

Hierarchical Bayesian Model for Estimating Spatial-Temporal Photovoltaic Potential in Residential Areas

Joel Villavicencio Gastelu¹, *Student Member, IEEE*, Joel David Melo Trujillo², *Member, IEEE*,
and Antonio Padilha-Feltrin³, *Senior Member, IEEE*

Abstract—This paper presents a Bayesian hierarchical model to estimate the spatial-temporal photovoltaic potential in residential areas. The proposed model offers a probabilistic approach that uses technical criteria of planners and favorable socioeconomic conditions for installing photovoltaic systems. Thus, the inhabitants' distrust of the photovoltaic solar energy choice is modeled via random distributions. The results are a spatial database that allows the creation of thematic maps to visualize the spatial distribution of photovoltaic potential in cities' residential areas for each year of the planning horizon. The proposed methodology was applied to a medium-sized city in Brazil. Maps which came from the application show the subareas with higher photovoltaic potential, where a range of impacts could appear on the distribution networks. Therefore, the results can contribute to multisenario planning and operation studies of low- and medium-voltage networks performed by utility companies.

Index Terms—Hierarchical Bayesian model, spatial-temporal analysis, photovoltaic systems, electrical distribution networks.

I. INTRODUCTION

THE increasing cost of power generation in countries dependent on fossil fuels, as well as international agreements for reducing its negative environmental impact, have motivated the search for alternative energy sources [1], [2]. One attractive alternative source is photovoltaic (PV) power generation in residential areas, as the price of PV systems is dropping, and residential consumers who choose this type of energy source are paying special tariffs, as observed in Japan [3] and Australia [4].

Most methods of estimating PV potential in urban areas have focused on the determination of the roof's area where solar panels are frequently installed [5]–[9]. Other approaches, in addition, have developed an economic analysis to determine the

domiciles with economic eligibility for installing PV systems [10], [11]. In general, the aim is to compute the PV power generation potential of a large area, such as a city, state or country, while ignoring the uncertainty of inhabitants in the decision-making process to invest in this new technology. These methods also provide a class of optimistic scenarios where all habitants, with favorable economic conditions and with available area for installation, can choose this new technology. The provided class of optimistic scenario has little information for multi-scenario planning studies of distribution systems [12].

Due to the insertion of new technologies, distribution planners are actually performing multi-scenario planning studies to take into account changes in electrical consumption patterns through the modeling of uncertainties that may occur in future plans. In this planning class, several types of scenarios should be created [12]. Therefore, the motivation of this work aims to aid the multi-scenario planning by characterizing the inhabitants' distrust of the photovoltaic solar energy choice.

In this way, the proposed work aims to model this type of uncertainty through a probabilistic approach. The estimation is done using subareas in order to locally evaluate the potential of PV power generation, identifying the subareas with higher PV potential. Thus, the potential calculated by the proposal provides support in the analysis of possible local impacts on medium- and low-voltage distribution networks. Networks with high values of PV potential have a high possibility of a reverse power flow occurring, which could result in a voltage rise in a period of low load [13], [14], as well as in malfunction of protection relays on distribution networks [15].

In the scope of probabilistic methodologies, hierarchical Bayesian models have been applied in several knowledge areas, such as epidemiology [16] and ecology [17], in order to improve the estimating process of growth rates with spatial and temporal variation within urban areas. Because of the similar behavior of growth rates and the spatial distribution and temporal variations of factors that influence the choice of inhabitants to install photovoltaic systems, hierarchical Bayesian models can also improve the estimating of PV potential in residential areas. Furthermore, Bayesian models allow the use of technical knowledge and planners experience in the estimating process of growth rate. In the estimation of photovoltaic potential, technical knowledge can help to better characterize the study zone by incorporating information of the acceptance of new technologies and energy consumption patterns of consumers.

Manuscript received March 24, 2017; revised June 20, 2017 and September 5, 2017; accepted October 22, 2017. Date of publication November 1, 2017; date of current version March 20, 2018. This work was supported in part by São Paulo Research Foundation (FAPESP) under Grants 2015/21972-6 and 2107/01909-3, in part by CNPq under Grant 307281/2016-7, and in part by CAPES. Paper no. TSTE-00277-2017. (Corresponding author: Antonio Padilha-Feltrin.)

J. Villavicencio is with Universidade Estadual Paulista—UNESP, Ilha Solteira 15385-000, Brazil (e-mail: joelitoilha@gmail.com).

J. D. Melo is with the Federal University of ABC—UFABC, Santo Andre 09210-170, Brazil (e-mail: joel.melo@ufabc.edu.br).

A. Padilha-Feltrin is with UNESP, Ilha Solteira 15385-000, Brazil, and also with UFABC, Santo Andre 09210-170, Brazil (e-mail: padilha@dee.feis.unesp.br).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSTE.2017.2768824

During the spatial estimation of PV potential, input data are necessarily aggregated per subarea. The extension of a subarea may vary depending on divisions created by the municipal urban planning department of each city or the demographic surveying agency of each country. In general, this demographic information is in the public domain and is aggregated to census tracts [18]; therefore, in this paper, it is assumed that all information is aggregated to census tracts or subareas. On the other hand, temporal estimation characterizes the evolution of factors that can influence the choice of inhabitants to install PV systems in their homes during each year of the study period.

A. Contributions

The proposal to estimate the spatial distribution of PV potential in residential areas and its evolution over time is undertaken in an innovative fashion, with a probabilistic focus through a hierarchical Bayesian model. The main contributions of the proposed method are explained below.

The proposed model permits creating PV penetration scenarios, considering factors that influence the inhabitants' choice of installing the PV systems. Such scenarios providing information about the changes in electrical consumption patterns that may occur in future plans.

The results of the proposal are a spatial database useful for creating tables and/or thematic maps with more of the major spatial resolution than previous similar methodologies. This resolution permits easily identifying residential subareas with high values of PV potential, where there is high probability of distribution networks being negatively impacted.

The characterization of the growth of the PV potential helps the distribution planners in the definition of the investments or reinforcements that must be realized in the electrical networks, facilitating efficient long term investment.

The results obtained with the proposed methodology indicate that despite the mean PV penetration in the study area is small, less than 1%, problems of loadability and voltage on low voltage networks can arise in micro areas of the study zone.

B. Work Organization

This paper is organized as follows: In Section II, a review of previous works about the PV potential estimation is performed. In Section III, the proposed methodology is presented, specifying input data and its spatial-temporal variation within the Bayesian inference process that is used to estimate PV potential. Moreover, details of this inference are explained in order to show how external disturbances might change the decision-making of inhabitants regarding whether to install PV systems. Results obtained from applying the proposed methodology to a medium-sized city are shown in Section IV, as well as a discussion comparing PV potential computed using the proposed methodology and other methodologies. Also shown is how the results of PV potential could be used to assess the impact of PV potential on the distribution networks. Finally, in Section V conclusions are presented.

II. PV POTENTIAL ESTIMATION

Several studies have been performed about the PV potential estimation of solar panels installed on roofs without taking into account the socioeconomic characteristics [5]–[9]. Factors

commonly considered in this estimation include: solar radiation, usable roof area and PV system efficiency. To estimate the usable roof area, remote sensing techniques were employed in order to obtain aerial photogrammetric data, which are processed using tools of geographic information systems (GIS). The spatial resolution in these cases generally is limited to buildings, neighborhoods or districts [5], [6]. The usable roof area was also estimated using information about the number of residential buildings and statistical data, in [7]–[9] where the spatial resolutions were states, regions or a country. Furthermore, the results of previous studies provided little [5], [6] or much generalized information [7]–[9], which is insufficient to make decisions in multi-scenario planning of distribution systems.

Because of the PV system cost, which can limit the number of domiciles for choosing this type of energy source, inhabitants' socioeconomic characteristics were taken into account, as in [10] and [11]. This consideration permitted determining domiciles with economic eligibility for installing PV systems and obtaining the PV potential in a large area. Nonetheless, a modeling of the inhabitants' uncertainty in the decision-making process of whether to invest in this new technology was not considered.

In order to better characterize the PV potential at the level of small subareas, this work presents a probabilistic model that considers the technical knowledge of the planners and the socioeconomic characteristic of the inhabitants.

III. PROPOSED METHODOLOGY

For an urban zone that has I subareas or census tracts, the proposed method estimates PV potential in each i th subarea ($i = 1, 2, \dots, I$) for each j th year ($j = 1, 2, \dots, T$) using (1).

$$\Psi_{ij} = H_{ij} \eta_{ij} S_{ij} p_{ij} \quad (1)$$

Where Ψ_{ij} is the PV potential (kW); H_{ij} , the maximum solar radiation level throughout the year (kW/m²); η_{ij} , the total conversion efficiency of PV systems; S_{ij} , the usable area on roofs (m²); and p_{ij} , the probabilities that inhabitants will install PV systems. The initial values for the first year of the study, H_{i1} , η_{i1} and S_{i1} , are proposed as input data and can be considered as constant in medium-term planning horizons, while they must be inserted as a time series for long-term planning.

The p_{ij} values characterize the uncertainties of inhabitants for installing PV systems. The Bayesian inference mechanism that employs technical knowledge and socioeconomic data is used to obtain a probability curve for the p_{ij} values. A simulation based on the Markov Chain Monte Carlo (MCMC) method [19], [20] is performed to determine the expected p_{ij} value. This p_{ij} value is the average value of the probability curve, which quantifies the possibility of installing the PV system.

The result of the proposed methodology is a spatial database that can be manipulated in GIS to obtain tables and/or thematic maps for aiding planners in the decision-making process of distribution planning studies.

A. Input Data

Similarly to methods available in the literature, the proposed method considers the following as input data: solar radiation level, total conversion efficiency of PV systems, and the usable area on roofs for installing solar panels. Moreover, the use of

socioeconomic data enhances the characterization of the population with favorable socioeconomic conditions for installing PV systems. All of the above information must be aggregated into census tracts.

The solar radiation level can be obtained using measuring instruments like radiometers, pyranometers and pyrheliometers, or through mathematical models [21].

Manufacturers using standard conditions specify the efficiency of solar panels. Performance rate should be considered due to factors that can reduce the efficiency of PV systems, e.g., dust on solar panels or the PV system being exposed to adverse environmental conditions and/or the system being installed differently from manufacturer specifications [5], [7]. Thus, the total conversion efficiency of PV systems can be calculated as the product of efficiency of the solar panels and performance rate [7].

The usable area on roofs can be calculated through remote sensing techniques and geographic information systems. However, in the absence of information to assess these techniques, the usable area can be determined as the product of the number of domiciles, their average area, and a roof utilization factor for PV systems [7], [8].

B. Calculation of Installation Probabilities of PV Systems

Installation probabilities, p_{ij} , characterize inhabitants' uncertainties for installing PV systems. This uncertainty is influenced by the inhabitants' behavior in neighboring subareas and by socioeconomic conditions for installing PV systems.

The spatial distribution of socioeconomic characteristics is generally heterogeneous because of different economic activities performed in urban areas [22], [23]. Thus, heterogeneous spatial distribution is expected in installing probabilities of PV systems. Moreover, price dropping of PV system components over time will make new families eligible to install PV systems, contributing to the increase in number of installed PV systems for each year of the planning horizon. Simultaneously, factors such as distrust of PV power generation should decrease over time due to marketing campaigns advocating its advantages.

In a Bayesian approach, the probability distribution curve of a parameter of interest is determined using two probability functions known as likelihood and prior probability. These functions are obtained from the available information and adjustment techniques for probability distribution curves, such as techniques explained in [19], [20].

The number of domiciles with favorable socioeconomic conditions for installing PV systems (candidate domiciles) in each census tract determines the likelihood function. Technical knowledge or experience of planners about the inhabitants' behavior for installing PV systems supports prior probability. Application of the Bayesian inference process yields a probability function for p_{ij} values that is known as posterior probability. An MCMC simulation is performed to determine the most representative value or average value of this curve, as in [19], [20].

In the following, the Bayesian inference process is explained step-by-step in the modeling of an urban area that has I subareas, where each i th census tract has a total of n_i domiciles in each year of the planning horizon.

In estimating the probability of inhabitants installing PV systems in the i th census tract, the p_i value is considered to be

common to every inhabitant in this subarea, since the socioeconomic characteristics of inhabitants are aggregated per census tract. This subarea has y_i candidate domiciles for installing PV systems. Therefore, the posterior probability of the installation probability, $f(\mathbf{p} | \mathbf{y})$, is obtained using (2).

$$f(\mathbf{p} | \mathbf{y}) = \frac{f(\mathbf{y} | \mathbf{p}) f(\mathbf{p})}{\int f(\mathbf{y} | \mathbf{p}) f(\mathbf{p}) d\mathbf{p}} \quad (2)$$

Where, $\mathbf{y} = y_1, y_2, \dots, y_I$ is the information about the candidate domiciles for installing PV systems, $\mathbf{p} = p_1, p_2, \dots, p_I$ are installation probabilities for each subarea, $f(\mathbf{y} | \mathbf{p})$ is the likelihood function and $f(\mathbf{p})$ is the prior probability.

If $f(y_i | p_i)$ is the probability distribution of candidate domiciles for installing PV system given the value of installation probabilities, then the likelihood function of the candidate domiciles for installing PV systems, $f(\mathbf{y} | \mathbf{p})$, is computed by (3) [17].

$$f(\mathbf{y} | \mathbf{p}) = \prod_{i=1}^I f(y_i | p_i) \quad (3)$$

In subarea i , just two types of behavior are expected by inhabitants in the n_i domiciles: or they decide to install the PV systems or they decide to do not install. Furthermore, the inhabitants' probability for installing the PV systems was denoted as p_i . In this circumstances, the number of candidate domiciles, y_i , from the set of n_i domiciles is modeled using a binomial distribution [24] that is indicated in (4).

$$f(y_i | p_i) = B(n_i, p_i) \quad (4)$$

Technical knowledge and planner experience reveal the installation costs of the PV system influence the inhabitants' decision to install this type of technology, p_i , which is a function of their socioeconomic characteristics [10], [11]. This decision also depends on a random component that is related to inhabitants' tendency towards the imitation. Hence, inhabitants in the i th subarea can be motivated for installing PV system when they observe inhabitants in neighboring subareas installing the PV systems. Taking into account the considered model in (4), the influence of socioeconomic variables and imitation tendency in the decision to install PV systems are considered in the calculation of the mean of the probability installation, as given in (5) [25].

$$Ln\left(\frac{p_i}{1-p_i}\right) = \mathbf{x}_i \boldsymbol{\theta} + \phi_i \quad (5)$$

Where $\mathbf{x}_i = (1, x_{i1}, \dots, x_{im})$ is a vector containing the socioeconomic characteristics that are considered influential in the values of y_i , m is the number of socioeconomic characteristics considered, $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_m)^T$ is a vector of weight parameters of \mathbf{x}_i values, and ϕ_i is a random component of spatial effects that takes into account the influence of neighboring subareas on the behavior of inhabitants in the i th subarea for installing PV systems. This influence has a stochastic character, which can't be modeled by the $\mathbf{x}_i \boldsymbol{\theta}$ regression component.

In Bayesian model, all parameters are represented by probability distributions [16]. Thus, the multivariate Gaussian distribution with mean λ_θ and diagonal matrix of variances Σ_θ , indicated in (6), can characterize the $\boldsymbol{\theta}$ parameters in conserva-

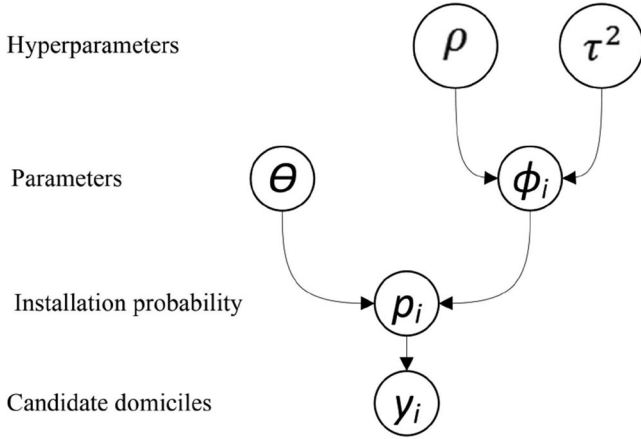


Fig. 1. Framework for hierarchical Bayesian model.

tive estimates [17].

$$f(\theta) = \mathcal{N}(\lambda_\theta, \Sigma_\theta) \quad (6)$$

Where $f(\theta)$ is the probability distribution of weight parameters of socioeconomic variables.

Since imitation tendency has a random characteristic, the component of spatial effects, ϕ_i , is modeled from a conditional autoregressive model from a multivariate Gaussian distribution that takes into account a binary neighborhood matrix W [20]. This matrix is adequately set in a way in which elements $w_{ik} = 1$ for i and k , belonging to subareas that share a border, and $w_{ik} = 0$ in other cases. The degree of influence among subareas is quantified by the ρ parameter, which is modeled using a uniform distribution. Variance of the Gaussian distribution, τ^2 , is modeled by an inverse-gamma distribution. The considered conditional autoregressive model is indicated in (7).

$$f(\phi_i | W, \tau^2, \rho) = \mathcal{N}\left(\frac{\rho \sum_{k=1}^I w_{ik} \phi_k}{\rho \sum_{k=1}^I w_{ik} + 1 - \rho}, \frac{\tau^2}{\rho \sum_{k=1}^I w_{ik} + 1 - \rho}\right) \quad (7)$$

In the theory of hierarchical Bayesian models, τ^2 and ρ parameters are known as hyperparameters in order to differentiate them from parameters of interest in the estimation process [19], [20]. Notice that the probability density functions of hyperparameters are fix, that is, they do not change during the estimation process.

The probability distribution of the spatial effects, $f(\phi_i)$ is calculated using the total probability theorem, as given in (8).

$$f(\phi_i) = f(\phi_i | W, \tau^2, \rho) f(\tau^2) f(\rho) \quad (8)$$

Where $f(\phi_i | W, \tau^2, \rho)$, $f(\tau^2)$ and $f(\rho)$ are the probability distributions of $\phi_i | W, \tau^2, \rho$, τ^2 and ρ , respectively.

From (5), the prior probability, $f(p)$, is obtained as a function of the probability distributions $f(\theta)$ and $f(\phi_i)$ given in (6) and (8), respectively.

Fig. 1 shows dependence among different parameters considered in the hierarchical Bayesian model.

After set the prior distributions and determining the likelihood function, the expected value of installation probabilities, the posterior distribution, $f(p | y)$, must be estimated. The MCMC simulation method was used in the proposed methodology. In this method is just used the product $f(y | p)f(p)$ of the posterior

distribution indicated in (2) for the construction of a Markov chain. The simulation process is divided into two stages. In the first stage, an initial number of samples ensures that the Markov chain converges to the form of the posterior distribution $f(p | y)$. This stage is known as the burn-in period of the chain. Then, in the second stage, the remaining samples are used to approximate the expected value of $f(p | y)$, using the Monte Carlo integration. Algorithms to perform this simulation as well as its convergence criteria are detailed in [27].

C. Calculation of PV Potential

PV potential in each i th subarea ($i = 1, 2, \dots, I$) is calculated for each j th year ($j = 1, 2, \dots, T$) of the planning horizon using the flowchart shown in Fig. 2. Input data represent the maximum solar radiation level (H_{i1}), total conversion efficiency of PV systems (η_{i1}), usable area for installing solar panels (S_{i1}), the total number of domiciles (n_i), the initial number of candidate domiciles for installing PV systems (y_i), socioeconomic variables (x_i), the parameters of the probability distributions $f(\theta)$, $f(\tau^2)$ and $f(\rho)$, the binary neighborhood matrix W , the total number of samples and the amount samples needed for to reach the convergence of the chain (burn-in period)

The proposed algorithm has five steps:

- 1) *Step 1:* the fixed and random components of the regression model shown in (5) are set. The fixed component is set by specifying the socioeconomic variables of the regression model and parameters of the probability distribution $f(\theta)$ indicated in (6). The random component is calculated according to (8), where the binary neighborhood matrix W and the parameters of the probability distributions $f(\tau^2)$ and $f(\rho)$ are used;
- 2) *Step 2:* the likelihood function, $f(y | p)$, of the number of candidate domiciles for installing PV systems is determined by (3) and (4), using the input data n_i and y_i ;
- 3) *Step 3:* from the steps 1 and 2, the product $f(y | p)f(p)$ of the posterior probability, $f(p | y)$, indicated in (2), is calculated.
- 4) *Step 4:* the expected value of installation probabilities, p_i , are obtained by the MCMC simulation using the product $f(y | p)f(p)$ calculated in step 3. The total number of samples and the amount samples needed for to reach the convergence of the chain (burn-in period) are considered in this step;
- 5) *Step 5:* PV potential (Ψ_i) is calculated using (1), considering the installation probabilities obtained in step 4 and the input data H_{i1} , η_{i1} .and S_{i1} .

The proposed algorithm stops when the time step reaches the number of years of the planning horizon (stop criterion). When this criterion is met, the photovoltaic potential is calculated.

The proposed algorithm to estimate the PV potential can be implemented using any statistical software such as R [28], OpenBUGS [29], or Stan [30].

IV. APPLICATION STUDY

Spatial-temporal analysis of PV potential was evaluated for a city from the southeastern region of Brazil that has approximately 200,000 inhabitants. Information on the city's socioeconomic characteristics is aggregated into 301 census tracts or subareas, originating from a national census performed in 2010 by the Geography and Statistic Brazilian Institute [18].

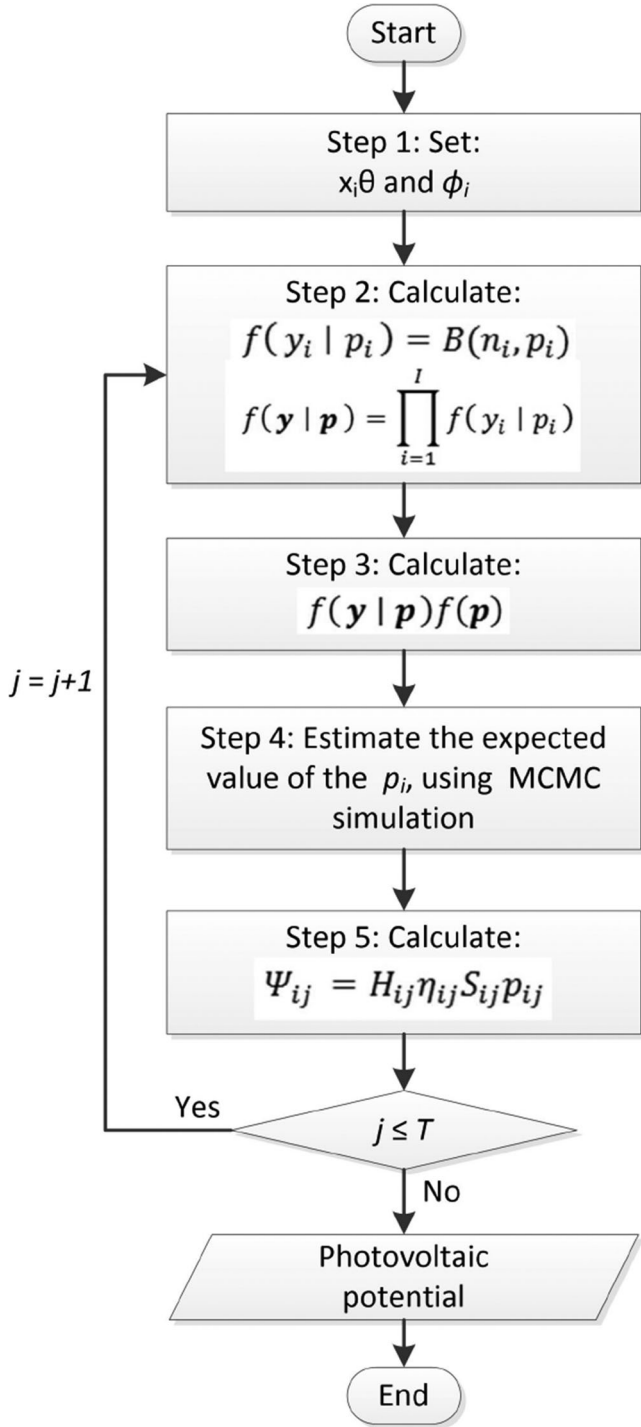


Fig. 2. Flowchart to calculate PV potential.

The chosen planning horizon is medium-term and comprises three years (2017, 2018 and 2019).

The estimation of PV potential per subarea and analysis year is obtained using the proposed algorithm. The maximum solar radiation level (H_{i1}) is 0.774 kW / m² [21]; total conversion efficiency of the PV systems (η_{i1}) is 12.3% [7]. The calculation of usable area on roofs (S_{i1}) considers only house-type domiciles and was calculated according to [7] with information provided by [31] and [32].

Calculation of installation probabilities (p_{ij}) uses the flowchart shown in Fig. 2. Household income level and home occupancy condition (owned or rented) were the chosen socioeconomic variables (x_i) because they are very influential in the decision of inhabitants regarding PV system installation. Although we have used two variables in the presented application, other variables can be added according to the planner experience or requirements of case in study, creating, thus, other classes of penetration scenarios. This information, together with data on the total number of homes per subarea (n_i), was obtained from [31].

Information on the initial number of candidate domiciles for installing PV systems (y_i) was obtained using an insertion analysis of new technology in the houses of the 301 census tracts. Three analysis scenarios (conservative, moderate and accelerated) and two tariff incentive schemes (net metering and feed-in) were considered in the above-mentioned insertion analysis, where a conservative setting under the net metering scheme represents the current scenario in the study city. Characteristics of analysis scenarios are indicated in the Appendix.

This application was implemented using the statistical software R [28] despite the proposed hierarchical Bayesian model can be implemented in any statistical software. Values used for λ_θ and Σ_θ were 0 and 1000, respectively. A value of 0.001 was used for scale and shape parameters of the inverse-gamma distribution τ_ϕ^2 , while the domain of uniform distribution ρ_ϕ was the interval [0–1]. The above values were chosen experimentally for reducing the impact of hyper parameters on estimations, [16] and [17], while offers a quickly convergence of the MCMC simulation. The number of samples and the burn-in period were experimentally set in 10000 and 2000, respectively. Computational time spent was approximately 7 seconds for each simulation.

The input data used in the application considers the experience of the incentives to renewable energy sources made in several countries. This input data results in a scenario class that replicates penetration values of PV systems in Brazil. However, if planners have another criterion or need to perform another class of analysis, then other information should be used.

A. Results

The Ψ_{ij} values obtained using the proposed method are represented in thematic maps in order to make the identification of the subareas with higher PV potential values easier. The calculated values are grouped in five spans that are represented by a color scale. The subareas with higher values are represented by the color orange, while subareas with lower values are shown in pink. Subareas in white were not considered in the estimation process, as information necessary to make the calculations was non-existent.

Under the net metering scheme, significant PV potential was not observed in most subareas for any analysis scenarios. This behavior was also observed for the conservative, moderate scenarios under the feed-in scheme, indicating that under these tariff conditions, PV system prices and incentives, hardly any households would be interested in this generation technology. However, in a scenario of greater incentives in subsidies and tariffs, like an accelerated one under the feed-in scheme, significant generation potentials are observed in some subareas, as

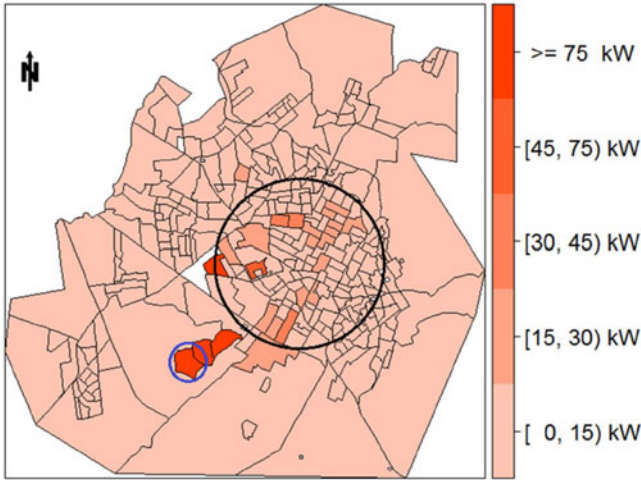


Fig. 3. Spatial distribution of PV potential (kW) under accelerated scenario and feed-in tariff. Year: 2017.

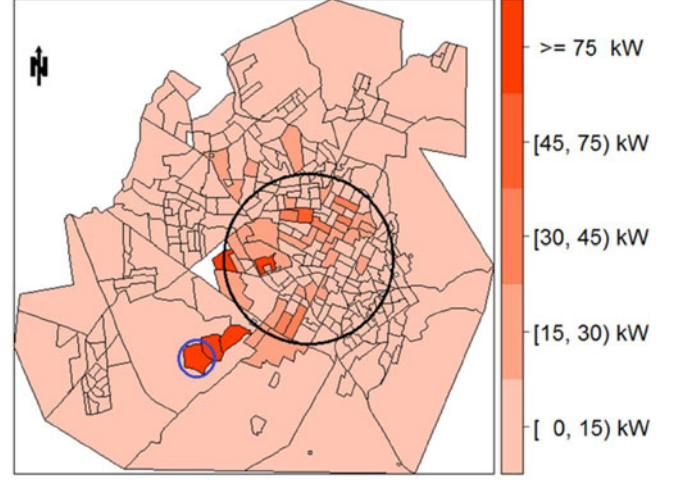


Fig. 5. Spatial distribution of PV potential (kW) under accelerated scenario and feed-in tariff. Year: 2019.

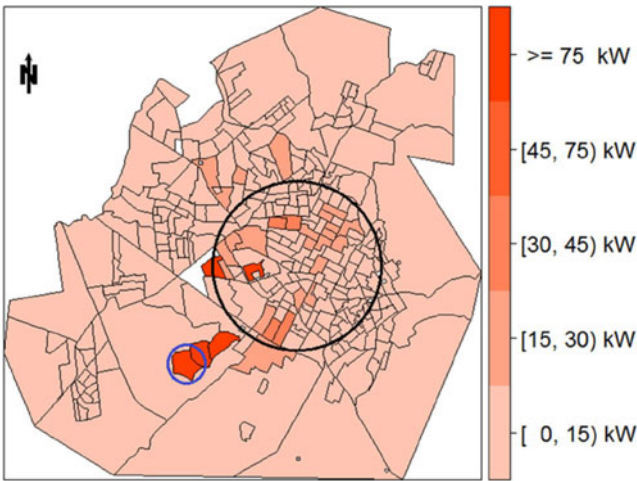


Fig. 4. Spatial distribution of PV potential (kW) under accelerated scenario and feed-in tariff. Year: 2018.

shown in Figs. 3, 4 and 5 for the years 2017, 2018 and 2019, respectively.

The observed spatial distribution of PV potential follows the expected heterogeneity, with higher values in the central and southern regions of the city in which a large quantity of high-income homes are located.

Due to the price of PV system components decreasing over time, new families will decide to install PV systems. This is observed by comparing Figs. 3, 4 and 5 in which the color of the subareas within the black color circumference become progressively more orange. However, such differences are small due both to the price of the local energy that is relatively low and the PV systems price that is relatively high in Brazil, limiting thus the feasibility of this type of project. The subarea with the highest value of PV has a potential of 144 kW for the year 2019 and is enclosed by a blue color circumference in Figs. 3–5.

For verification purposes, it was considered the estimation of the PV potential performed in the work [33] for a large area: the Southeastern region of Brazil with approximately 80 million of inhabitants. The PV penetration for the accelerated scenario

set in [33] was estimated in 0.35% for the year 2019. The same above scenario was considered for estimating the PV penetration in the study area and the proposed methodology was used. The mean penetration obtained for the study city resulted in 0.87% for the year 2019. Both works found a very small penetration, less than 1%, but the proposed methodology by small areas identifies that some subareas may require more detailed studies. The above results show the importance of a local estimative (by small areas) of the PV potential during the impact analysis on the distribution network.

B. PV Penetration Scenarios

For the purpose of confrontation information in different PV penetration scenarios, spatial distribution of PV potential in the city for the year 2019 was calculated using two preview methodologies different from the one proposed.

The first of these methodologies (MAX) does not take into account the inhabitants' socioeconomic characteristics, which limit the investment in this type of project. Thus, this methodology considers in its PV potential estimation process that the solar panels will be installed throughout the usable area on roofs [7]. The result obtained with this methodology is shown in Fig. 6. The maximum PV potential observed in the subareas was 1262 kW.

In Fig. 6, most of the subareas have higher PV potential than those obtained using the proposed Bayesian model showed in Fig. 5.

The other methodology (PRE) takes into account socioeconomic characteristics of the families, but not the uncertainty of the inhabitants for installing PV systems [10]. The result of this methodology is shown in Fig. 7. The maximum PV potential found in the subareas was 354 kW.

In Fig. 7, the subareas that appear with the highest estimated PV potentials correspond to those of Fig. 5, provided by the proposed method (BAY), but with different values of PV potential. There are also significant differences in subareas that have intermediate values.

A comparison of the results obtained by all three methods (MAX, PRE and BAY) is shown in Table I.

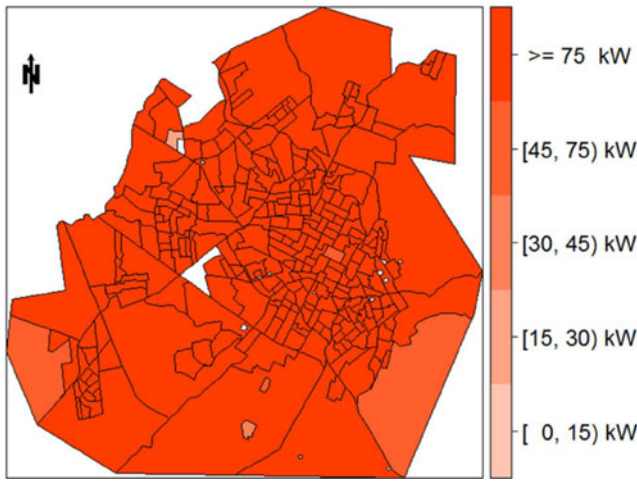


Fig. 6. PV potential (kW) according [7]. Year: 2019.

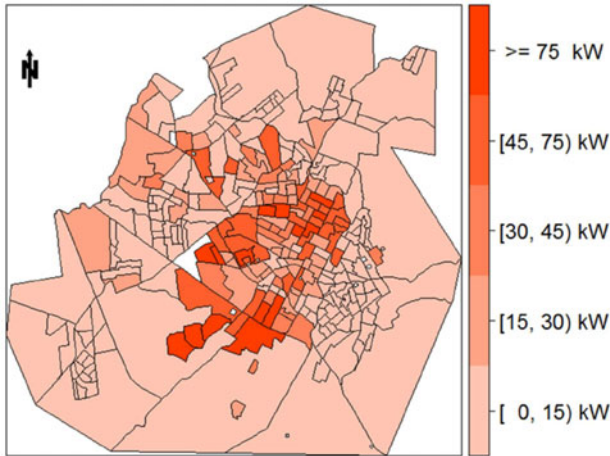


Fig. 7. PV potential (kW) according to [10]. Year: 2019.

TABLE I
RESULT COMPARISON FOR YEAR 2019

Method	Total PV (MW)	Number of subareas			
		PV > 300 kW	300 kW > PV > 150 kW	150 kW > PV > 50 kW	PV < 50 kW
MAX	150.82	241	30	12	5
PRE	6.53	3	2	29	254
BAY	2.35	0	0	5	283

Comparison of results reveals the importance of properly considering socioeconomic variables and modeling the inhabitants' uncertainties in each area of the utility concession. In Table I, it is observed that global generation potential on the roofs of residences in the entire city is 150.82 MW, but just 6.53 MW emerge with economic evaluation of interest, and this potential can be reduced to 2.35 MW after evaluating the population's installation probabilities. Thus, the proposed method provides more conservative results on the amount of energy available to be generated in the city using PV systems.



Fig. 8. Spatial distribution of the distribution transformers in the study subarea.

Currently, the development of the PV technology is in an initial phase in Brazil, with few PV systems installed throughout of the country. Therefore, values indicated in Table I show that the estimations obtained using the proposed model are more conservative than the estimations obtained by [7] or [10], where all available area can be used in the installation of photovoltaic panels or inhabitants with favorable economic conditions can choose the photovoltaic solar energy. A conservative scenario provides useful information to planners in the characterization of possible events that may occur in the future plans of distribution utilities with less impact on the elements of the electrical network than the optimistic scenarios.

C. Impact of PV Potential on Distribution Networks

Results, from a Brazilian city and obtained by the proposed method, can be used to assess the impact of the PV potential on the distribution networks. For this goal, the estimated PV potential by the proposed methodology must be associated to the distribution transformers. Then, a power flow calculation can be simulated on the distribution networks in order to observe the voltage impact. Two steps could be followed to perform the analysis of voltage impact in a low-voltage network:

- 1) *Step 1*: because each subarea can have more than one distribution transformer, the estimated PV potential must be associated with these transformers;
- 2) *Step 2*: a power flow is performed in the low-voltage network of the distribution transformers. In this power flow, the consumers' demands and the PV generation must be considered.

For exemplification purposes, the subarea enclosed by the blue color circumference on the map of PV potential shown in Fig. 5 (for the year 2019) is considered. This subarea has the highest PV potential in the studied area. It was chosen because subareas with high PV potential have more probability to present a negative voltage impact. A more detailed view of this subarea is shown in Fig. 8.

This subarea is a residential subarea and it has 19 distribution transformers (labeled from DT-01 to DT-19 and indicated in

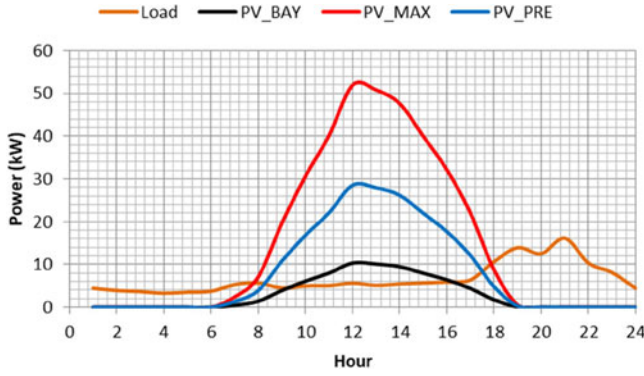


Fig. 9. Load curve on the transformer and PV generation for the year 2019.

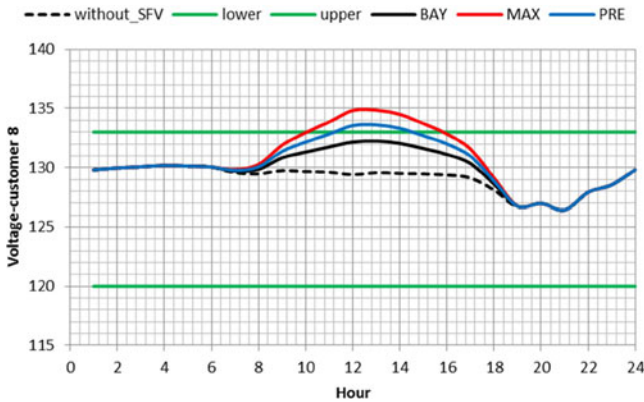


Fig. 10. Voltage profile on customer 8.

red in Fig. 8) which majorly has rated powers of 30 kVA and 45 kVA. The DT-05 transformer was chosen for analysis purposes of voltage impact. Electrical characteristics of the DT-05 transformer are: 30 kVA, 11.4/0.22 kV. The low-voltage feeder of the DT-05 transformer fed 20 consumers. The consumers are fed using the phase conductor and neutral conductor, that is, to a nominal voltage of 127 V.

To associate the estimated PV potential with the transformers in the subarea, it was considered that the PV systems have a nominal power of 4 kW. Then, the number of PV systems in the subarea was calculated by dividing the PV potential of the subarea and the nominal power of the PV systems. Finally, this quantity of PV systems was associated with each transformer using a roulette technique.

The power flow calculation was performed considering the demand curve of the consumers and the PV generation associated with the DT-05 transformer for the year 2019. The load curve on this transformer and the generated power by the PV systems according to the method BAY, MAX and PRE are shown in Fig. 9.

After simulations, it was observed that customer 8 presented greater variation on the voltage profile. This variation is shown in Fig. 10, where, the green lines indicate the operation limits, which are $\pm 5\%$ of the nominal voltage (127 V). The voltage obtained considering the PV potentials above mentioned are represented by the black, red and blue lines, respectively. The dashed line represents the voltage when the connection of the PV systems is not considered.

From the results it can be observed that the estimation methods of the PV potential indicate different situations of voltage impact on the network. Thus, optimistic estimations of the PV potential will have a greater impact on the network than a conservative estimation, like the one used in the proposed methodology, during the impact study of the network.

V. CONCLUSION

A new methodology for estimating spatial distribution of PV potential in urban areas has been presented. The methodology is based on two essential premises: uncertainty of consumers in the purchase of new technology; and the need to obtain information from small subareas of cities.

The results per subarea allow for the calculation of PV potential of the entire city. However, unlike many previous studies that demonstrate the PV potential of large areas (cities, states or countries), this work presents an evaluation by subarea with results that can be shown in tables and thematic maps. This information makes easy the identification of low- and medium-voltage networks that can be damaged (for example, on the voltage profile) by the connection of solar panels.

Subareas with increasing trends in PV potential value deserve special attention, as they may turn into subareas with high potential values in the future, thus requiring more planning and operation studies.

The spatial database of the PV potential, obtained by the proposed methodology, can expedite the work of electrical network planners in terms of decision-making related to the necessary investment for planning and operations of distribution networks.

APPENDIX

The characteristics of the analysis scenarios to determine the initial number of candidate domiciles are presented in Table II.

TABLE II
FEATURES OF ANALYSIS SCENARIOS

	Analysis scenarios		
	Conservative	Moderate	Accelerated
Subsidy	0%	20%	40%
Drop in investment	4%	5%	6%

Two tariff incentive schemes were considered — net metering and feed-in — in each analysis scenario. The tariff incentive values and the cost of residential electricity (in December 2016) are shown in Table III.

TABLE III
FEATURES OF TARIFF SCHEME AND ELECTRICITY TARIFF

	Net metering	Feed-in
Tariff incentive	0.20 US\$/kWh	0.40 US\$/kWh
Price of electricity	0.20 US\$/kWh [34]	

REFERENCES

- [1] REN21, Paris, France, “Renewables 2015 global status report,” 2015.
- [2] A. Zahedi, “Australian renewable energy progress,” *Renewable Sustain. Energy Rev.*, vol. 14, no. 8, pp. 2208–2213, Oct. 2010.

- [3] K. Ogimoto, I. Kaizuka, Y. Ueda, and T. Oozeki, "A good fit: Japan's solar power program and prospects for the new power system," *IEEE Power Energy Mag.*, vol. 11, no. 2, pp. 65–74, Mar./Apr. 2013.
- [4] B. Mountain and P. Szuster, "Solar, solar everywhere: Opportunities and challenges for Australia's rooftop PV systems," *IEEE Power Energy Mag.*, vol. 13, no. 4, pp. 53–60, Jul./Aug. 2015.
- [5] L. Bergamasco and P. Asinari, "Scalable methodology for the photovoltaic solar energy potential assessment based on available roof surface area: Application to Piedmont region (Italy)," *Solar Energy*, vol. 85, no. 5, pp. 1041–1055, May 2011.
- [6] D. Palmer, I. Cole, T. Betts, and R. Gottschalg, "Assessment of potential for photovoltaic roof installations by extraction of roof tilt from light detection and ranging data and aggregation to census geography," *IET Renewable Power Gener.*, vol. 10, no. 4, pp. 467–473, 2016.
- [7] K. Mainzer, K. Fath, R. McKenna, J. Stengel, W. Fichtner, and F. Schultmann, "A high-resolution determination of the technical potential for residential roof-mounted photovoltaic systems in Germany," *Solar Energy*, vol. 105, pp. 715–731, Jul. 2014.
- [8] N. Martín-Chivelet, "Photovoltaic potential and land-use estimation methodology," *Energy*, vol. 94, pp. 233–242, Jan. 2016.
- [9] L. Kurdgelashvili, J. Li, C.-H. Shih, and B. Attia, "Estimating technical potential for rooftop photovoltaics in California, Arizona and New Jersey," *Renewable Energy*, vol. 95, pp. 286–302, Sep. 2016.
- [10] R. Miranda, A. Szklo, and R. Schaeffer, "Technical-economic potential of PV systems on Brazilian rooftops," *Renewable Energy*, vol. 75, pp. 694–713, Mar. 2015.
- [11] K. Fath, J. Stengel, W. Sprenger, H. R. Wilson, F. Schultmann, and T. E. Kuhn, "A method for predicting the economic potential of (building-integrated) photovoltaics in urban areas based on hourly radiance simulations," *Solar Energy*, vol. 116, pp. 357–370, Jun. 2015.
- [12] H. Lee Willis, *Power Distribution Planning Reference Book*, 2nd ed. New York, NY, USA: Marcel Dekker, 2004.
- [13] R. Tonkoski, D. Turcotte, and T. H. M. EL-Fouly, "Impact of high PV penetration on voltage profiles in residential neighborhoods," *IEEE Trans. Sustain. Energy*, vol. 3, no. 3, pp. 518–527, Jul. 2012.
- [14] J. D. Watson, N. R. Watson, D. Santos-Martin, A. R. Wood, S. Lemon, and A. J.V. Miller, "Impact of solar photovoltaics on the low-voltage distribution network in New Zealand," *IET Gener., Transmiss. Distrib.*, vol. 10, no. 1, pp. 1–9, 2016.
- [15] P. Mohammadi, and S. Mehraeen, "Challenges of PV integration in low-voltage secondary networks," *IEEE Trans. Power Del.*, vol. 32, no. 1, pp. 525–535, Feb. 2017.
- [16] A. B. Lawson, *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*, 2nd ed. Boca Raton, FL, USA: Chapman & Hall, Mar. 2013.
- [17] F. Korner-Nievergelt, T. Roth, S. von Felten, J. Guélat, B. Almasi, and P. Korner-Nievergelt, *Bayesian Data Analysis in Ecology Using Linear Models With R, BUGS, and Stan*, 1st ed. San Diego, CA, USA: Elsevier, Apr. 2015.
- [18] Instituto Brasileiro de Geografia e Estatística – IBGE, "Base de informações do censo demográfico." 2010. [Online]. Available: http://www.ipea.gov.br/redeipea/images/pdfs/base_de_informacoes_por_setor_censitario_universo_censo_2010.pdf (in Portuguese)
- [19] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. Boca Raton, FL, USA: Chapman & Hall, Nov. 2013.
- [20] S. Banerjee, B. P. Carlin, and A. E. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. Boca Raton, FL, USA: Chapman & Hall, 2015.
- [21] S. C. Bhatia, *Advanced Renewable Energy Systems, Part – I*, 1st ed. Philadelphia, PA, USA: Woodhead, 2014.
- [22] M. Batty, *Cities and Complexity: Understanding Cities With Cellular Automata, Agent-Based Models and Fractals*, 1st ed. Cambridge, MA, USA: MIT Press, 2007.
- [23] D. Lee and R. Mitchell, "Boundary detection in disease mapping studies," *Biostatistics*, vol. 13, no. 3, pp. 415–426, 2012.
- [24] D. Collett, *Modelling Binary Data*, 2nd ed. Boca Raton, FL, USA: Chapman & Hall, Sep. 2002.
- [25] J. Jiang, *Linear and Generalized Linear Mixed Models and Their Applications*, 2nd ed. New York, NY, USA: Springer, May 2007.
- [26] D. Lee, "An R package for bayesian spatial modeling with conditional autoregressive priors," *J. Statist. Softw.*, vol. 55, no. 13, pp. 1–24, Nov. 2013.
- [27] C. Robert and G. Casella, *Introducing Monte Carlo methods With R*, 1st ed. New York, NY, USA: Springer, 2010.
- [28] R Core Team, "R: A language and environment for statistical computing," *R Found. Statist. Comput.* 2017. [Online]. Available: <http://www.R-project.org>
- [29] OpenBUGS Foundation, "OpenBUGS. The BUGS project," 2005. [Online]. Available: <http://www.openbugs.net>
- [30] Stan Development Team, "Stan modeling language users guide and reference manual," version 2.17.0. 2017. [Online]. Available: <http://mc-stan.org>
- [31] Instituto Brasileiro de Geografia e Estatística - IBGE, "Censo demográfico 2010: agregado por setores censitários, Rio de Janeiro." 2011. [Online]. Available: <http://downloads.ibge.gov.br/index.htm>. (in Portuguese).
- [32] E. Ghisi, A. Montibeller, and R. W. Schmidt, "Potential for potable water savings by using rainwater: An analysis over 62 cities in southern Brazil," *Building Environ.*, vol. 41, no. 2, pp. 204–210, Feb. 2006.
- [33] N. Kagan *et al.*, *Redes Elétricas Inteligentes no Brasil: Análise de Custos e Benefícios de um Plano Nacional de Implantação*, 1st ed. Rio de Janeiro, Brazil: Synergia, 2013. (in Portuguese).
- [34] Agência Nacional de Energia Elétrica – ANEEL, "Informações técnicas - relatórios do sistema de apoio a decisão: consumidores, consumidores, consumo, receita e tarifa média – região, empresa e classe de consumo." 2016. [Online]. Available: http://relatorios.aneel.gov.br/_layouts/xlview.aspx?id=/RelatoriosSAS/RelSAMPRegiaoEmp.xlsx&Source=http%3A%2F%2Frelatorios%2Eaneel%2Egov%2Ebr%2FRelatoriosSAS%2FForms%2FAIItems%2Easpx&DefaultItemOpen=1. (in Portuguese).

Joel Villavicencio Gastelu received the B.Sc. degree in electrical engineering from UNI, Lima, Peru, in 2012. He received the M.Sc. degree in electrical engineering in 2016 from UNESP, Ilha Solteira, Brazil, where he is currently working toward the Ph.D. degree in electrical engineering. His research interests include renewable energy and optimization of power systems.

Joel David Melo Trujillo (S'10–M'14) received the B.S. degree in electrical engineering from UNMSM, Lima, Peru, in 2006, and the M.Sc. and Ph.D. degrees in electrical engineering from UNESP, Ilha Solteira, Brazil, in 2010 and 2014, respectively. He is currently an Associated Professor at the Federal University of ABC—UFABC. His main interests include power network planning and spatial analysis.

Antonio Padilha-Feltrin (SM'06) received the B.Sc. degree in electrical engineering from UNIEFEI, Itajubá, Brazil, and the M.Sc. and Ph.D. degrees in electrical engineering from UNICAMP, São Paulo, Brazil. From 1995 to 1997, he was Visiting Faculty at the University of Wisconsin, Madison, WI, USA, and from 2012 to 2013 at Imperial College London, U.K. He is currently a Professor at UNESP, Ilha Solteira, and a Senior Visiting Professor at UFABC, Brazil. His main interests include the analysis and control of distribution and transmission systems. He is an Editor for the IEEE TRANSACTION ON SUSTAINABLE ENERGY.