frontiers | Frontiers in Energy Research

OPINION
published: 08 July 2022
doi: 10.3389/fenrg.2022.956543

Statistical Machine Learning Model for Uncertainty Analysis of Photovoltaic Power

Xueqian Fu[1,2,3,4]*, *Chunyu Zhang*[1,2,3,4] *and Xianping Wu*[1,2,3,4]

[1]*College of Information and Electrical Engineering, China Agricultural University, Beijing, China,* [2]*National Innovation Center for Digital Fishery, China Agricultural University, Beijing, China,* [3]*Key Laboratory of Smart Farming Technologies for Aquatic Animal and Livestock, Ministry of Agriculture and Rural Affairs, Beijing, China,* [4]*Beijing Engineering and Technology Research Center for Internet of Things in Agriculture, Beijing, China*

Keywords: photovoltaic power, statistical machine learning, generative adversarial networks, uncertainty, weather

OPEN ACCESS

Edited by:
Bo Yang,
Kunming University of Science and Technology, China

Reviewed by:
Lei Gan,
Hohai University, China
Limei Zhang,
Hebei Agricultural University, China

***Correspondence:**
Xueqian Fu
fuxueqian@cau.edu.cn

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *30 May 2022*
Accepted: *07 June 2022*
Published: *08 July 2022*

Citation:
Fu X, Zhang C and Wu X (2022) Statistical Machine Learning Model for Uncertainty Analysis of Photovoltaic Power. Front. Energy Res. 10:956543. doi: 10.3389/fenrg.2022.956543

INTRODUCTION

Under the goal of carbon peaking and carbon neutrality, the installed capacity of photovoltaics continues to grow explosively. By 2020, the cumulative installed capacity of photovoltaic power generation in China will reach 203 GW, ranking first in the world (Liu et al., 2021). Recently, China's National Development and Reform Commission has put forward the work requirements for the large-scale development of distributed photovoltaics in the whole county (city and district), which will further promote the development of distributed photovoltaic power generation and accelerate the clean and green transformation of energy structure (Long et al., 2022). In order to promote the green, safe, and efficient operation of distribution networks, it is necessary to focus on the key technologies of new distribution network modeling, planning, and regulation based on distributed photovoltaics. However, distributed photovoltaic power generation has strong uncertainty. How to accurately describe the uncertainty in the distributed photovoltaic power grid has become a key problem in distribution network modeling (A. R. Jordehi, 2018).

It can be found from the previous research results that probability theory has been proven to be an effective method to simulate uncertainty in the power grid. Lee and Baldick (2016)presented the wind load mode based on a generalized dynamic factor model. Sun et al. (2020)generated weather scenario data by deriving the conditional probability density function of current weather prediction using Bayesian theory. Compared with the probability model, the planning model based on statistical machine learning (SML) has good controllability and scalability (Fu, 2022). The SML model can more accurately describe the uncertainty of photovoltaic power generation to a certain extent by fully mining the multi-feature and high-dimensional raw data (Fu et al., 2021). Fu et al. (2015) combined multi-objective particle swarm optimization (MOPSO) and support vector machine (SVM) to solve the uncertain optimization problem of distributed generation in the distribution network. Huang et al. (2020)established a deep joint generation model of radiation-electric load-temperature scenario based on denoising variational autoencoder, according to different weather scenarios, which improved the reliability of energy supply in extreme scenarios). Fu et al. (2020) comprehensively used Markov chain and copula function to capture the uncertainty and correlation of weather scenarios. The aforementioned research shows us the potential of SML in power grid modeling and planning and proves that SML can be an effective tool to deal with uncertainty.

At present, the uncertainty modeling methods for photovoltaic output are mainly probability modeling methods based on traditional statistics and scenario generation methods. In recent years, the probability generation model in SML represented by generative adversarial networks (GAN) has been favored by researchers. It has been applied to the uncertainty of new energy power generation (Wang et al., 2019), providing a new idea for us to solve the uncertainty of photovoltaic output. Among the factors affecting photovoltaic output, the weather is the direct factor that leads to strong

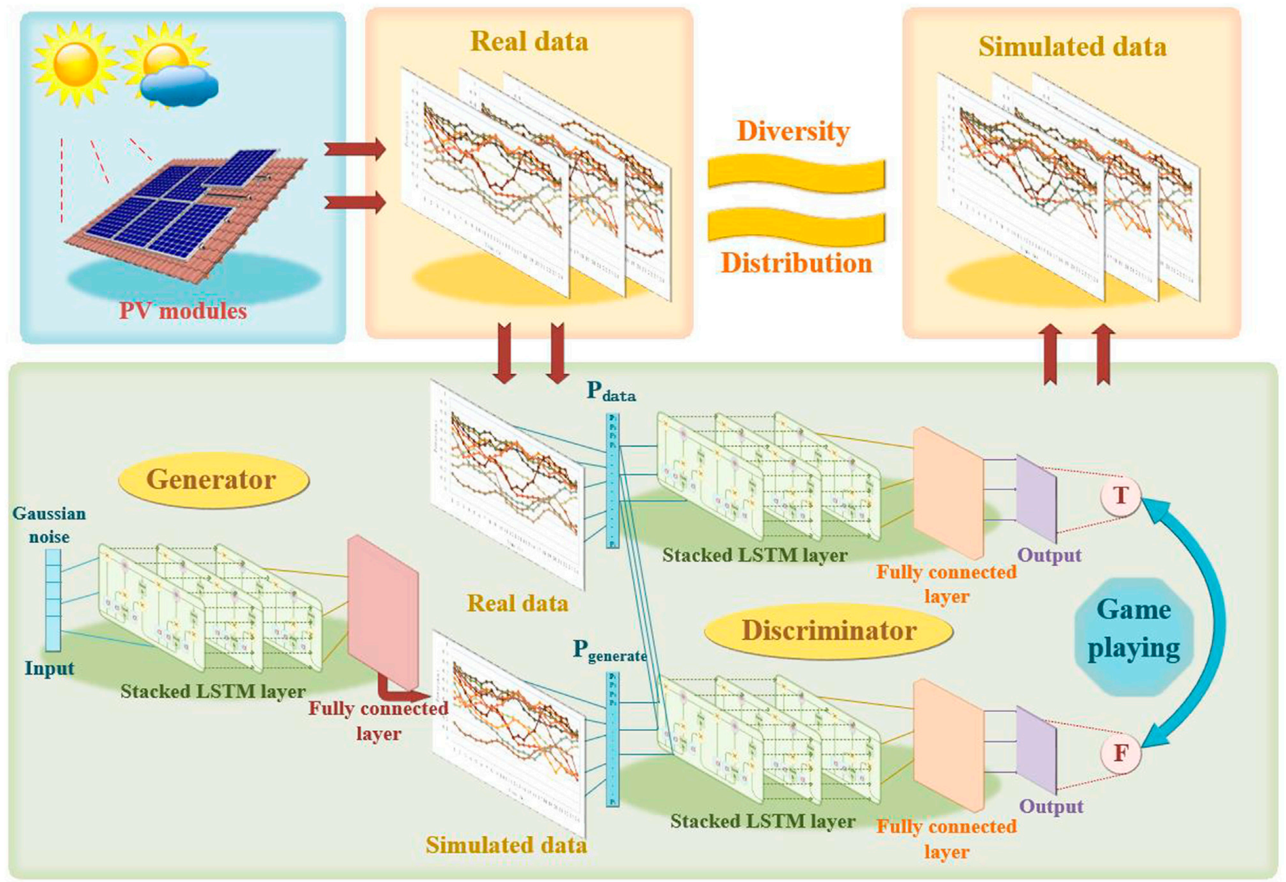

FIGURE 1 | Photovoltaic power generation uncertainty simulation method based on SML.

fluctuation and uncertainty of photovoltaic output (Fu and Zang., 2019). The uncertainty caused by weather affects the safe and stable operation of the power grid to a certain extent and is not conducive to the coordination and arrangement of power generation plans by the power grid dispatching department.

TECHNOLOGY METHODOLOGIES

This section consists of two parts. First, the photovoltaic power generation model is introduced in the *Photovoltaic power generation model* section. Second, in the *Photovoltaic power generation uncertainty simulation method based on SML* section, a photovoltaic power generation uncertainty simulation method based on SML is proposed.

Photovoltaic Power Generation Model

Solar radiation and temperature have a significant impact on the output of photovoltaic power generation. However, the micro weather scenario of photovoltaic power station is a complex nonlinear system, which has the characteristics of randomness, volatility, and uncontrollability. Due to the existence of season, day and night, overcast, sunny, rain, and snow, the solar radiation and temperature are uncertain. In addition, clouds have an important impact on the transmission of solar radiation. Due to the distribution differences in local cloud height, thickness, and shape, the solar radiation under the same cloud amount is not necessarily the same, so the temperature has greater uncertainty. At the same time, the micro weather scenario of the photovoltaic power station will also interact and affect other weather changes outside, making it more difficult to analyze the uncertainty of solar radiation and temperature.

Solar radiation directly affects the power output of photovoltaic cells. The photocurrent and photogenerated electromotive forces of photovoltaic cells are different under different solar radiation. The level of photovoltaic power directly depends on the intensity of solar radiation. Therefore, the uncertainty of solar radiation will directly affect the uncertainty of photovoltaic power generation. Solar radiation and air temperature directly affect the temperature of photovoltaic cells, affect the temperature drift of photovoltaic cells, and have a great impact on the work of photovoltaic cells, thus indirectly affecting the power output of photovoltaic cells. Therefore, the uncertainty of air temperature will also affect the uncertainty of photovoltaic power generation. To sum up, the power output of photovoltaic power generation is mainly affected by solar radiation and photovoltaic cell temperature. The power output of photovoltaic cells can be modeled according to these two variables (Rohani and Nour., 2014).

$$P = \frac{\delta U Re}{Rs}\left[1 + \eta\left(T_b - Ts\right)\right], \quad (1)$$

where δ represents the power derating factor (%); U is the rated capacity of photovoltaic cell array (kW); Re represents the solar radiation in the real environment (kW/m^2), with uncertainty; Rs is the inherent solar radiation under standard test conditions (STC) (kW/m^2); η represents temperature coefficient of a photovoltaic cell (%/°K); Tb and Ts, respectively, represent the real photovoltaic cell temperature and the cell temperature (°K) under STC, and Tb is directly affected by the temperature and solar radiation, resulting in strong uncertainty of Tb.

The air temperature and solar radiation directly affect the temperature of the photovoltaic cell. The air temperature and solar radiation are used to model the temperature sequence of photovoltaic cells corresponding to the time step. The formula is as follows (Brihmat and Mekhtoub., 2014):

$$T_b = \frac{Te + \frac{Re}{Rs}\left[1 - \frac{(1-\eta Ts)}{\tau\beta}\right]\left(T_n - T_{e,n}\right)}{1 + \frac{\alpha\eta R_e}{\gamma\beta R_{e,n}}\left(T_n - T_{e,n}\right)}, \quad (2)$$

where Te represents the temperature in the real environment (°K), with uncertainty; Tn refers to normal operating cell temperature (NOCT) (°K) of the photovoltaic module. Te,n and Re,n are the ambient temperature (°K) and solar radiation (kW/m^2) of corresponding Tn, respectively, which are fixed values; α, β, and γ are constants, which, respectively, represent the maximum power point efficiency, solar absorptivity of the photovoltaic cell array, and solar transmittance (%) of covering on photovoltaic cell array under STC. The temperature of photovoltaic cells is directly affected by air temperature and solar radiation. The physical constraints of two weather variables directly lead to the uncertainty of photovoltaic cell temperature and affect the photovoltaic output. Therefore, the air temperature and solar radiation that cause uncertainty are the focus of our attention.

Photovoltaic Power Generation Uncertainty Simulation Method Based on SML

The traditional probability model has some defects in dealing with uncertainty. Therefore, we proposed a photovoltaic power generation uncertainty simulation model based on SML, which uses GAN to simulate the photovoltaic output, as shown in **Figure 1**.

The aforementioned figure shows the specific modeling process, and the lower part of the figure is the uncertainty simulation of photovoltaic power generation using GAN. In GAN, the generator and discriminator are the key components. Because of the timing characteristics of photovoltaic power data, the generator and discriminator are internally composed of stacked long short-term memory (stacked LSTM). As shown in the figure, LSTM has a complex internal structure, which can alleviate the problems of gradient explosion and gradient disappearance during model training. Finally, both the generator and discriminator are followed by a fully connected layer to output the final results. The generator randomly samples from the real photovoltaic power data samples to learn the probability density function (PDF) of photovoltaic power data, so as to convert a random Gaussian noise into photovoltaic power data. The discriminator is a two-category model, which is responsible for distinguishing the authenticity of photovoltaic power data. Its goal is to judge the data generated by the generator as "false" and the real photovoltaic power data samples as "true" as far as possible. The goals of the generator and discriminator are contradictory. However, in the continuous confrontation game between the generator and discriminator, the generation ability of the generator is continuously improved to generate more real photovoltaic data. Jensen–Shannon divergence is usually used to calculate the distribution distance in GAN.

$$\mathrm{JS}\left(P_{\mathrm{data}}, P_{generate}\right) = \frac{1}{2}\int P_{data}\log\frac{2P_{data}}{P_{data} + P_{generate}}\mathrm{dx} + \frac{1}{2}\int P_{generate}\log\frac{2P_{generate}}{P_{\mathrm{data}} + P_{generate}}\mathrm{dx}, \quad (3)$$

where $Pdata$ is the PDF of the real photovoltaic power generation data, and $Pgenerate$ is the PDF of the data generated by the generator.

The traditional probability algorithms are complex and time-consuming, while GAN directly fits the probability distribution of real sample data, we do not need to explicitly specify the probability model or fit the characteristics of probability distribution and do not need to carry out complex sampling and manual labeling of data, which greatly reduce the computational burden. After the repeated confrontation game between the generator and the discriminator, the generator can learn the uncertainty and correlation of photovoltaic power generation data and generate a large number of high-quality and diverse photovoltaic power generation data. Therefore, with the help of SML, we can more efficiently analyze the uncertainty of distributed photovoltaic power generation.

DIFFICULTY ANALYSIS AND TREATMENT

In the traditional probability model calculation, how to ensure the correlation between the two variables is a difficulty, which usually needs to be handled by Copula function and other methods. However, in the neural network, just insert the data of multiple variables into a data sample at the same time and train the neural network, the neural network will actively establish a functional relationship between multiple variables and solve the problem of correlation.

The data of photovoltaic power generation in each time step is continuous. The data on timestamp are not isolated and static. It is affected by the data on the previous time step or even further time step. In view of the timing characteristics of photovoltaic power data, we introduce LSTM. LSTM is a variant of recurrent neural network (RNN). The RNN has the ability of memory. It can find the law from the historical sequence data and save this law in the network. On this basis, LSTM adds a cell state, which increases the capacity and pertinence of memory. Therefore, the use of LSTM can better capture the timing characteristics of photovoltaic power generation data.

In the simulation, how to make the data have diversity is difficult. The simulated data should not only comply with the real data distribution but also have diversity, so as to reflect the uncertainty of photovoltaic power data. The traditional tool for GAN to calculate the distribution distance is Jensen–Shannon divergence, but this method has many defects. The binary loss calculated by the discriminator through Jensen–Shannon divergence may be difficult to give full play to the learning ability of the generator and capture the timing distribution law in the photovoltaic power data. Based on the traditional loss function, the new loss function is added in this article.

$$\mathrm{Loss} = \sum_{t}^{n}\left(\left|\sqrt{\mathrm{D}(P_{\mathrm{data}})} - \sqrt{\mathrm{D}(P_{generate})}\right| + \left|\mathrm{E}(P_{data}) - \mathrm{E}(P_{generate})\right|\right), \tag{4}$$

where D represents the variance of the data, and E represents the mean of the data. For a series of data, the mean can be used to evaluate the size of the data, and the variance can be used to evaluate the dispersion of the data. Compared with only using Jensen–Shannon divergence as the loss function, adding a new loss function will improve the authenticity and diversity of the simulated photovoltaic power data and better reflect the uncertainty in the simulation data.

DISCUSSION

Due to the influence of weather and other uncontrollable factors, photovoltaic power data have high uncertainty and also have high-dimensional characteristics in time and space dimensions. In the traditional probability model, the methods used to describe the uncertainty of system variables mainly include the simulation method, analytical method, and approximate method. However, these algorithms are complex and time-consuming, so it is difficult to model and analyze the photovoltaic power generation data. On this basis, SML, as an effective combination of probabilistic model and machine learning, simplifies the calculation process by virtue of functional correlation. We use the GAN to model the uncertainty of photovoltaic power generation. We do not need to use Markov chain to repeatedly sample, infer in the learning process, or explicitly specify the probability model or fit the probability distribution characteristics, which greatly reduces the burden of solving probability problems and has good advantages in calculation speed.

Many existing achievements are based on weather scenario modeling, rather than direct modeling of photovoltaic power generation (Rohani and Nour., 2014). The photovoltaic power generation model uses solar radiation and air temperature to simulate photovoltaic power generation. If the power data on photovoltaic power generation are modeled directly, the constraints of weather uncertainty on power generation will be ignored. Simulating the weather scenario first can impose more strict physical constraints on the photovoltaic power generation model and ensure its universality.

CONCLUSION

Photovoltaic power generation is an important way to achieve the goal of carbon peaking and carbon neutrality. However, the randomness and uncontrollability of weather variables make distributed photovoltaic power generation have strong uncertainty. At the same time, the output of a photovoltaic power station is an uncertain variable. The universal access to photovoltaic power stations makes the uncertainty of photovoltaic power generation become a high-dimensional uncertainty problem. Therefore, how to better analyze the uncertainty of photovoltaic output has become a key problem. In order to solve this problem, we proposed the uncertainty simulation method of photovoltaic power generation based on SML, which improves the accuracy and credibility of the analysis results, enhances the calculation efficiency, generalization ability, and interpretability of the algorithm, effectively analyzes the uncertainty in photovoltaic power generation, and proves that SML can become a powerful tool to analyze and solve the uncertainty problem in distributed photovoltaic power generation.

In the development of the artificial intelligence era, deep learning and reinforcement learning in the field of SML have an active development trend, which will provide us with more new ideas and methods to analyze the uncertainty of photovoltaic power generation in the future. Deep learning is often applied to photovoltaic prediction (Wen et al., 2021), and reinforcement learning can be applied to control decision-making in photovoltaic power generation (Zhang et al., 2021). At the same time, with the upgrading of computer hardware, we can face high-dimensional and large-scale data with ease, which provides the possibility for us to solve the high-dimensional uncertainty of photovoltaic power generation.

AUTHOR CONTRIBUTIONS

XF conceived and wrote the general idea of the article. CZ drew and wrote the article. XW processed the formula and verified the article. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This study is supported by the National Natural Science Foundation of China under Grant 52007193, Yantai City School-Local Integration Development Project under Grant 2020XDRHXMPT10 and The 2115 Talent Development Program of China Agricultural University.

REFERENCES

Brihmat, F., and Mekhtoub, S. (2014). "PV Cell Temperature/PV Power Output Relationships Homer Methodology Calculation," in Conférence Internationale des Energies Renouvelables" CIER'13"/International Journal of Scientific Research & Engineering Technology (Pune: International Publisher & C. O), 0. No. 02.1

Fu, X., Chen, H., Cai, R., and Yang, P. (2015). Optimal Allocation and Adaptive VAR Control of PV-DG in Distribution Networks. *Appl. Energy* 137, 173–182. doi:10.1016/j.apenergy.2014.10.012

Fu, X., Guo, Q., and Sun, H. (2020). Statistical Machine Learning Model for Stochastic Optimal Planning of Distribution Networks Considering a Dynamic Correlation and Dimension Reduction. *IEEE Trans. Smart Grid* 11 (4), 2904–2917. doi:10.1109/tsg.2020.2974021

Fu, X. (2022). Statistical Machine Learning Model for Capacitor Planning Considering Uncertainties in Photovoltaic Power. *Prot. Control Mod. Power Syst.* 7 (1), 1–13. doi:10.1186/s41601-022-00228-z

Fu, X., Wu, X., and Liu, N. (2021). Statistical Machine Learning Model for Uncertainty Planning of Distributed Renewable Energy Sources in Distribution Networks. *Front. Energy Res.* 847. doi:10.3389/fenrg.2021.809254

Fu, X., and Zhang, X. (2019). Estimation of Building Energy Consumption Using Weather Information Derived from Photovoltaic Power Plants. *Renew. Energy* 130, 130–138. doi:10.1016/j.renene.2018.06.069

Huang, N., Wang, W., and Cai, G. (2020). Optimal Configuration Planning of Multi-Energy Microgird Based on Deep Joint Generation of Source-Load-Temperature Scenarios. *CSEE J. Power Energy Syst.* doi:10.17775/CSEEJPES.2020.01090

Jordehi, A. R. (2018). How to Deal with Uncertainties in Electric Power Systems? A Review. *Renew. Sustain. energy Rev.* 96, 145–155. doi:10.1016/j.rser.2018.07.056

Lee, D., and Baldick, R. (2016). Load and Wind Power Scenario Generation through the Generalized Dynamic Factor Model. *IEEE Trans. power Syst.* 32 (1), 400–410. doi:10.1109/TPWRS.2016.2562718

Liu, B., Song, C., Wang, Q., and Wang, Y. (2021). Forecasting of China's Solar PV Industry Installed Capacity and Analyzing of Employment Effect: Based on GRA-BiLSTM Model. *Environ. Sci. Pollut. Res.* 29, 4557–4573. doi:10.1007/s11356-021-15957-1

Long, H., Fu, X., Kong, W., Chen, H., Zhou, Y., and Yang, F. (2022). Key Technologies and Applications of Rural Energy Internet in China. *Inf. Process. Agric* (In Press). doi:10.1016/j.inpa.2022.03.001

Rohani, G., and Nour, M. (2014). Techno-economical Analysis of Stand-Alone Hybrid Renewable Power System for Ras Musherib in United Arab Emirates. *Energy* 64, 828–841. doi:10.1016/j.energy.2013.10.065

Sun, M., Feng, C., and Zhang, J. (2020). Probabilistic Solar Power Forecasting Based on Weather Scenario Generation. *Appl. Energy* 266, 114823. doi:10.1016/j.apenergy.2020.114823

Wang, F., Zhang, Z., Liu, C., Yu, Y., Pang, S., Duić, N., et al. (2019). Generative Adversarial Networks and Convolutional Neural Networks Based Weather Classification Model for Day Ahead Short-Term Photovoltaic Power Forecasting. *Energy Convers. Manag.* 181, 443–462. doi:10.1016/J.ENCONMAN.2018.11.074

Wen, H., Du, Y., Chen, X., Lim, E., Wen, H., Jiang, L., et al. (2021). Deep Learning Based Multistep Solar Forecasting for PV Ramp-Rate Control Using Sky Images. *IEEE Trans. Ind. Inf.* 17, 1397–1406. doi:10.1109/tii.2020.2987916

Zhang, X., Li, C., Li, Z., Yin, X., Yang, B., Gan, L., et al. (2021). Optimal Mileage-Based PV Array Reconfiguration Using Swarm Reinforcement Learning. *Energy Convers. Manag.* 232, 113892. doi:10.1016/J.ENCONMAN.2021.113892

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.