Optimizing Photovoltaic Power Forecasting Through Machine Learning Algorithms

Senye Zhang
Electrical and Computer Engineering
Lakehead University
Thunder Bay, Canada
ORCID:0009-0005-0121-2771

***Abstract*—This paper focuses on optimizing real-time PV power forecasting to support efficient grid dispatch. By leveraging local weather data, including solar irradiance and temperature, along with historical PV power output data, machine learning models such as Long Short-Term Memory (LSTM) and Random Forest were employed to improve the accuracy of PV generation predictions. The results indicate that these models effectively reduce prediction errors, enhancing the reliability of PV power forecasts. This research provides a foundation for optimizing grid dispatch and addressing the complexities of integrating renewable energy into the power grid.**

***Index Terms*—photovoltaic power forecasting, machine learning, solar energy prediction, renewable energy, power grid scheduling optimization, LSTM, random forest, energy management, stochastic optimization, time series forecasting**

I. INTRODUCTION

The increasing deployment of photovoltaic (PV) systems in Yantai, Shandong Province, has significantly reshaped the region's energy landscape. With the rapid expansion of PV capacity, the traditional peak electricity demand patterns, once driven by cooling loads at midday, have shifted. Due to the large influx of solar power, the grid now faces new challenges, including midday power backfeed and negative load conditions in some areas. Accurate PV power forecasting is essential to ensure stable grid operation and optimize dispatch strategies. However, existing forecasting methods often struggle to capture the variability of PV generation under dynamic weather conditions. This paper addresses the gap by utilizing machine learning algorithms to predict real-time PV power output based on local meteorological data, aiming to improve grid stability and inform future dispatch strategies.

II. LITERATURE REVIEW

Current methods for photovoltaic (PV) power forecasting can be categorized into three main types: statistical, physical, and machine learning methods. Statistical methods, such as ARMA, ARIMA, and exponential smoothing, rely on historical time series data and are effective for short-term forecasting, but struggle with nonlinear relationships. Physical methods are based on environmental variables like solar irradiance and temperature to model PV output, offering insights into weather impacts, but often lack flexibility for dynamic environments and require precise data. Machine learning methods, including Support Vector Machines (SVM), Random Forest, Gradient Boosting, and Long Short-Term Memory (LSTM) networks, are widely used for capturing complex and nonlinear relationships in PV power generation, providing high accuracy but requiring substantial amounts of training data. A growing trend is the use of hybrid methods, which combine physical models with machine learning techniques to enhance both accuracy and adaptability, addressing the variability and complexity in PV power forecasting more effectively.

Identify applicable funding agency here. If none, delete this.

III. METHODOLOGY

Machine learning methods have become increasingly popular in the field of photovoltaic (PV) power forecasting due to their ability to handle large datasets and capture complex, nonlinear relationships between various input features such as weather conditions and power output. Models like Support Vector Machines (SVM), Random Forest, Gradient Boosting, and Long Short-Term Memory (LSTM) networks have shown great potential in accurately predicting PV power generation, especially in dynamic and uncertain environments. These methods can incorporate diverse factors such as solar irradiance, temperature, and time-series data, making them more adaptive to changing conditions compared to traditional methods. However, there are still several limitations in the current research. First, machine learning models often require vast amounts of high-quality historical data for training, which may not always be available. Second, while these models can perform well in controlled environments, their accuracy may decline when faced with extreme weather conditions or in regions with insufficient data. Additionally, many machine learning models act as "black boxes," making it difficult to interpret how predictions are made, which can be problematic for energy operators who require transparency and reliability. Lastly, the computational complexity of training advanced machine learning models can be high, potentially limiting their practical deployment in real-time grid operations.

A. Data Collection and Preprocessing

Define abbreviations and acronyms the first time they are used in the text,

B. Machine Learning Models

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as "3.5-inch disk drive".

- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: "Wb/m^2" or "webers per square meter", not "webers/m^2". Spell out units when they appear in text: ". . . a few henries", not ". . . a few H".
- Use a zero before decimal points: "0.25", not ".25". Use "cm^3", not "cc".)

C. Model Evaluation Metrics

Number equations consecutively. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \tag{1}$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(1)", not "Eq. (1)" or "equation (1)", except at the beginning of a sentence: "Equation (1) is . . ."

D. LaTeX-Specific Advice

Please use "soft" (e.g., `\eqref{Eq}`) cross references instead of "hard" references (e.g., `(1)`). That will make it possible to combine sections, add equations, or change the order of figures or citations without having to go through the file line by line.

Please note that the `{subequations}` environment in LaTeX will increment the main equation counter even when there are no equation numbers displayed. If you forget that, you might write an article in which the equation numbers skip from (17) to (20), causing the copy editors to wonder if you've discovered a new method of counting.

LaTeX does not have precognitive abilities. If you put a `\label` command before the command that updates the counter it's supposed to be using, the label will pick up the last counter to be cross referenced instead. In particular, a `\label` command should not go before the caption of a figure or a table.

Do not use `\nonumber` inside the `{array}` environment. It will not stop equation numbers inside `{array}` (there won't be any anyway) and it might stop a wanted equation number in the surrounding equation.

E. Some Common Mistakes

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o".
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an "inset", not an "insert". The word alternatively is preferred to the word "alternately" (unless you really mean something that alternates).
- The abbreviation "i.e." means "that is", and the abbreviation "e.g." means "for example".

An excellent style manual for science writers is [**?**].

F. Authors and Affiliations

The class file is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

G. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is "Heading 5". Use "figure caption" for your Figure captions, and "table head" for your table title. Run-in heads, such as "Abstract", will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

H. Figures and Tables

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1", even at the beginning of a sentence.

TABLE I
TABLE TYPE STYLES

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a]Sample of a Table footnote.

Fig. 1. Example of a figure caption.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

IV. Results and Discussion

A. Forecasting Accuracy Comparison

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

B. Impact of Weather Variables on Prediction

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

C. Implications for Power Grid Scheduling Optimization

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

V. Conclusion

VI. Future Work

Acknowledgment

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

References

Please number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [**?**]—do not use "Ref. [**?**]" or "reference [**?**]" except at the beginning of a sentence: "Reference [**?**] was the first ..."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [**?**]. Papers that have been accepted for publication should be cited as "in press" [**?**]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [**?**].

References

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.