A Data-Driven Framework for Optimal Placement of Grid-Connected Solar Generation

Ana M. Ospina, *Student Member*
Departament of Electrical and Computer Engineering
Texas A&M University, College Station, TX
ana.ospina@tamu.edu

Mladen Kezunovic, *Life Fellow*
Departament of Electrical and Computer Engineering
Texas A&M University, College Station, TX
kezunov@tamu.edu

***Abstract*— This work presents a decision making approach for selecting an optimal placement of the grid-connected solar generation using Geographical Information System (GIS) as the decision making tool. A terrain analysis for solar radiation assessment, as well as buildings and vegetation spatial data are analyzed in order to determine the shadow impact that can be anticipated for medium or large-scale PV installation. In addition, different historical weather conditions are considered and integrated into the model to show the impact of this variable on the solar generation output. Some details of the methodology, testbed development and results related to the selection of potential sites for PV installation are presented. To illustrate the process and proposed methodology, an example using large scale synthetic networks is implemented.**

Keywords—PV Generation, Power System, Geographical Information System, Decision Making, Optimal Placement.

I. Introduction

The high penetration of PV generation connected into the grid needs new operational and planning methodologies that deal with the variability and unpredictability of the solar resource [1]. The large amount of grid-connected solar generation rises significant concern about how to ensure the stability and preserve the resiliency of the power system [2]. Currently, the optimal placement for renewable energy has been explored from different perspectives. One of them is a multi-objective optimization using techniques such as: technical and economic impacts models [3]; simulated annealing algorithm [4], [5]; stochastic integer programming problem [6]; genetic algorithm [7], [8]; Grey Wolf Optimizer (GWO) algorithm [9]; genetic algorithm (GA) based on Newton–Raphson power flow [10], hybrid techniques [11] and so on. Furthermore, methods like Monte Carlo [12], and robust optimization-based PV placement strategy [13], have been used with satisfactory results. More advanced techniques, such as evolutionary programing and integer-encoding evolutionary algorithms, have been also introduce in order to tackle the issue of placement of the different type of distributed generation into the grid [14], [15].

Having more detailed models that deal with an extensive amount of data, constrains and uncertainties has helped the development of interdisciplinary tools that integrate detailed system information such as grid topology, resources measurement, digital elevation models, combined in just one methodological approach. In addition, applying the current operational and planning methodologies for generation resources, which are highly dependent on the physical and electrical constraints, is difficult and highly time-consuming. The transition to a novelty and clean energy economy requires a proper integration between high-quality of renewable energy data and geographic information system (GIS), as is stated in [16]. These data sets are crucial for making informed decisions – ranging from policy and investment risks to reliable power sector planning [16].

In order to solve the optimal placement problem, other approaches have been studying Geographic Information System (GIS) framework. Such studies have been focused mainly on the use of spatial analysis techniques. In [17] the authors used GIS spatial analysis techniques jointly with wavelet theory and with a different types of collected data, such as: field monitoring data, Digital Elevation Models (DEM), electrical status of the power system, classified land use data, population, and other data sources, in order to define suitable places for the development of solar generation. Other example of the use of GIS is presented in [18], where the authors introduced an on-site algorithm for solar power generation, testing spatial-temporal data method and modeling to develop a solar generation placement in India using a web-base application. The technique called Fuzzy Analytic Hierarchy Process (AHP) has been adopted in few studies and authors showed its efficiency in finding the best suitable site for installation of solar power plant in practical cases in India [19] and Saudi Arabia [20]. Finally, references [21] and [22], introduce different applications of how to use GIS for the analysis of owner distribution system.

A data-driven framework allows the operator or designer of the system to: (i) import data from outside sources, (ii) easily have access and modify the data, and (iii) query the database in real time or historical. It is crucial to take into the account that data can vary in type and quality, be expensive to obtain, and require specific skills and resources to process and interpret [16]. With the purpose of performing a detailed analysis that leads to the decision about the optimal placement, the realistic models of the grid and PV integration options allowing consideration of maximum number of constraint that are crucial for planning and operational, are necessary. In this work, we propose a methodology to create an input layer using GIS that assists the decision-making process for the optimal location of PV generation based on solar radiation, vegetation location, placement of supporting structures, historical weather patterns, and grid operating regimes. Our approach presents a systematic, practical and easily implementable methodology using public data that has not been utilized in previous approaches. Due to the high correlation of different types of data, this approach is

This publication was made possible by BD Spokes: SPOKE SOUTH: Collaborative: Smart Grids Big Data: NSF1636772. The statements made herein are solely the responsibility of the authors.

very versatile and provides information about optimal location of grid-connected solar generation.

This paper is organized as follows. In Section II, the methodology components are presented. The GIS framework as a decision-making tool is introduced in Section III. Section IV provides the spatial correlation of the data. The methodology evaluation and the results are presented in the Section V and VI, respectively. Finally, the conclusion and future work is shown in Section VII.

II. Modeling Components

In the optimal placement of grid-connected solar generation it is necessary to take into account the continuity of the evolving conditions that affect this type of variable resources. In this context, this paper presents a data-driven framework as a tool for dealing with these unpredictable and variable conditions. An overview of the data sources are shown in Table 1. A range of historical data in Table 1 is collected by a variety of technologies, such as: 1) satellite-based surface radiation, 2) satellite algorithm for shortwave radiation budget, 3) mapped type techniques using land cover and abiotic variables, and 4) geolocation using ZIP code.

Table 1: Data sources and characteristics.

Source	Data Type	Data description
Radiation (Multi year PSM Global Normal Irradiance) [29]	Raster	Global Normal Irradiance (GNI): The solar radiation values represent the resource available to solar energy systems. Surface cells are 4 km in size approximately.
Brazos County, Bryan and College Station City data [30]	Vector	Shapefiles of county and cities boundary, building zones, parks, future developments, commercial areas, among other. Polygons represent different features of data.
Texas ecological Mapping system data [31]	Raster	Vegetation data is Post Oak Savanah specifically for Brazos County. Polygons for each type of tree (Veg_ID) area with shape area and shape length. The Veg_ID can be join with the height of the trees.
Power system data [32]	Vector	Point with the substation for Brazos county using synthetic power grids data.

When we are dealing with big data in solving power systems problems, the data management process is crucial. In this application we integrate the *data ingestion*, *data cleansing*, and *data curation* as it is shown in Fig. 1, following the definitions presented in [27] and [28]. In this case, we are dealing with different types of data that require spatial and temporal analysis.

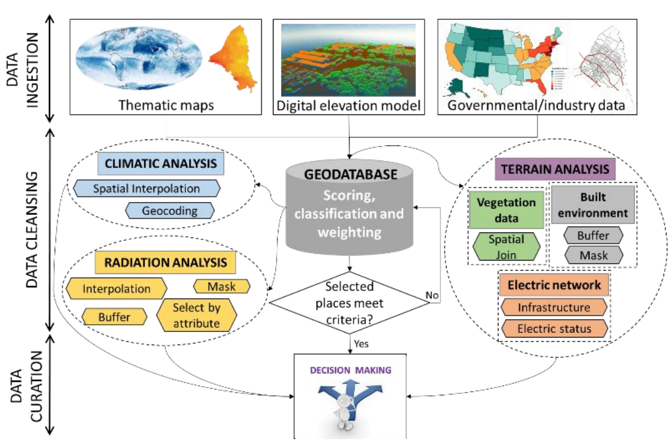

Fig. 1: The integration of the modeling components to create an optimal placement of grid-connected solar generation.

III. Geographical Information System as a Decision Making Support Tool

A. Terrain Analysis

The terrain analysis using GIS is an accurate and adequate approach to determining the potential advantages or disadvantages of integration of solar generation into distribution and transmission systems [22]. Using different public sources of information, it is possible to determine the optimal location for PV generation under different conditions as illustrated in Fig. 2.

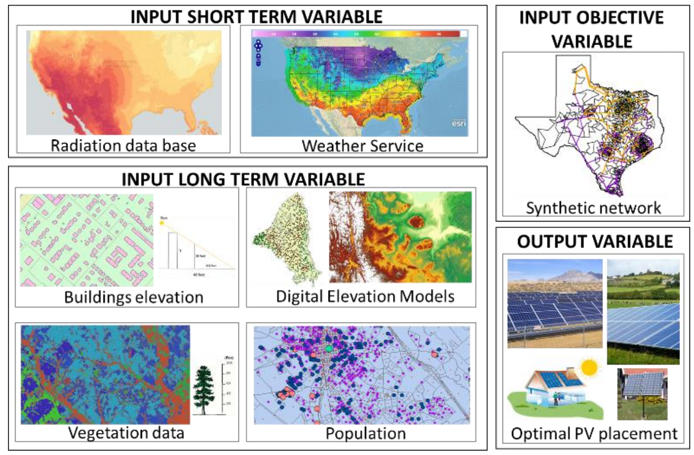

Fig. 2: Terrain analysis using Geographic Information System.

B. Decision Making Process

Multi-criteria optimization analysis is frequently used in the decision-making process for power system planning and operation. In this case, using the results of terrain analysis, it is possible to evaluate the impacts that PV generation can have on the current distribution power system [21]. Fig. 3 shows the decision making process applied to a synthetic network [32].

The use of high quality data from renewable energy sources and other geospatial data can improve the impact of integration of high levels of variable resources. Geospatial data and analysis are crucial to supporting the integrated assessment that brings together renewable energy resources, geographic, economic, and other considerations for data-driven renewable energy target setting [16].

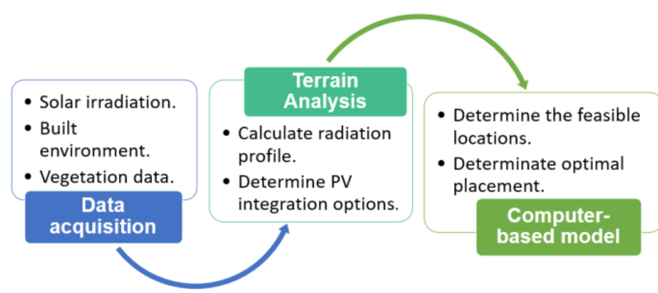

Fig. 3: Decision-Making process for optimal location of PV generation.

IV. Spatial Correlation of Data

Fig. 4 presents the spatial correlation of the data sets for the optimal placement of grid-connected solar generation. The location of the power system substation is stored in the geodatabase. This geodatabase is first loaded with the historical radiation data that are geocoded into a raster file created from a spatial interpolation (Inverse Distance Weighting – IWD process). Vegetation data are obtained from Texas parks & Wildlife website [31], and processed by applying spatial joins and selections to create shapefiles. Those shapefiles are then added to the existing database. For each vegetation section that meet the pre-specified requirements, the feasible zones are evaluated form the historical radiation data in a radius that can vary. Radius primary depends on the size of grid connected solar generation that it is being integrated to the grid. Weather data is associated with the location and overlaid in the area of interest. After this procedure, the parameters of the weather data need to be spatially interpolated, with the objective of performing the estimation of the values at each location in the area under study.

The final output of the spatial correlation is a set of data reflecting radiation, weather parameters, built environment properties, and vegetation with all the related features integrated in the outage shapefile. This file contains the locations of the feasible regions to integrated medium or large solar farms into the grid, and all in an optimal approach.

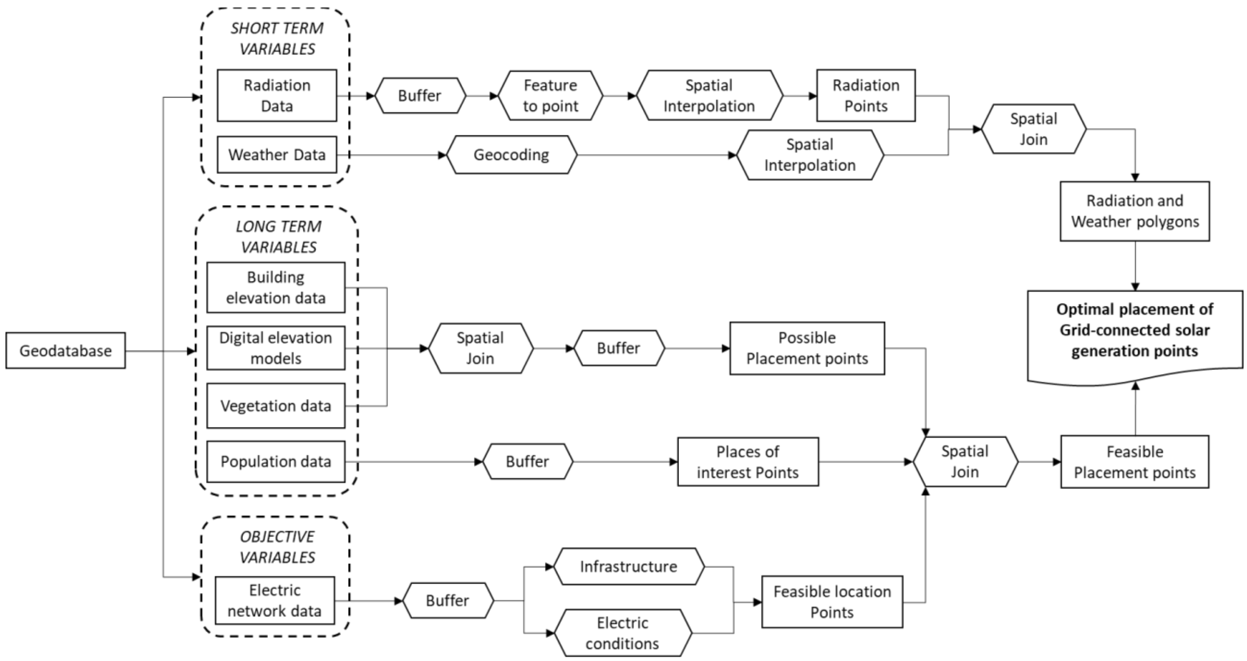

Fig. 4: Spatial correlation of data for optimal placement of the grid-connected solar generation.

V. Methodology Evaluation

In order to develop the methodology, the probabilistic behavior of power system is taken into account as a high risk factor. This situation is getting even worst when the increased amount of variable renewable energy sources, such as solar energy are being integrated into the grid [23] and [24].

The proposed large-scale testbed architecture is presented in Fig. 5 and Fig. 6, taking into account the work presented in [25] and [26]. This testbed presents an example of how the majority of the electric grid data can be spatio-temporarily correlated with interdisciplinary sources of data, such as weather, radiation, vegetation, built environment, and other decision-making tools. Fig. 5 present the general idea behind the testbed, and the representative architecture for the weather testbed is shown in Fig. 6. In this specific application of grid-connected solar generation, we are using an optimization process, aim to optimally locate medium or large scale solar farms. The heuristic optimization method implement in GIS follows the flow chart present in Fig. 4. The objective function is find the maximum amount of optimal placement that meet the customized criteria. The testbed is implemented using some commercial solutions. First, the public information is loaded to the ArcGIS platform (see Table 1). The spatial-correlation methodology as depicted in Fig. 4, is implemented in ArcGIS and it allow us to integrate and spatio-temporarily correlate the standard types of data related to radiation, vegetation, build environment, and weather.

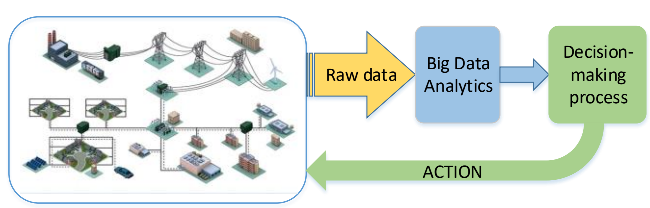

Fig. 5: The large scale testbed for study of weather impacts on the electricity grid.

This solar generation placement application shows how it is possible, using public data, to perform an optimal placement analysis of grid-connected solar generation into the grid as a first step prior to the decision making process.

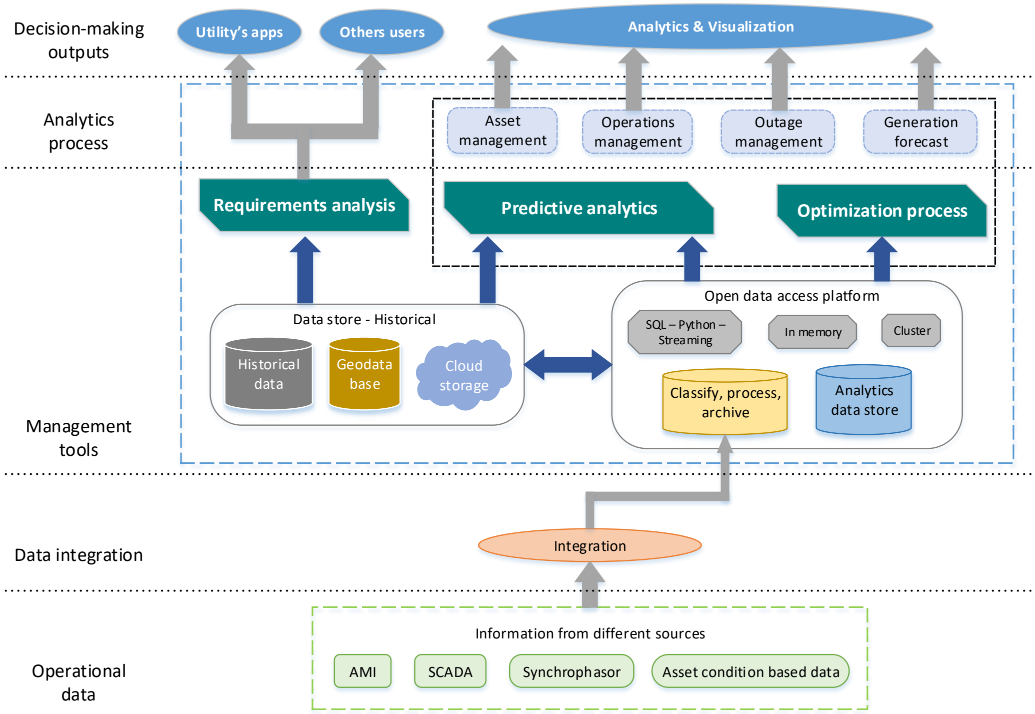

Fig. 6: Representative architecture for integration of weather data and analytics in an electric grid.

VI. Results and Discussion

As an example, the previous methodology is tested on a scenario in Brazos County. The scenario integrate the radiation data form NREL [29], with the public data available on the governmental website for the county of Brazos, and the cities of Bryan and College Station [30]. These information layers are integrated with the vegetation data for the Post Oak Savanah ecosystem form the official Texas ecological Mapping system data [31]. Finally, the resulting layer is populated with the substation locations in Brazos County using the information of the synthetic power grids data in Texas [32].

The goal of this methodology is to find the optimal placement for grid-connected solar generation, by correlating multiple factor such as radiation, build environment, vegetation layer and power system infrastructure. Fig. 7 presented the result obtained in this case study. As it can be seen from Fig. 7, the proposed methodology was able to identify the possible/feasible locations for the medium or large solar farms.

The main benefit of the proposed methodology is that it allows to show how one can correlate different features of interest using the publicly available data. These data sets need to be handled before the decision making process involves the integration of high variable resources, such as solar or wind, into the grid. The users of this study can be utilities companies, electric system operator and planner, as well as land developers. Also, this outcome might be useful for individuals who intend to put a PV installation on their land. With the provided information, utility companies or others can use their power systems data location to determine which of these options are optimal regarding the location of the current power system (distribution or transmission) and its actual physical state.

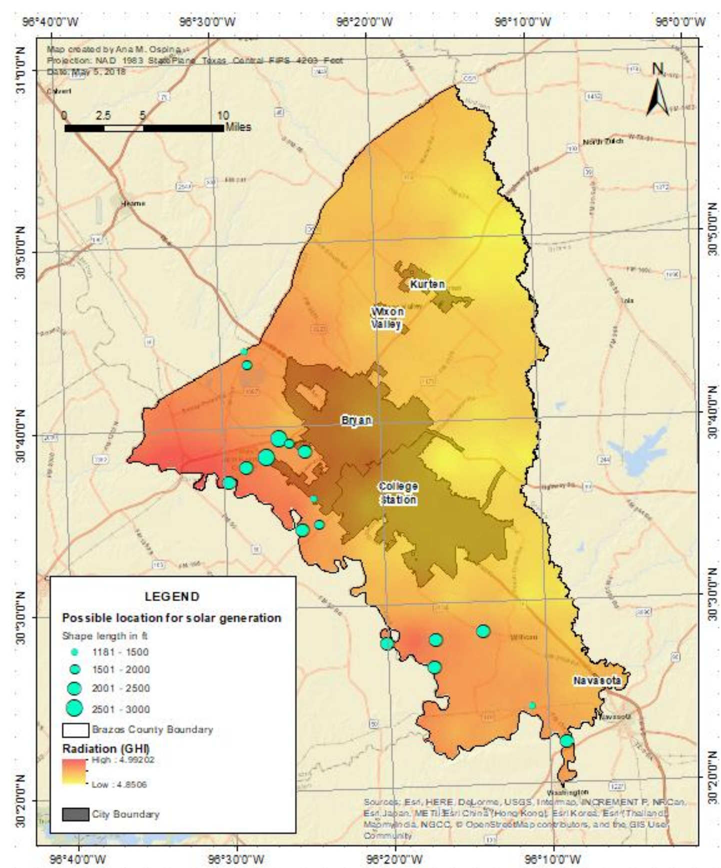

Fig. 7: Application example in Brazos County for optimal placement of grid-connected solar generation.

VII. Conclusion

This paper presents a systematic, practical and easily implementable methodology to optimally locate PV generation from GIS information using public data. In particular, the contributions are:

- Novel approach that can be used for the optimal placement of medium and large scale solar farm and can help with the decision making process previous to an investment analysis.
- Due to the high correlation of different types of data, this methodology is very versatile and provides the feasible information about optimal location correlated with the power system infrastructure.
- The proposed type of models are crucial to carry out planning and operation studies and are required when assessing the impacts of integration of non-conventional energy sources, such as solar or winding generation, into the grid in order to improve power systems reliability and resiliency.

References

[1] Mohamed Saleh, Petrus Pijnenburg, and Eduardo Castillo-Guerra, "Load aggregation from generation-follows-load to load-follows-generation: Residential". *IEEE Transactions on Industry Applications*, 53:833-842. March/April 2017.

[2] Benjamin Kroposki. "Integrating high levels of variable renewable energy into electric power systems". *Journal of Modern Power System and Clean Energy*, 5(6):831 – 837, November 2017.

[3] A. Medina, J. C. Hernandez and F. Jurado, "Optimal Placement and Sizing Procedure for PV Systems on Radial Distribution Systems," 2006 International Conference on Power System Technology, Chongqing, 2006, pp. 1-6.

[4] T. Sutthibun and P. Bhasaputra, "Multi-objective optimal distributed generation placement using simulated annealing," ECTI-CON2010: The 2010 ECTI International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Chiang Mai, 2010, pp. 810-813.

[5] K. Dharageshwari and C. Nayanatara, "Multiobjective optimal placement of multiple distributed generations in IEEE 33 bus radial system using simulated annealing," 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015], Nagercoil, 2015, pp. 1-7.

[6] Chaorui Zhang and Ying Jun Zhang, "Optimal solar panel placement in microgrids," 2016 IEEE International Conference on Smart Grid Communications (SmartGridComm), Sydney, NSW, 2016, pp. 376-381.

[7] M. Bhattacharya and D. Das, "Multi-objective placement and sizing of DGs in distribution network using genetic algorithm," 2016 National Power Systems Conference (NPSC), Bhubaneswar, 2016, pp. 1-6.

[8] S. Chowdhury and T. Matlokotsi, "Optimal placement and sizing of renewable distributed generation in electricity networks considering different load models," 2017 52nd International Universities Power Engineering Conference (UPEC), Heraklion, 2017, pp. 1-6.

[9] V. Kalkhambkar, R. Kumar and R. Bhakar, "Joint optimal sizing and placement of renewable distributed generation and energy storage for energy loss minimization," 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, 2017, pp. 1-9.

[10] M. Ahmadi, M. E. Lotfy, R. Shigenobu, A. Yona and T. Senjyu, "Optimal sizing and placement of rooftop solar photovoltaic at Kabul city real distribution network," in IET Generation, Transmission & Distribution, vol. 12, no. 2, pp. 303-309, 30 1 2018.

[11] K. M. L. Prasanna, A. Jain and R. J. R. Kumar, "Optimal distributed generation placement using hybrid technique," 2017 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Bangalore, 2017, pp. 1-6.

[12] G. Guerra and J. A. Martinez, "A Monte Carlo method for optimum placement of photovoltaic generation using a multicore computing environment," 2014 IEEE PES General Meeting | Conference & Exposition, National Harbor, MD, 2014, pp. 1-5.

[13] Z. Wang, B. Chen, J. Wang, J. Kim and M. M. Begovic, "Robust Optimization Based Optimal DG Placement in Microgrids," in IEEE Transactions on Smart Grid, vol. 5, no. 5, pp. 2173-2182, Sept. 2014.

[14] C. Camacho-Gómez, R. Mallol-Poyato, S. Jiménez-Fernández, L. Cornejo-Bueno and S. Salcedo-Sanz, "Optimal placement of distributed generation in micro-grids with binary and integer-encoding evolutionary algorithms," 2016 IEEE Congress on Evolutionary Computation (CEC), Vancouver, BC, 2016, pp. 3630-3637.

[15] D. K. Khatod, V. Pant and J. Sharma, "Evolutionary programming based optimal placement of renewable distributed generators," in IEEE Transactions on Power Systems, vol. 28, no. 2, pp. 683-695, May 2013.

[16] Sadie Cox, Anthony Lopez, Andrea Watson, Nick Grue and Jennifer E. Leisch. National Renewable Energy Laboratory (NREL), "Renewable Energy Data, Analysis, and Decisions: A Guide for Practitioners", Technical Report, March 2018.

[17] R. Li, X. Liu and Y. Sun, "Integrated Assessment Model for Solar Energy Resource Based on GIS," 2010 International Conference on Multimedia Technology, Ningbo, 2010, pp. 1-4.

[18] B. C. Surve, "Web based spatio -Temporal data modeling and application development for Solar Power plant siting using GIS for India," 2014 Conference on IT in Business, Industry and Government (CSIBIG), Indore, 2014, pp. 1-5.

[19] R. Guptha, H. Puppala and S. Kanuganti, "Integrating fuzzy AHP and GIS to prioritize sites for the solar plant installation," 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Zhangjiajie, 2015, pp. 465-470.

[20] H. Z. A. Garni and A. Awasthi, "A fuzzy AHP and GIS-based approach to prioritize utility-scale solar PV sites in Saudi Arabia," 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, 2017, pp. 1244-1249.

[21] Yusuke Kakumoto, Yuki Koyamatsu, Atsushi Shiota, Yaser Qudaih, and Yasunori Mitani. "Application of Geographic Information System to power distribution system analysis". *Energy Procedia, 100:360 – 365*, 2016. 3rd International Conference on Power and Energy System Engineering, CPESE 2016, 8-10 September 2016, Kitakyushu, Japan.

[22] P. Quesada, A. Arguello, J. Quiros-Tortos and G. Valverde, "Distribution Network Model Builder for OpenDSS in Open Sources GIS Software", 2016, *IEEE T&D PES Latin America*, Morelia, Mexico.

[23] Li, Wenyuan, "Risk Assessment of Power System Models, Methods, and Applications", *IEEE Press Series on Power Engineering. A Jhon Wiley & Sons, Inc,* 2005.

[24] Li, Wenyuan, "Probabilistec Tranmission System Planning", *IEEE Press Series on Power Engineering. A Jhon Wiley & Sons, Inc,* 2011.

[25] T. Dokic, M. Kezunovic, Z. Obradovic, and S. Roychoudhury, "Systematic framework for integration of weather data into prediction models for the electric grid outage and asset management applications". In 2018 *50th Hawaii International Conference on System Sciences (HICSS)*, January 2018.

[26] Jeffrey S. Katz, Chapter 3 – "*The Role of Big Data and Analytics in Utility Innovation*", Editor(s): Reza Arghandeh, Yuxun Zhou, Big Data Application in Power Systems, Elsevier, 2018, Pages 43-56.

[27] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics", *Inter. Journal Information Man.*, 137-144, 2015.

[28] J. Hu and A. V. Vasilakos, "Energy Big Data Analytics and Security: Challenges and Opportunities," in *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2423-2436, Sept. 2016.

[29] National Energy Renewable Laboratory. National Solar Radiation Data Base http://rredc.nrel.gov/solar/old_data/nsrdb/

[30] County of Brazos. http://gis.bisconsultants.com/brazoscad/

[31] Texas Parks & Wildlife. Texas ecological Mapping system data. https://tpwd.texas.gov/gis/programs /landscape-ecology/by-ecoregion-vector

[32] Texas A&M Univesiry, Synthetic networks project. https://electricgrids.engr.tamu.edu/.