A procedure for accuracy increasing of a solar power plant short-term generation forecast

Natalya Lisitskaya[1], Semen Frid[1], Sofya Kiseleva[1,2]*

[1]*Joint Institute for High Temperatures of the Russian Academy of Sciences, Russia*

[2]*Lomonosov Moscow State University, Russia*

**Corresponding author's e-mail: k_sophia_v@mail.ru*

Abstract

A method has been developed for refining a short-term forecast of solar power plant performance using machine learning, which allows reducing financial losses of inaccurate forecast in most cases to 2-7% of the generated electricity cost and is approximately an order of magnitude more accurate than the dynamic simulation of solar power plant operation based on numerical weather prediction data.

Keywords: solar power plants, short-term forecast, financial losses, forecast errors

Introduction

The growing fraction of renewable generation in power network makes it difficult to maintain a balance of power and energy in them. Due to the non-guaranteed electric power output at exactly the specified time by solar power plants (SPP), it becomes relevant to develop methods for short-term forecasting of the performance of these plants [1]. Particular attention is paid to the forecast for the generation of solar power plants for the day ahead with an hourly resolution, which is due to the current procedure for electricity trading in the wholesale market in many countries of the world, including Russia. The rules of the Russian wholesale electricity and capacity market establish permissible deviations of actual production from hourly volumes in the price application for the day ahead market (DAM) for SPPs in the amount of 10% of the installed capacity of generating equipment. The financial losses of the generating organization consist of lost profits (results from the fact that excess energy is paid at the rate of 1 rouble/MWh) and losses from negative imbalance (they can be interpreted as the need to purchase on the balancing market the missing energy for wholesale market supply).

A significant reduction in financial losses was achieved by solar power plants modeling using numerical weather prediction (forecasting the level of insolation and temperature in the surface air layer) with a combination of machine learning methods.

Materials, methods and results

Of the many existing forecasting error metrics [1], the root mean square error (RMSE) normalized to the installed capacity of generating equipment was used. In addition to it, financial losses from forecast errors were also used directly. Preliminary estimates showed that a halving of the RMSE forecast leads to a 5-9-fold decrease in losses.

The calculation was carried out throughout Russia (3241 grid points 1°×1° in latitude and longitude) excluding the territories served by the Far East power network and technologically isolated power networks for the period from January 2020 to July 2021.

The calculated output of the photovoltaic panel (PVP) operating at the maximum power point was used as the actual SPP output. PVPs of the most widespread types in Russia (STP380S-B60/Wnh [2] and HVL-320/HJT [3]), oriented to the south and installed at an angle of inclination to the horizon less than the latitude by 12°, were considered. The generation calculation was carried out by dynamic simulation with the TRNSYS system [4] using a single-diode five-parameter PVP model. The calculations use actinometric data of NASA Clouds and the Earth's Radiant Energy System (CERES) [5] (according to [6] CERES satellite observations are the most accurate available for the territory of Russia) and averaged over the CERES spatial grid (1°×1° in latitude and longitude) ERA5 data [7] for ambient air temperature and wind speed.

An analysis of generation forecasting methods showed that for a forecast horizon of more than 6 hours, the forecast is almost always based on a numerical weather prediction (NWP). The generation forecast was calculated using the data of the NWP ICON model [8] averaged over the CERES grid, from which, according to daily forecasts for the next day, continuous time series of hourly data were built. To form the SPP forecast order for the DAM, submitted before 13:30 Moscow time, taking into account the delay in uploading the ICON data to the server, the results of this NWP, performed at 06:00 UTC, were used.

Before TRNSYS simulations the hourly sums of solar radiation from CERES and ICON were upper bounded by the values for a clear day.

The RMSE of the predicted output from the calculated one was 72-85 Wh (19-22%) for the STP380S-B60/Wnh PVP, and 60-75 Wh (19-23%) for the HVL-320/HJT PVP. A pronounced influence of the solar power plant location on the accuracy of the generation forecast was not found.

The purpose of updating the short-term forecast was to reduce the financial losses of the generating company. Since the Russian wholesale electricity and capacity market operates only in price zones, only the territories included in the first (747 1°×1° points) and second (550 points) price zones were considered. The indicators of the balancing market, averaged for the price zones, together with the day-ahead market prices, were taken from the website of system operator. A direct forecast method was implemented in which the predictive power calculated from the NWP data was used along with other characteristics for machine learning. In order to save time and resources, 4 classic machine learning models [9] were used separately for each point: linear and polynomial regressions, random forest and boosting [10].

An annual period was chosen for training the models, which avoids data imbalance. The remaining data from the sample (from January 18 to July 31, 2021) were used for cross-validation of models according to the following scheme: forecasting for 72 hours, calculation of metrics for the second half of the forecasting horizon, shift by 24 hours, which were added to the training set. The procedure was repeated 193 times. The model was evaluated for the entire period of cross-validation. Financial losses were summed up, the RMSE was calculated. The python libraries sklearn and xgboost were used.

More accurate regression results are observed for random forest (Random Forest Regressor of the sklearn library) and boosting on 35 features, although both methods require more time to calculate than linear or polynomial regression.

Analysis of the results of refinement of the production forecast by modeling using machine learning methods showed that the smallest financial losses in most cases are obtained by random forest method (64%), linear (26%) and polynomial (9%) regression, boosting – in 1%. For fast modeling and preliminary evaluation of the result, linear regression shows good results both in terms of speed and quality. For practical use, preference can be given to a random forest with saving the trained model and calling it to get a forecast, because these models have the ability to parallelize processes and do not require such costs as neural networks.

The developed short-term forecasting method allows to reduce financial losses in most cases to 5-15 rubles for the control 193 days per one PVP, which is about 2-7% of the cost of generated electricity and is approximately an order of magnitude lower than the loss estimate based on the results of dynamic simulation of the SPP operation according to ICON NWP forecast data.

An analysis of the influence of the location of solar power plants on the accuracy of the performance forecast showed that a large amount of financial losses from forecasting inaccuracy is typical for the polar part of the first price zone, which is unsuitable for the construction of SPPs, the smallest losses are observed in the southern part of the Krasnoyarsk Territory (Altai, Khakassia, Tyva) and in the northern part of the Irkutsk region.

Conclusion

The developed method for refining the short-term forecast makes it possible to reduce financial losses from inaccurate forecasting for most geographic points in the territory of the price zones of the Russian Federation to 2-7% of the cost of generated electricity, which is approximately an order of magnitude lower than the loss estimate based on the results of dynamic simulation of the SPP operation using ICON forecast data.

Acknowledgments

The study was supported by RFBR (project № 19-08-00877).

References

[1] Antonanzas J. et al. 2016 *Solar Energy* **136** 78-111
[2] Characteristics of STPXXXS-B60/Wnh. https://www.suntech-power.com/wp-content/uploads/download/product-specification/EN_Ultra_S_mini_STP385S_B60_Wnh.pdf (27.03.2022)
[3] HVL-320-335/HJT https://www.enfsolar.com/pv/panel-datasheet/crystalline/47300 (27.03.2022)
[4] Duffie, J. and Beckman, W. 2013 *Solar Engineering of Thermal Processes*, New York: Wiley
[5] Clouds and the Earth's Radiant Energy System https://ceres.larc.nasa.gov (27.03.2022)
[6] Yang D., Bright J.M. 2020 *Solar Energy* **210** 3-19
[7] Copernicus. ERA5 hourly data on single levels from 1979 to present https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview (27.03.2022)
[8] Dwd. Our Services. NWP forecast data. https://www.dwd.de/EN/ourservices/nwp_forecast_data/nwp_forecast_data.html (27.03.2022)
[9] Hyndman R J and Athanasopoulos G 2018 *Forecasting: Principles and Practice* Melbourne: OTexts
[10] Brownlee J 2018 *XGBoost With Python* Machine Learning Mastery