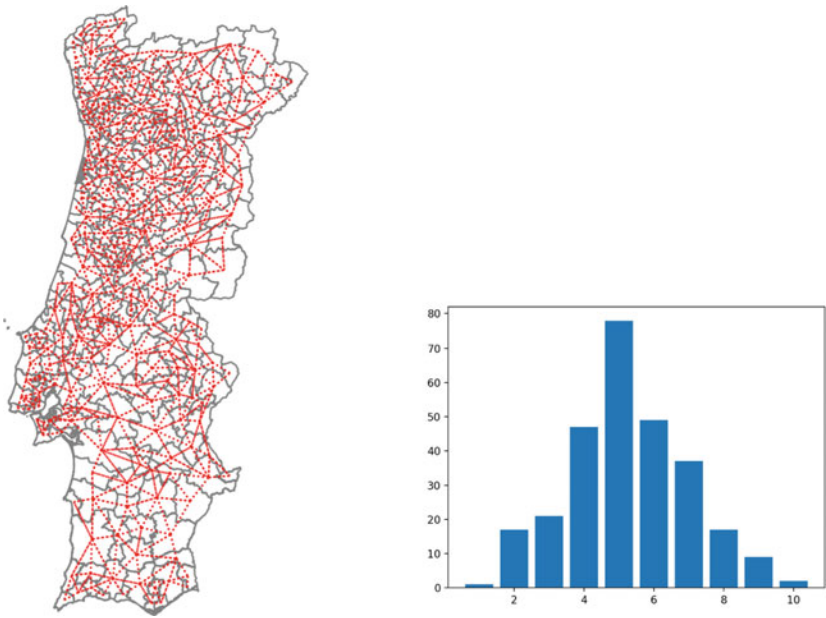


Fig. 9 Distribution of number of panels per municipality



(a) Centroids and edges used in rook and queen weights matrix

(b) Histogram number of neighbors in the rook and queen weights matrix

Fig. 10 Rook and Queen contiguity in Portuguese municipalities

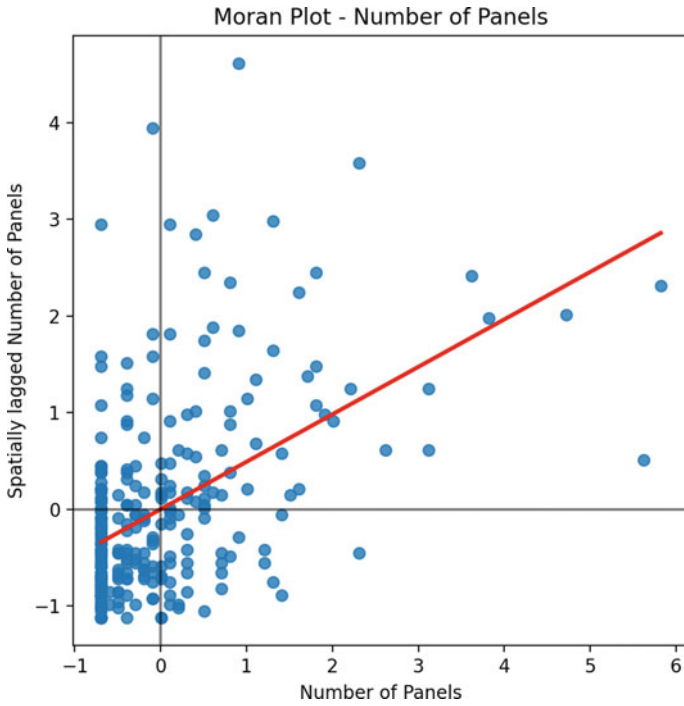


Fig. 11 Distribution of the number of panels against its lag

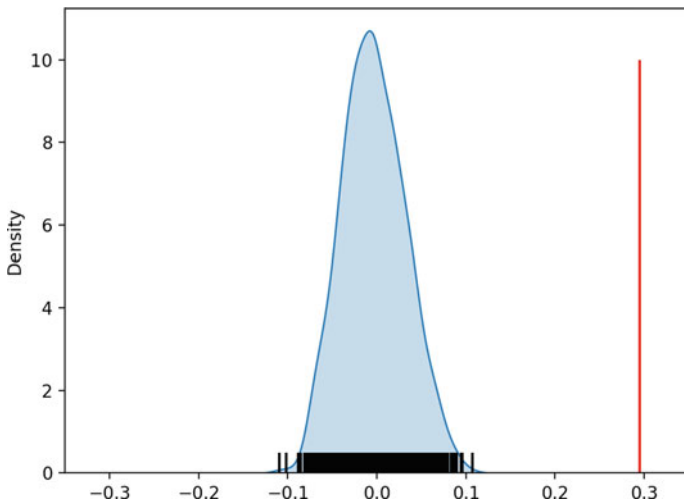


Fig. 12 Distribution of simulated Moran's I statistics for number of panels and vertical line showing the estimated value (in red)

Table 6 Results of global spatial correlation statistics for variables Number of Panels and Installed Power with respective pseudo p-values in brackets

	Number of panels	Installed power
Moran I	0.2951 (0.001)	0.2610 (0.001)
Geary's C	0.6455 (0.001)	0.6758 (0.001)
Getis and Ord's G	0.0597 (0.001)	0.0583 (0.001)

Figure 4 shows four plots that bring the different perspectives of the results for LISA for Installed Power together.

The upper-left map shows the result for local spatial autocorrelation represented by the LISA statistics. The municipalities that show high local spatial correlation in Installed Power are represented in yellow. There are some differences in the municipalities with high local spatial correlation when it comes to the Number of Panels, as can be seen in Fig. 13, namely there are less municipalities in the north of Portugal with this characteristic. The upper-right maps show the location of the LISA statistic in the quadrant of the Moran scatter plot. Comparing these two maps one can see that the positive association in the north interior part of Portugal is due to low adoption in these municipalities, while in the coastal south part of Portugal the positive association is due to the high adoption of solar PV. However, it is important to introduce the underlying statistical significance of the local values when analyzing this. Positive forms of local spatial autocorrelation are of two types: significant HH (high-high) clustering, i.e., hot spots, or LL (low-low) clustering, i.e., cold spots. Locations with significant but negative local autocorrelation are either doughnuts (low value is neighbored by locations with high values) or diamonds (high value is neighbored by locations with low values). In the last map, in bright red are the locations with an unusual concentration of high installed power surrounded also by similar locations. In light red there are the first type of spatial outliers, areas that have high installed power despite being surrounded by areas with low values. These correspond to some areas in the interior of Portugal. In darker blue one can see the spatial clusters of low power. In light blue there is another type of spatial outlier, areas with low installed power nearby areas with high.

The core idea of LISA statistics is to identify cases in which the comparison between the value of an observation and the average of its neighbors is either more similar (HH, LL) or dissimilar (HL, LH) than one would expect from chance. Figures 14 and 15 show the distribution of LISA values for the dependent variables, indicating a skewed distribution due to the dominance of the positive forms of spatial association.

The maps representing the values for the G statistics, which can be seen in Fig. 16, show similar results to LISA.

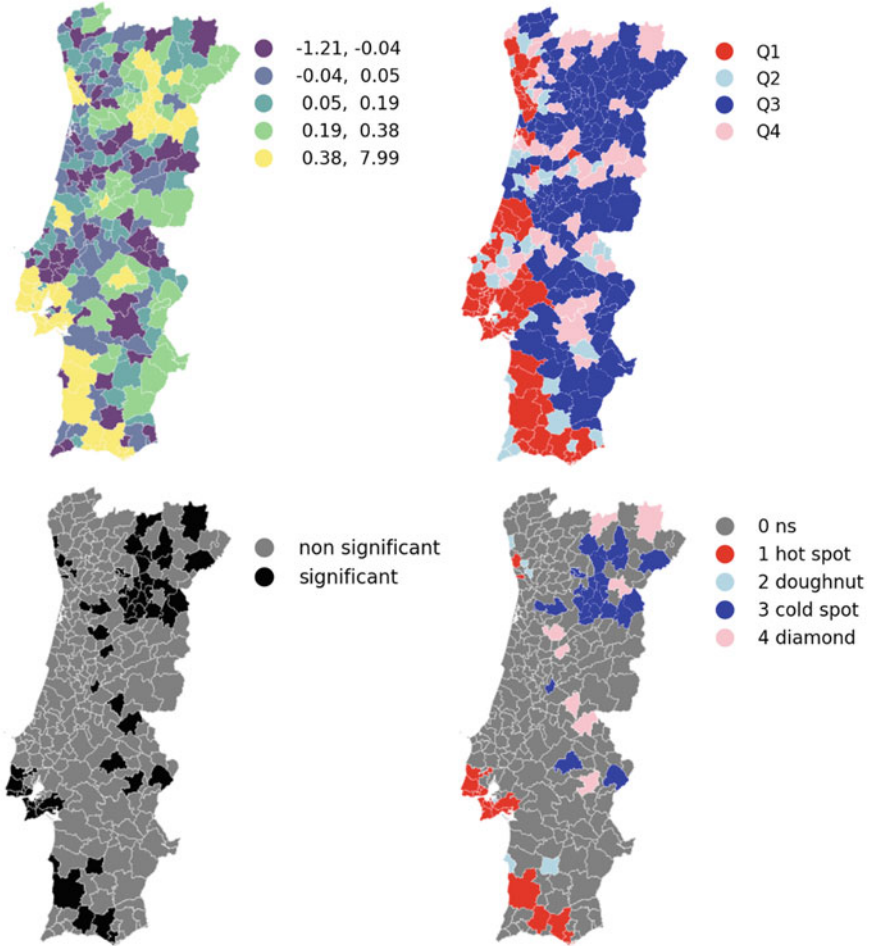


Fig. 13 LISA statistics for number of panels across municipalities

3.2 Models

Binary Target Variable In this section, the results of the non-spatial and the SAR Probit models to estimate the adoption of Installations are presented. A summary of these results is shown in Table 7.

The variables Purchase Power, Subsidies, Rental Agreements, Gross income Gini coefficient, Votes in the most voted, Family housing, and Abstention were not significant to explain whether a certain municipality adopts solar PV installations. Table 7 shows that all of the fitted models' coefficients are statistically significant. The exception lies in the SAR Probit model's spatial lag coefficient ρ , thus indicating that

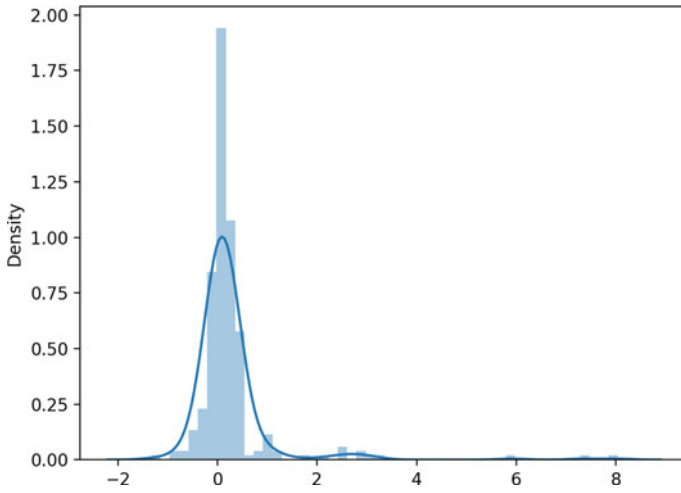


Fig. 14 Distribution of LISA values for the installed power

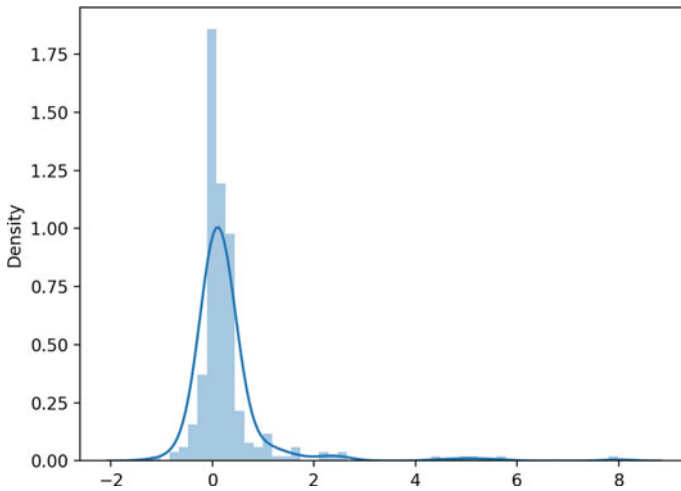


Fig. 15 Distribution of LISA values for the number of panels

the decision to adopt PV installations in one location does not seem to directly affect this decision in other locations. Regarding the log-likelihood statistics shown at the end of Table 7, they seem to show a negligible difference between the Probit and the SAR Probit model.

When considering the marginal effects presented in Table 8, one can see that while Population, Temperature, and Gross Income contribute positively to the probability of installing PV systems, Energy Consumption, Art Exhibitions, and Housing buildings seem to have a negative contribution. Regarding the SAR Probit marginal effects,

G statistic for Number of Panels

G* statistic for Number of Panels

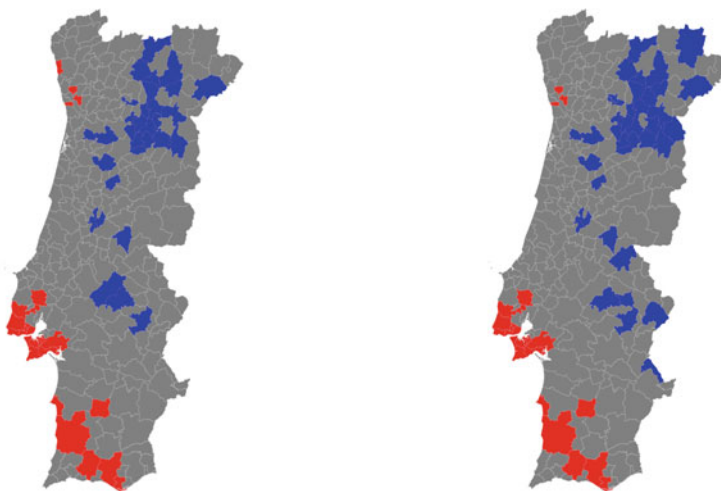


Fig. 16 Distribution of G statistic values for the number of panels

Table 7 Summary of the installation (binary) model (SE between brackets)

	Probit	SAR Probit
(Intercept)	-2.779** (1.150)	-2.835** (1.262)
Population	0.00002** (0.00001)	0.00002** (0.00001)
Energy consumption	-0.0003* (0.0002)	-0.0004** (0.0002)
Temperature	0.166** (0.082)	0.181** (0.085)
Art exhibitions	-10.674** (5.416)	-11.660** (5.686)
Gross income	0.0002** (0.0001)	0.0002** (0.0001)
Housing buildings	-0.011** (0.005)	-0.012** (0.006)
rho		-0.026 (0.132)
AIC	280.806	283.2535
BIC	306.1991	312.2745
Log likelihood	-133.403	-133.6267
Note:	* p<0.1; ** p<0.05; *** p<0.01	

Table 8 Marginal effects

	Probit	SAR Probit		
		Direct	Indirect	Total
Population	0.000005	0.000002	-0.000001	0.000002
Energy consumption	-0.000094	-0.000201	-0.000026	-0.000207
Temperature	0.045030	0.010450	-0.012240	0.009950
Art exhibitions	-2.902000	-5.627000	-0.728900	-5.725000
Gross income	0.000065	0.000024	-0.000016	0.000023
Housing buildings	-0.002939	-0.005602	-0.000640	-0.005565

it is clear that the direct effects are larger in the case of every explanatory variable, except for temperature. The direct effect of Temperature in the probability to install is positive, while the indirect effect is negative in a similar magnitude.

Discrete Target Variable In this section, the results of the OLS and SLX models to estimate the Number of Panels are presented. A summary of these results is shown in Table 9.

The variables Purchase Power and Rental agreements are not included in either of the models, together with the spatially lagged variables Population, Rental Agreements, Gross Income Gini, Housing Buildings, Votes in the most voted, and Abstention. When comparing both models, one can see that the introduction of the spatially lagged variables improves the fit, as the adjusted R^2 increases by 0.032. Both models estimate that the Number of Panels increases when Population, Gross Income, Housing buildings, and Abstention increase. Both estimate that the dependent variable decreases when Energy Consumption, Art exhibitions, Votes in the most voted, and Family housing increase. Subsidies and Temperature are not present in the SLX model, but are statistically significant in the OLS model and their lagged variant is also present in the SLX model. This means that although the temperature and subsidies received in each municipality do not seem to influence the number of solar panels acquired in the same municipality, their values contribute to explain the variance of this phenomenon in neighboring municipalities. While the lagged temperature has a positive influence on the number of panels in neighboring municipalities, lagged subsidies result in the opposite behavior, although the latter is not statistically significant. The value of the Gini coefficient of gross income is not considered relevant for the OLS, but it is statistically significant at a 90% significance level in the SLX model. Spatially lagged Purchase Power also seems to negatively influence the number of Panels acquired in neighboring locations, even though the Purchase Power of a certain location does not explain the number of panels in the same location. Energy consumption, Gross income, Family housing, and Art exhibitions, all seem to influence the total number of panels in both the locations they relate to and their neighbors, although spatially lagged Art exhibitions are not statistically significant.

Table 9 Summary of the number of panels models

	OLS	SLX
(Intercept)	-24.894*** (9.033)	-14.345*** (3.645)
Population	0.0001*** (0.00001)	0.00004*** (0.00000)
Energy consumption	-0.002*** (0.001)	-0.001*** (0.0002)
Subsidies	1.275*** (0.436)	
Temperature	1.210*** (0.402)	
Art exhibitions	-48.850** (23.676)	-15.889** (7.229)
Gross income Gini coefficient		0.161* (0.082)
Gross income	0.002*** (0.0004)	0.0003* (0.0002)
Votes in the most voted	-0.233** (0.110)	-0.060* (0.033)
Housing buildings	0.179*** (0.044)	0.062*** (0.013)
Family housing	-0.207*** (0.036)	-0.077*** (0.011)
Abstention	0.260*** (0.087)	0.060** (0.028)
L. Purchase Power		-0.132*** (0.043)
L. Energy consumption		0.003*** (0.001)
L. Subsidies		-0.369 (0.231)
L. Temperature		0.478** (0.188)
L. Art exhibitions		-23.992 (15.767)
L. Gross income		0.001** (0.0005)
L. Family housing		0.038*** (0.013)
R ²	0.4924	0.533
Adjusted R ²	0.473	0.505
Residual Std. Error	7.246	2.156
F Statistic	25.902***	18.655***
AIC	1902.82	1234.495
Log likelihood	-939.41	-599.2476
Note:	*p<0.1; **p<0.05; ***p<0.01	

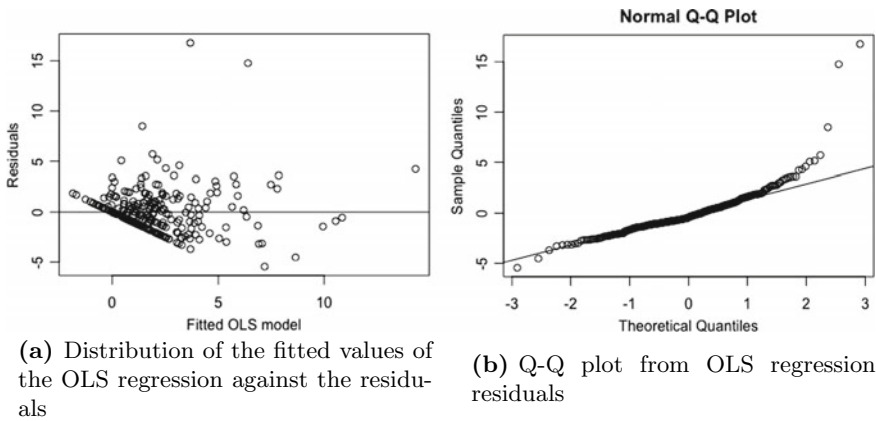


Fig. 17 Residuals of the OLS regression

Gross income in a certain location has a positive influence on the number of panels in this same location as well as its neighbors, while Energy Consumption and Family housing have opposite effects when comparing their influence on their location and its neighbors.

Continuous Target Variable In this section, the results of the OLS, SAR, SEM, GSM, SLX, SDM, SDEM, and Manski models to estimate the Installed Power are presented. Summaries of these results are shown in Tables 10 and 11.

To analyze these results, it is important to firstly analyze the residuals of the regressions and inspect the chance of heteroscedasticity as well as normality of the residuals. Analyzing Fig. 17a the residuals seem to be heteroscedastic. Looking at the Q-Q plot in Fig. 17b, the residuals tend to stray from the line near the tails, especially the right tail, which can indicate that they are not normally distributed.

Spatial autocorrelation is at least partly the suspected cause of some heteroscedasticity and non-normality found in the residuals, thus the results for spatial dependence tests in the OLS residuals were produced and can be found in Table 3. Moran's I value for global spatial autocorrelation in the residuals of the estimated model of 0.09 is statistically significant, indicating that spatial autocorrelation seems indeed to exist in the residuals of this regression. Both statistics that test for spatial error dependence (LMerr and RLMerr) are statistically significant at a 95% significance level, as well as the portmanteau test (SARMA). On the other hand, test statistics LMlag and RLMlag, which test for a missing spatially dependent variable, are not statistically significant. This seems to indicate that there is in fact spatial dependence in the residuals, but the cause is rather the spatial error dependence and not so much a spatially lagged dependent variable.

An analysis of the residuals of the other regressions, namely SAR, SEM, GSM, SLX, SDM, SDEM, and Manski, shown in the Appendix in Figs. 18, 19, 20, 21, 22, 23, and 24 reveals that their distributions remain very similar despite the different model specifications.

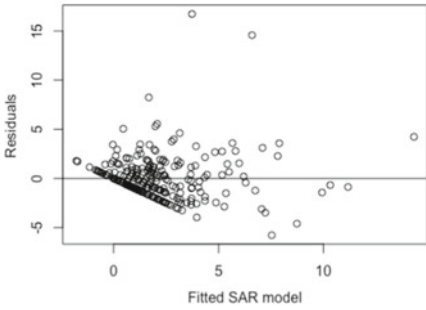
Table 10 Summary of the installed power models

	OLS	SAR	SEM	GSM
(Intercept)	-8.248*** (2.821)	-7.649*** (2.781)	-7.268** (3.190)	-7.300** (3.434)
Population	0.00003*** (0.00000)	0.00003*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)
Energy consumption	-0.001** (0.0002)	-0.001** (0.0002)	-0.001*** (0.0002)	-0.001*** (0.0002)
Temperature	0.442*** (0.120)	0.410*** (0.121)	0.432*** (0.142)	0.456*** (0.161)
Art exhibitions	-16.745** (7.430)	-16.050** (7.269)	-14.229** (7.242)	-0.135* (7.172)
Gross income	0.001*** (0.0001)	0.001*** (0.0001)	0.001*** (0.0001)	0.001*** (0.0001)
Votes in the most voted	-0.068** (0.035)	-0.068** (0.034)	-0.065* (0.034)	-0.063* (0.034)
Housing buildings	0.068*** (0.013)	0.068*** (0.013)	0.069*** (0.013)	0.068*** (0.013)
Family housing	-0.079*** (0.011)	-0.078*** (0.011)	-0.084*** (0.011)	-0.087*** (0.011)
Abstention	0.081*** (0.027)	0.080*** (0.026)	0.081*** (0.029)	0.080*** (0.030)
Rho		0.10802 (0.079751)		-0.15075 (0.14566)
Lambda			0.28883*** (0.084845)	0.40034*** (0.12999)
R ²	0.4668			
Adjusted R ²	0.449			
Residual Std. Error	2.274 (df = 268)			
F Statistic	26.072*** (df = 9; 268)			
Nagelkerke Pseudo R ²		0.470	0.480	0.483
AIC	1257.633	1258.062	1252.472	1253.096
BIC	1297.536	1301.593	1296.004	1300.255
Log likelihood	-617.8163	-617.031	-614.236	-613.548
σ^2		4.949	4.787	4.6737
Wald Test (df = 1)		1.835	11.589***	160.61***
LR Test (df = 1)		1.571	7.160***	8.5366**
Breusch Pagan	14.162	14.624	15.836*	15.685*
Note:	*p<0.1; **p<0.05; ***p<0.01			

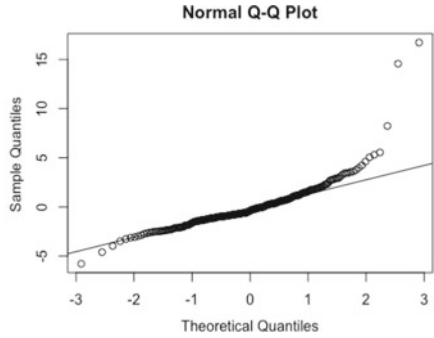
Table 11 Results Installed Power with spatially lagged explanatory variables

	SLX	SDM	SDEM	Manski
(Intercept)	-14.345*** (3.645)	-10.865*** (3.415)	-12.184*** (3.896)	-8.3895** (3.7336)
Population	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)
Energy consumption	-0.001*** (0.0002)	-0.001*** (0.0002)	-0.001*** (0.0002)	-0.001*** (0.0002)
Art exhibitions	-15.889** (7.229)	-16.293** (6.938)	-14.632** (6.894)	-12.267* (6.614)
Gross income Gini	0.161* (0.082)	0.144* (0.079)	0.166** (0.083)	0.153* (0.084)
Gross income	0.0003* (0.0002)	0.0003* (0.0002)	0.0003* (0.0002)	0.0003** (0.0002)
Votes in the most voted	-0.060* (0.033)	-0.063* (0.032)	-0.060* (0.032)	-0.054* (0.032)
Housing buildings	0.062*** (0.013)	0.061*** (0.013)	0.066*** (0.013)	0.068*** (0.013)
Family housing	-0.077*** (0.011)	-0.078*** (0.011)	-0.080*** (0.011)	-0.082*** (0.011)
Abstention	0.060** (0.028)	0.060** (0.027)	0.061** (0.029)	0.047 (0.031)
L. Purchase Power	-0.132*** (0.043)	-0.146*** (0.041)	-0.157*** (0.044)	-0.185*** (0.049)
L. Energy consumption	0.003*** (0.001)	0.002*** (0.001)	0.003*** (0.001)	0.003*** (0.001)
L. Subsidies	-0.369 (0.231)			
L. Temperature	0.478** (0.188)	0.306* (0.164)	0.295 (0.203)	
L. Rental agreements				3.251* (1.910)
L. Art exhibitions	-23.992 (15.767)			
L. Gross income	0.001** (0.0005)	0.001*** (0.0004)	0.001*** (0.0005)	0.002*** (0.0005)
L. Family housing	0.038*** (0.013)	0.032*** (0.012)	0.024* (0.013)	
Rho		0.20689** (0.085)		-0.29091* (0.1604)
Lambda			0.28517*** (0.085074)	0.55031*** (0.11073)

(continued)

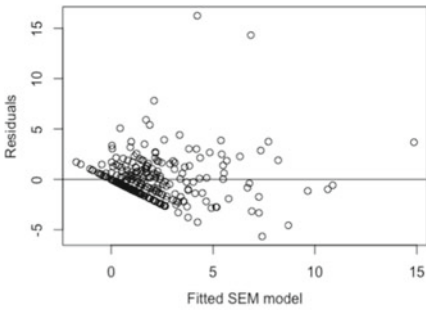


(a) Distribution of the fitted values of the SAR regression against the residuals

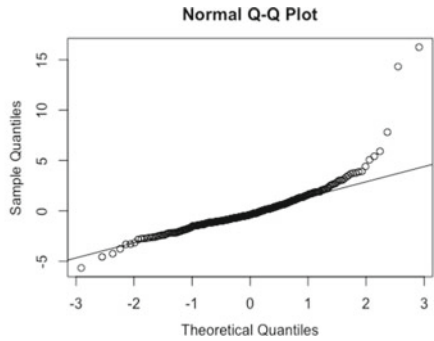


(b) Q-Q plot from SAR regression residuals

Fig. 18 Residuals of the SAR regression

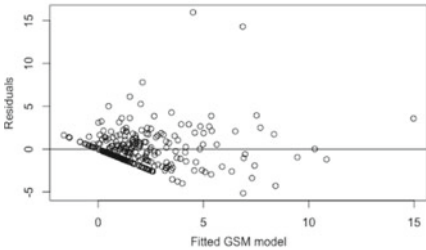


(a) Distribution of the fitted values of the SEM regression against the residuals

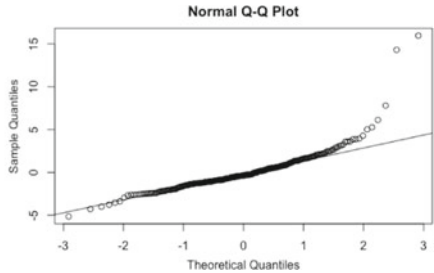


(b) Q-Q plot from SEM regression residuals

Fig. 19 Residuals of the SEM regression



(a) Distribution of the fitted values of the GSM regression against the residuals

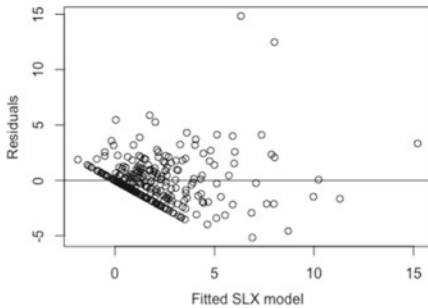


(b) Q-Q plot from GSM regression residuals

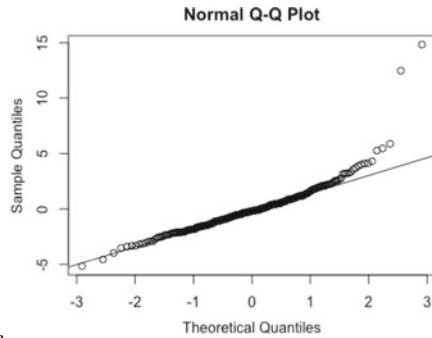
Fig. 20 Residuals of the GSM regression

Table 11 (continued)

	SLX	SDM	SDEM	Manski
R ²	0.534			
Adjusted R ²	0.505			
Residual Std. Error	2.156			
F Statistic	18.65***			
Nagelkerke Pseudo R ²		0.5317	0.5397	0.54044
AIC	1234.495	1233.565	1228.771	1228.324
BIC	1299.792	1295.235	1290.441	1289.994
Log likelihood	-599.248	-599.783	-597.386	-597.162
σ^2		4.359	4.242	3.984
Wald Test (df = 1)		4.128**	11.236***	238.26***
LR Test (df = 1)		3.244*	8.038***	173.08***
Breusch Pagan	32.139***	31.411***	32.088***	32.922***
Note:	*p<0.1; **p<0.05; ***p<0.01			



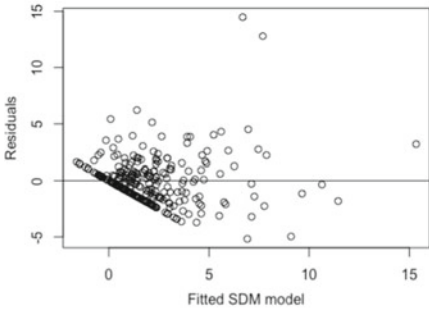
(a) Distribution of the fitted values of the SLX regression against the residuals



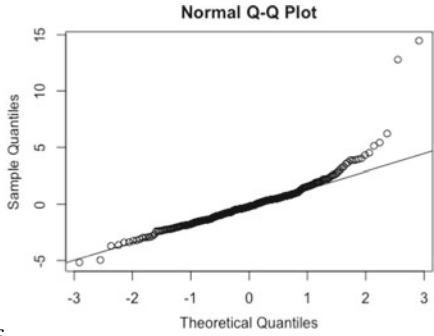
(b) Q-Q plot from SLX regression residuals

Fig. 21 Residuals of the SLX regression

All models that include a spatial term seem to produce a better fit, considering the Pseudo R-squared, than the non-spatial OLS estimation. The SAR regression produces only a slight improvement from the OLS estimation, and the ρ coefficient for the spatially lagged dependent variable is not statistically significant. Hence it seems that the Installed Power in one municipality does not affect the Installed Power in its neighbors directly. On the other hand, the λ coefficient for the spatial dependence in the error is positive and statistically significant. This indicates that similar unobserved characteristics result in similar decisions regarding Installed Power in nearby municipalities. This may be the result of a concentration of solar initiatives, local PV supplier activities, marketing campaigns or even other socioeconomic and demographic variables that were not taken into account. It is interesting to notice that

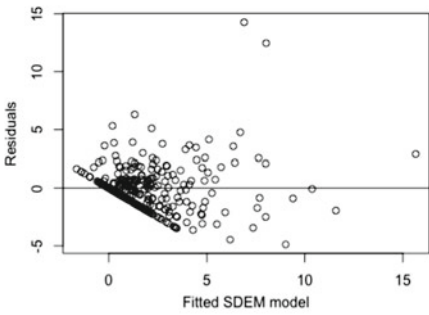


(a) Distribution of the fitted values of the SDM regression against the residuals

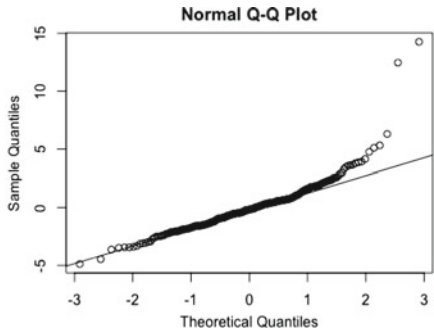


(b) Q-Q plot from SDM regression residuals

Fig. 22 Residuals of the SDM regression

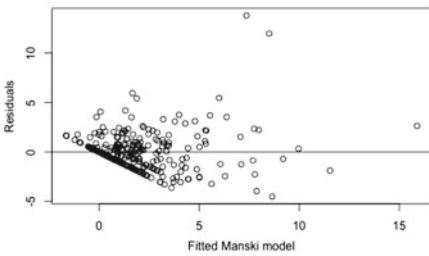


(a) Distribution of the fitted values of the SDEM regression against the residuals

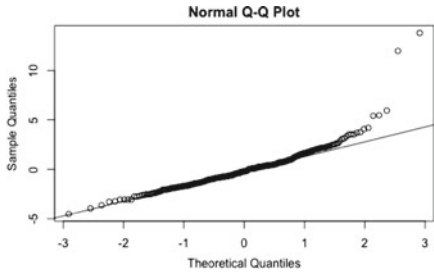


(b) Q-Q plot from SDEM regression residuals

Fig. 23 Residuals of the SDEM regression



(a) Distribution of the fitted values of the Manski regression against the residuals



(b) Q-Q plot from Manski regression residuals

Fig. 24 Residuals of the Manski regression

the λ coefficient is higher and still statistically significant at 99% significance level for the GSM model. This means that although the spatially lagged dependent variable by itself does not seem to help to model the data, it increases the influence of the spatial component in the error term. The model fit, however, measured by the pseudo R-squared increases only slightly when compared to the SEM model. These four models include the same explanatory variables and all excluded the variables Purchase Power, Temperature, Subsidies, Rental Agreements, and Gross Income Gini. All of the remaining variables are statistically significant, meaning that they contribute to explain Installed Power. Population, Temperature, Gross Income, Housing buildings, and Abstention provide a positive contribution towards Installed Power, so that when their values increase, so does the chosen Installed Power. On the other hand when Energy consumption, Art exhibitions, Votes in the most voted, or Family housing increases, the Installed Power decreases.

When adding some of the explanatory variables with a spatial lag to the OLS model (SLX model), the adjusted R-squared increases when compared to the four previous models and many of these variables are significant, showing that indeed some characteristics of neighbor municipalities seem to influence Installed Power. As was the case when lagged explanatory variables were not considered, including a spatially lagged dependent variable (SDM) improves the fit slightly. It is further improved when instead a spatially dependent error term is considered (SDEM) and even more when both spatial components are included (Manski). In the SDM, however, ρ is statistically significant at a 95%, which did not happen in SAR, meaning that when the spatial lag of explanatory variables is considered, the Installed Power in nearby municipalities seems to influence the Installed Power of an individual municipality directly. In SDEM λ also has a positive statistically significant influence on the Installed Power. Similar to the case without lagged explanatory variables, the λ coefficient increases when the spatially lagged dependent variable is added (Manski). However, the coefficient of this variable, ρ , becomes negative with a similar magnitude (0.2 and -0.3), impacting the Installed Power in the opposite way when comparing to the SDM.

The Breusch-Pagan test reveals the presence of heteroscedasticity, by rejecting the null hypothesis of homoscedasticity, in residuals of SEM, GSM, and all models that include lagged explanatory variables.

As to the β estimates, they generally do not change drastically in magnitude when comparing OLS to spatial models, what also indicates that the spatial association does not account for a great part of the model.

As mentioned in Sect. 2.2, to analyze in a more precise way the influence of each explanatory variable in models with a spatial autoregressive component (SAR, GSM, SDM, Manski), a distinction should be made between direct and indirect impacts. These can be found in the Appendix in Tables 12, 13, 14, and 15, but such a detailed analysis was considered out of the scope of this study.

Likelihood of Adoption Distribution Map The likelihood of adoption distribution map, which represents the estimated probability of PV solar installations being adopted in a certain municipality, was produced with the predicted values for Installed

Power of the SDEM regression. The estimated value for the Installed Power for each region was divided by the total estimated value. The resulting map and its normalized version can be seen in Fig. 5.

The municipalities that have a probability higher than 50% of adopting PV systems belong mainly to five clusters. Sintra being the municipality with the highest probability is also surrounded by other municipalities with high adoption probability, namely Cascais, Oeiras, Seixal, and Loures. On the north of Portugal, there is another cluster containing Vila Nova de Gaia, Porto, and Matosinhos. In the south, there is another cluster made up from Santiago do Cacém and Odemira. Furthermore, there are two municipalities that are isolated that form their own single clusters, namely Braga and Coimbra. Hence, these are the municipalities towards which selling efforts should be focused.

Table 12 Impacts SAR model

	Direct	Indirect	Total
Population	0.00003	0.000004	0.00005
Energy consumption	-0.0006	-0.0001	-0.0006
Temperature	0.4112	0.0489	0.4601
Art exhibitions	-16.0826	-1.9114	-17.9939
Gross income	0.0006	0.0001	0.0006
Votes in the most voted	-0.069	-0.008	-0.0767
Housing buildings	0.0681	0.0081	0.0762
Family housing	-0.0782	-0.0093	-0.0875
Abstention	0.0802	0.0095	0.0897

Table 13 Impacts GSM model

	Direct	Indirect	Total
Population	0.000038	-0.000005	0.000033
Energy consumption	-0.000783	0.000105	-0.000678
Temperature	0.457296	-0.061306	0.395990
Art exhibitions	-13.55226	1.816842	-11.73542
Gross income	0.000564	0.000076	0.000488
Votes in the most voted	-0.062910	-0.084338	-0.054476
Housing buildings	0.068347	-0.009163	0.059184
Family housing	-0.087023	0.011666	-0.075356
Abstention	0.080357	-0.010773	0.069584

Table 14 Impacts SDM model

	Direct	Indirect	Total
Population	0.000036	0.000007	0.000043
Energy consumption	-0.000772	-0.000514	-0.000924
Art exhibitions	-16.37473	-3.209359	-19.58409
Gross income Gini	0.1444413	0.028310	0.1727511
Gross income	0.000287	0.000056	0.000343
Votes in the most voted	-0.063155	-0.012378	-0.075532
Housing buildings	0.061716	0.012096	0.073812
Family housing	-0.078398	-0.015366	-0.093763
Abstention	0.059877	0.011736	0.071612
L. Purchase Power	-0.146840	-0.028780	-0.175620
L. Energy consumption	0.002395	0.000469	0.002865
L. Temperature	0.307829	0.060333	0.368161
L. Gross income	0.001155	0.000226	0.001381
L. Family housing	0.031831	0.006239	0.038070

Table 15 Impacts Manski model

	Direct	Indirect	Total
Population	0.000038	-0.000009	0.000029
Energy consumption	-0.000630	0.000148	-0.000482
Art exhibitions	-12.42235	2.920080	-9.502266
Gross income Gini	0.154449	-0.036306	0.118143
Gross income	0.000336	-0.000079	0.000257
Votes in the most voted	-0.054868	0.012898	-0.041971
Housing buildings	0.069294	-0.016289	0.053005
Family housing	-0.082595	0.019415	-0.063179
Abstention	0.047514	-0.011169	0.036345
L. Purchase Power	-0.187175	0.043999	-0.143176
L. Energy consumption	0.003149	-0.000740	0.002409
L. Rental agreement	3.292395	-0.773932	2.518463
L. Gross Income	0.001714	-0.000403	0.001311

4 Conclusions and Discussion

In this study, the problem of modeling the adoption of domestic solar PV systems was addressed. To do so, related data as well as socioeconomic and demographic data from each municipality was gathered. After the conclusion was reached that spatial correlation was present in the data, several models were run to try to model this behavior. Adoption was considered using three variables, namely simply whether each municipality had any installation at all, how many panels were installed and the installed power.

The purchase power and rental agreements of each municipality do not seem to add explanatory value to any of the models. Rental agreements, on the other hand, were inserted in the model to identify municipalities where many people own their house and can in fact decide on adoption of solar PV systems. For that reason, it was unanticipated. Energy consumption per capita seems to have a negative influence on the installed power, which was also not anticipated. It does, interestingly, seem to have a positive significant influence on neighboring municipalities. Temperature, as expected, has a positive significant influence on the installed power in neighboring municipalities. Municipalities that have less votes in the most voted party for government tend to have more solar power installed. One interpretation could be that these municipalities have larger environmental concerns and this is usually not represented in the most voted parties. Abstention has a positive influence, which was not expected. Intuitively one would think that more education results in less abstention and education was shown to be a positive influence on solar panel adoption.

Art exhibitions seem to be the major predictor for PV adoption, but this is most likely due to unobserved characteristics. Art exhibitions are only available on highly urban areas and these do not have high PV systems adoption rates, as apartment buildings are more common. As expected, gross income has a positive influence on adoption. A higher income naturally allows families to have space in their budget for environmentally conscious products. This positive relationship between economic status and PV installations is also reinforced by the negative influence that having a high rate of subsidies beneficiaries exerts on installed power in some model specifications. Another variable that refers to this economic factor is the Gini coefficient of gross income. Here a greater income inequality results in an increase in installed power, which is likely related to the fact that municipalities with a large total gross income result in a large Gini coefficient. Number of housing buildings has a positive influence on installed power, whereas one could have expected that an increase in housing buildings would diminish PV installations. Family housing on the other hand has a positive influence, both directly and indirectly through spatial lag, which intuitively makes sense.

The SDEM model, which considers spatially correlated explanatory variables and spatial effects in the error component is the final selected model, which means that the spatial lag is negligible. Thus, the total installed power that the population in a particular municipality in Portugal chooses to adopt does not seem to be directly dependent on the installed power of neighbor municipalities. Rather, it seems directly and indi-

rectly dependent on some observed demographic and socioeconomic variables of its neighbors, as well as unobserved characteristics (not controlled).

Considering the adoption likelihood map in Fig. 5, the focus should primarily go to Sintra. Other municipalities with high adoption likelihood can be seen in Fig. 6. When deciding on where to allocate efforts to promote solar adoption, following this order of municipalities should be optimal to accelerate Portugal's transition to renewable energies.

Acknowledgements This work is funded by national funds through the FCT—Fundação para a Ciência e a Tecnologia, I.P., under the scope of the projects UIDB/00297/2020 and UIDP/00297/2020 (Center for Mathematics and Applications).

Appendix: Complementary Figures and Tables

See Figs. 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 and 24.

Tables 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 and 15.

References

1. Schelly, C.: Residential solar electricity adoption: what motivates, and what matters? A case study of early adopters. *Energy Res. Soc. Sci.* **2**, 183–191 (2014)
2. Richter, L. -L.: Social effects in the diffusion of solar photovoltaic technology in the UK. Technical Report CWPE 1357 & EPRG 1332, Cambridge Working Papers in Economics (2013)
3. Bollinger, B., Gillingham, K.: Peer effects in the diffusion of solar photovoltaic panels. *Mark. Sci.* **31**(6), 900–912 (2012)
4. Graziano, M., Gillingham, K.: Spatial patterns of solar photovoltaic system adoption: the influence of neighbors and the built environment. *J. Econ. Geogr.* **15**(4), 815–839 (2014)
5. Schelly, C., Letzelter, J.C.: Examining the key drivers of residential solar adoption in upstate New York. *Sustainability* **12**(6), 2552 (2020)
6. Baginski, J.P., Weber, C.: Coherent estimations for residential photovoltaic uptake in Germany including spatial spillover effects. HEMF Working Paper 02/2019, Essen (2019)
7. Müller, S., Rode, J.: The adoption of photovoltaic systems in wiesbaden, Germany. *Econ. Innov. New Technol.* **22**(5), 519–535 (2013)
8. Rode, J., Weber, A.: Does localized imitation drive technology adoption? a case study on rooftop photovoltaic systems in Germany. *J. Environ. Econ. Manag.* **78**, 38–48 (2016)
9. Schaffer, A.J., Brun, S.: Beyond the sun—socioeconomic drivers of the adoption of small-scale photovoltaic installations in Germany. *Energy Res. Soc. Sci.* **10**, 220–227 (2015)
10. Tobler, W.R.: A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **46**, 234–240 (1970)
11. Anselin, L.: Local indicators of spatial association—LISA. *Geogr. Anal.* **27**(2), 93–115 (1995)
12. Elhorst, J.P.: Applied spatial econometrics: raising the bar. *Spat. Econ. Anal.* **5**(1), 9–28 (2010)
13. LeSage, J., Pace, R.K.: Limited dependent variable spatial models. In: *Introduction to Spatial Econometrics*, pp. 279–320. Chapman and Hall/CRC, Boca Raton, FL (2009)