# Photovoltaic System Power Output Forecasting Using Vector Auto-Regression (VAR)

Agussalim Syamsuddin, Andrew Cahyo Adhi, Achmad Widodo, Toni Prahasto and Amie Kusumawarhani



November 26, 2024

# Photovoltaic system power output forecasting using vector auto-regression (VAR)

1st Agussalim Syamsuddin
*Research Division*
*PT PLN (Persero) Research Institute*
Jakarta, Indonesia
agussalim.s@pln.co.id

2nd Andrew Cahyo Adhi
*Research Division*
*PT PLN (Persero) Research Institute*
Jakarta, Indonesia
andrew.cahyo@pln.co.id

3rd Achmad Widodo
*Dept. of Mechanical Engineering*
*Diponegoro University*
Semarang, Indonesia
achmadwidodo@lecturer.undip.ac.id

4th Toni Prahasto
*Dept. of Mechanical Engineering*
*Diponegoro University*
Semarang, Indonesia
toniprahasto@lecturer.undip.ac.id

5th Amie Kusumawardhani
*Dept. of Economy and Business*
*Diponegoro University*
Semarang, Indonesia
amiekusumawardhani
@lecturer.undip.ac.id

***Abstract*—As the penetration of solar energy sources into a power system increases, the significance of precise short-term forecasts for solar power plants becomes paramount. However, the erratic and non-periodic nature of solar poses challenges in accurately predicting the output power. The main objective of this paper is to study power output forecasting in photovoltaic systems based on vector autoregression (VAR) and SCADA data. The study of forecasting aims to predict the condition of specified parameters e.g., power output for the next time ahead to set up the planning for anticipation action if the predicted parameters fall into bad conditions. In this study, the data utilized to train the VAR algorithm originates from seven months of historical solar power plant operations collected from 1 December 2022 to 30 September 2023. The trained VAR is then applied to forecast one month ahead of the power output. The results are evaluated with performance forecasting measures namely nRMSE and nMAE which are relative measures with a basis on maximum power output generated in photovoltaic systems. Referring to the normalized RMSE, the maximum active power in this study reached 1350 kW for July 2023, and then the nRMSE and nMAE were 0.06 and 0.04, respectively. The nRMSE and nMAE give small values that mean the proposed method is admissible and could be applied to real solar power plant systems in forecasting the output power.**

***Keywords—photovoltaic; forecasting; performance; VAR; SCADA***

## I. INTRODUCTION

Since the issue of sustainable and renewable energy has become massive as well as concerns in environmental conservation, the demand for green electricity is uphill. The use of fossil fuels in power plants has been having negative impacts on the environment and life. Air pollution can break human health because of dissolved contaminants in the air that result from the imperfect burning of fossil fuels. In addition, energy security is also a factor in why the need for green energy continues to increase. Recently, the European Union announced the REPowerEU plan for developing more affordable, secure, and sustainable energy sources such as solar power to displace the consumption of natural gases [1]. Therefore, studies in the field of power plants through the photovoltaic system have received a lot of attention from researchers [2]. Numerous studies have predicted solar power generation using various methods, one of which is the application of forecasting methods using machine learning (ML) to support the predictive maintenance of PV systems as reported in [3]. They presented a review of the state-of-the-art comparing the technique to forecast solar irradiance and ambient temperature and their relationship with predictive maintenance models. This review included an evaluation of the suitability of the model for PV systems taking on their forecasting horizon, computational cost, and improvement opportunities for future work in the PV industry.

Time series analysis and machine learning methods have been widely used in forecasting active power in solar power plants. The work reported in [4] used the time series method and neural network model to predict ultra-short-term solar power. The neural network also has been applied with an auto-regressive integrated moving average (ARIMA) in the prediction of power consumption [5]. The ARIMA prediction model performance was claimed could be improved by applying multi-fractal noise to the noise of the time series in [6]. Other researchers used the autoregressive method in forecasting the power of PV systems are reported in [7], [8].

The application of VAR in power output forecasting is found in [9]. They demonstrate a correlation between predicted and actual solar irradiance, indicating the effectiveness of the VAR model for this task. A study in [10] applied two spatiotemporal forecasting methods, namely, space-time kriging and vector autoregressive models. They find that the forecasting accuracies do not increase by including more stations beyond the threshold distance. This result allows for the design of simplified monitoring networks and improved forecasting techniques. The application of VAR in weather forecasting was reported in [11]. This VAR model was built to forecast three important weather variables for 61 cities around the United States. Weather forecasting is crucial to both the demand and supply sides of the electricity system. The results showed that the proposed time series approach is appropriate for very short-term forecasting of hourly solar radiation, temperature, and wind speed.

This paper deals with the study of forecasting methods applied to photovoltaic systems in predicting active power. Data used to validate the proposed method is real data acquired from solar power plants through a supervisory control and data acquisition (SCADA) system. Sensor parameters involved in system development are solar irradiance, ambient temperature, humidity, wind direction, and wind speed while the calculated parameter used is active power AC. Historical data of such parameters are utilized to train a time series analysis method so-called vector autoregression (VAR) to predict the future condition of active power.

The structure of this paper is summarized as follows: Section 2 describes an overview of the basic theory of VAR and data acquisition of PV system parameters. Section 3 presents the architecture of the proposed method addressed to achieve the aims of the study. Section 4 describes the results of the study including a discussion of some findings. Finally, the end of the paper consists of conclusions and gives some perspective issues that relate to solar power plants.

## II. MATERIALS

### A. *Theoretical Base*

The model of vector auto-regression (VAR) is a multivariate forecasting algorithm. It is used when two or more time series variables are affecting each other. Each variable is modeled as a function of its past value, or time-delayed series value, considered by the autoregression model [12]. The VAR terminology involves the generalization of the univariate autoregression model to a vector of variables. The VAR model is a stochastic process that, as a linear function of its past values and the past values of all other variables in the group, represents time-dependent variables. Thus, the VAR model is formed as an equation of stochastic differences. In general, an autoregression model is a linear time series equation with a set of lag values combined. The set of lag values in the time series is used to predict the current value and future value.

VAR is an extension of the auto-regression model, which is a univariate time-series analysis model, to a multivariate model and reflects the correlation variables [13]. VAR model has parameter p which is a non-negative integer representing the order of the maximum lag. And usually denotes as VAR(p). The order of the VAR model is determined by using Akaike's information criterion (AIC) [4], the corrected AIC (AICc), and Schwartz's criterion (SBC) [5]. It is given by

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + \varepsilon_t\ , \quad t = 1, 2, \dots, T \tag{1}$$

$$Z_t = \begin{bmatrix} Z_{1t} \\ \vdots \\ Z_{nt} \end{bmatrix},\ \phi_t = \begin{bmatrix} \phi_{11}^{(i)} & \cdots & \phi_{1n}^{(i)} \\ \vdots & \ddots & \vdots \\ \phi_{n1}^{(i)} & \cdots & \phi_{nn}^{(i)} \end{bmatrix}, \quad i = 1, 2, \dots, p\ ,$$

$$\varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix} \tag{2}$$

In a typical vector auto-regression (VAR) model, the VAR(p) equation is denoted as in Equation (1), where $\boldsymbol{\phi}$ is the coefficients of lag from $t$-1 to $p$. Order $p$ is the $p$-lag value of $Z$, and they are the predictors in the equation. The error term is $\varepsilon t$ [14]. Vector autoregression (VAR) is the main element and particular case of the moving average (MA) model, including the autoregression integrated moving average (ARIMA) and the autoregression moving average (ARMA) time series models. The vector autoregression model has a more complex stochastic structure. The vector autoregression model (VAR) consists of an equation of two or more interlocking equations of stochastic difference, which contains two or more evolving random variables. VAR, AR, ARMA, and ARIMA algorithms are similar in that they require a series of observations to train the model before using the model for forecasting. However, the difference between VAR and the rest of the algorithms is that VAR is suitable for multivariate data, whereas AR, ARMA, and ARIMA are univariate models. The linear regressive variable Z is affected by its past value or predictors, but not the other way around. On the contrary, VAR is bidirectional, and its variables affect each other [15].

### B. *PV System Data*

The data used in this work was collected from a 3.5 MWp solar power plant that consists of 7,778 PV modules. SCADA system provides one-minute operational data acquired from environmental sensors: solar irradiance (pyranometer), ambient temperature, and wind speed and direction. The configuration of the PV system is summarized as follows:

1) *Inverter: #1- #9 (F1)*
   a) Zone B: 13 Strings
   b) Zone C: 149 Strings
2) *Inverter: #10 - #18 (F2)*
   a) Zone A: 111 Strings
   b) Zone B: 46 Strings
   c) Zone C: 4 Strings

The data used for validation of the proposed method was collected from 1 December 2022 to 30 September 2023 with data number 437,759 points. Some of the data will be used to build the VAR model and partly to test the model. Not all the weather parameters measured by environmental sensors are used to train the VAR model. Only two parameters of data: irradiance and ambient temperature were used to train the VAR model. The presentation of data daily yield power AC calculated in the inverter is shown in Figure 1. In addition, data on daily ambient temperature and irradiance during hours are shown in Figure 2 and Figure 3.

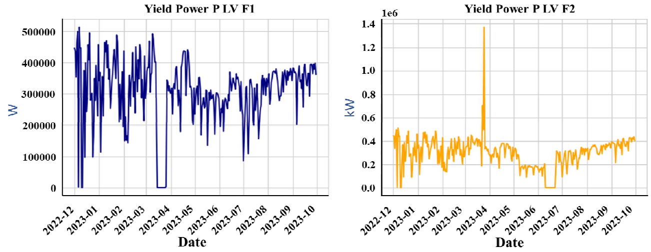


Fig. 1. Data daily yield power AC calculated from the inverter.

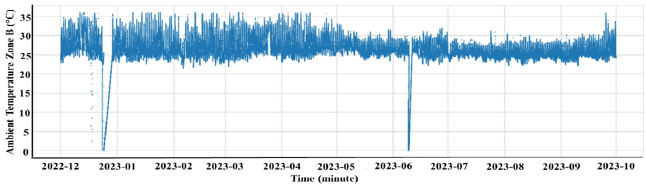


Fig. 2. Data on daily ambient temperature.

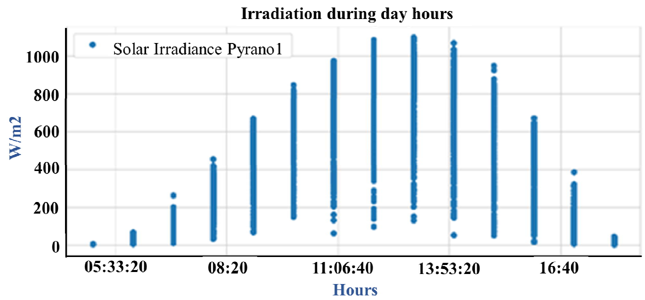


Fig. 3. Data of irradiance during hours.

This dataset seems to have a high variance. However, not all data points are used to build and validate the model because those tend to be burdensome, irrelative, and do not represent normal behavior of physical parameters. Therefore, the following steps are executed to reduce data size as well as extract meaningful information from data parameters.

- Remove data with active power near zero and irradiance slightly greater than zero.
- Remove data that happened during very low sunlight approximately at the time before 5.30 AM and after 6.45 PM.
- Resample data into one-hour intervals with averaging.

The resampling of one-hour averaging will decrease the variability in the dataset and can contribute to improving the accuracy of the model constructed. Applying the above procedure to the data can reduce the data size become 3,520 points of hourly measurement.

## III. Method

Figure 4 depicts the diagram for active power forecasting using the VAR model. The training process for VAR model building is carried on by feeding historical data into the VAR algorithm. Historical data stored in the database provides information on power plant conditions through parameters in time series format. The inverter as a critical component in the PV system can calculate and convert active power resulting from PV modules in DC to AC as time series data. VAR as a tool for time series analysis is employed to predict the active power AC for the next step-ahead time in the future. The forecasting method in VAR employs long-term memory for time series analysis.

As presented in Eqs. (1) and (2), to calculate Z_t, VAR will use the past values of Z_(t-1)and Z_(t-2) and so on till order parameter p is completed. For example, the vector autoregressive model of order 1 dan 2 is denoted as VAR(1) and VAR(2), and the lag two values for all variables are added to the right side of the equation. For a VAR(p) model, the first p lags of each variable in the system would be used as regression predictors for each variable. Training of VAR is also addressed to update the parameter p using previous historical data. Active power is then predicted using updated VAR and the prediction results will be evaluated by using matrix performance evaluation such as RMSE and MAE. If the prediction reaches good agreement with the actual active power this means the VAR prediction model can be regarded as a valid model and saved in the database.

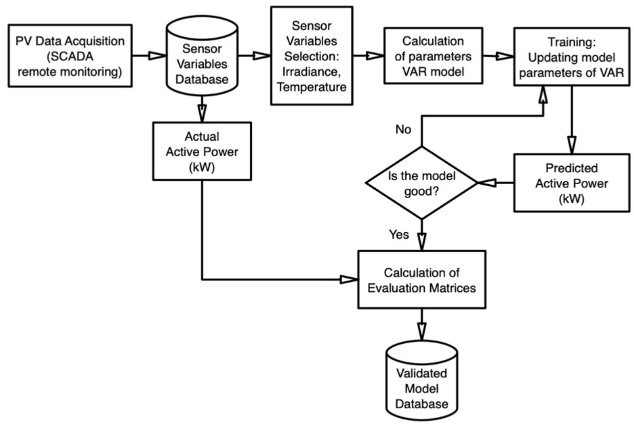


Fig. 4. Schematic diagram of active power forecasting.

## IV. Result and Discussion

The VAR model is trained by historical data of active power acquired from 1 December 2022 to 30 June 2023 and the trained VAR is used to predict the remaining data of 1 - 30 July 2023. The data that are required to obtain the forecasted active power values are historical active power, historical data of irradiance, and ambient temperature. The result of forecasting using the VAR model is shown in Figure 4.

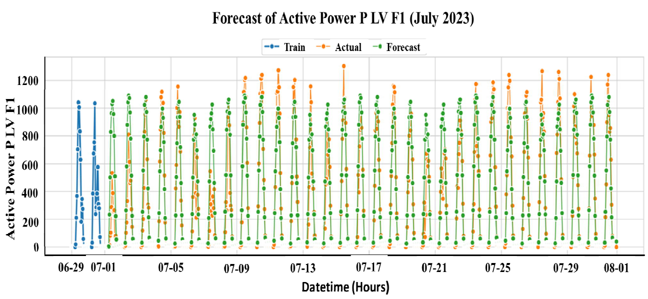


Fig. 5. Active power forecasting of July 2023.

Observing Figure 5, the blue line refers to partly training data, the orange line is actual active power and the green line represents forecasted data of active power. Performance prediction matrices through RMSE and MAE give 79.2 and 48.7, respectively. The predicted actual power can recognize the peaks almost at a similar time when compared to the actual one, but some of them failed to reach the maximum values of peaks actual power. Performance prediction that is expressed using RMSE and MAE seems good enough since the horizon of prediction is almost one month ahead. Performance prediction or forecasting can also be expressed in normalized form as suggested in [16]. They employed VAR for forecasting photovoltaic power and measured the performance of forecasting using normalized RMSE (or nRMSE) and MAE. The basis of normalization is the capacity of solar power plants. Referring to the normalized RMSE, the maximum active power in this study reached 1350 kW for July 2023, and then the nRMSE and nMAE were 0.06 and 0.04, respectively.

## V. Conclusion

In this paper, the study of power output forecasting of photovoltaic systems is modeled using vector autoregression (VAR). The data used in the experimental works was acquired from a real solar power plant system through a SCADA system. The VAR forecasting model is constructed by training the VAR algorithm with historical operational data acquired from 1 December 2022 to 30 June 2023. The trained VAR prediction model is then applied to testing data for power output prediction of July 2023. The results show that the performance prediction measured by nRMSE and nMAE falls into relatively small values based on maximum active power production. The proposed method seems to have a good performance and could be implemented in real solar power plant systems for predicting output power conditions.

## Acknowledgment

The authors would like to thank PT PLN Indonesia Power for supporting this study through permission to use real PV system operating data. Thanks also goes to Yasir Abdur Rahman and Ghiffari Awliya for computer programming assistance.

## References

[1] E. C. (EU), "REPowerEU: Affordable, Secure and Sustainable Energy for Europe," 17 November 2023. [Online]. Available: https://ec.europa.eu/info/strategy/priorities-2019-2024/european green-deal/repowereu-affordable-secure-and-sustainable-energy-europe_en.

[2] A. K. Azad, "Bibliometric analysis of photovoltaic thermal (PV/T) system: From citation mapping to research agenda," Energy Reports, vol. 8, no. 2, pp. 2699-2711, 2022.

[3] J. R. Vergara, L. B. Bosman, E. Wollega and W. D. Leon-Salas, "Review of forecasting methods to support photovoltaic predictive maintenance," Cleaner Engineering and Technology, vol. 8, p. 100460, 2022.

[4] L. Wang, H. Hu, X. -Y. Ai and H. Liu, "Effective electricity energy consumption forecasting using echo state network improved by differential evolution algorithm," Energy, vol. 153, p. 801–815, 2018.

[5] V. Kushwaha and N. M. Pindoriya, "A SARIMA-RVFL hybrid model assisted by wavelet decomposition for very short-term solar PV power generation forecast," Renew. Energy, vol. 2019, pp. 124-139, 2019.

[6] R. Baile, J. F. Muzy and P. Poggi, "Short-term forecasting of surface layer wind speed using a continuous random cascade model," Wind Energy, vol. 14, pp. 719-734, 2011.

[7] R. J. Bessa, A. Trindade, and V. Miranda, "Spatial-temporal solar power forecasting for smart grids," IEEE Transactions on Industrial Informatics, vol. 11, no. 1, pp. 232-241, 2014.

[8] H. Kim and D. Lee, "Probabilistic Solar Power Forecasting Based on Bivariate Conditional Solar Irradiation Distributions," IEEE Transactions on Sustainable Energy, vol. 12, no. 4, pp. 2031-2041, 2021.

[9] F. Z. Najdawi and R. Villarreal, "Utilizing the Vector Autoregression Model (VAR) for Short-Term Solar Irradiance Forecasting," Energy and Power Engineering, vol. 15, pp. 353-362, 2023.

[10] D. Yang, Z. Dong, T. Reindl, P. Jirutitijaroen and W. M. Walsh, "Solar irradiance forecasting using spatio-temporal empirical kriging and vector autoregressive models with parameter shrinkage," Solar Energy, vol. 103, pp. 550-562, 2014.

[11] Y. Liu, M. C. Robert, and R. Sioshansi, "A vector autoregression weather model for electricity supply and demand modeling," J. Mod. Power Syst. Clean Energy, vol. 6, no. 4, p. 763–776, 2023.

[12] Stock, J.H.; Watson, M.W. Vector autoregressions. J. Econ. Perspect., 15, 101–115, 2001.

[13] H. Akaike, "A new look at the statistical model identification," IEEE Trans. Autom. Control, vol. 19, p. 716–723, 1974.

[14] Zivot, E.; Wang, J. Vector autoregressive models for multivariate time series. In Modeling Financial Time Series with S-PLUS®; Springer: Berlin/Heidelberg, Germany, pp. 385–429, 2006.

[15] Hyndman, R.; Athanasopoulos, G. Forecasting: Principles and Practice, 2nd ed.; OTexts: Melbourne, Australia, 2018.

[16] A.-H. Jung, D.-H. Lee, J.-Y. Kim, C. K. Kim, H.-G. Kim and Y.-S. Lee, "Regional Photovoltaic Power Forecasting Using Vector Autoregression Model in South Korea," Energies, vol. 15, no. 21, p. 7853, 2022.