

One-step and Multi-step Performance Ratio Prediction of Solar Power Plants Using Time Series ARIMA

Steven Bandong, Endra Joelianto, Edi Leksono, Ayu Purwarianti, and Irsyad N. Haq

Abstract— The Internet of Things is very useful for real-time data acquisition and field monitoring. The results can be analyzed to improve system reliability and efficiency. One object of the system that requires this is a renewable energy system such as a Solar Power Plant. Performance ratio is a measure of the performance of solar power plants. Its value will determine the increase in energy costs over time. It is therefore important that its value can be predicted in the future. However, because of its fluctuating value, it needs to be decomposed before the trend value is taken. The value of this trend will be predicted later. ARIMA is used as a method for predicting time series. One-step and multi-step predictions are applied and compared. The ARIMA prediction results are compared with other methods such as SVM and Multiple Linear Regression (MLR). Obtained by ARIMA, the prediction results are more accurate than SVM and MLR which are marked with the smallest RMSE value, 0.008 and the closest R^2 to 1, 0.98. One-step prediction also shows higher prediction accuracy compared to multi-step predictions.

Index Terms— Multi-Step Prediction, One-Step Prediction, Performance Ratio, Solar Power Plant

I. INTRODUCTION

THE growing technology called the Internet of Things (IoT) originated from the proposal advised by Ashton, a Professor from MIT Auto-ID Center in 1999. Currently, IoT has become the backbone for a combination of services from various applications. As a result, many electronic tools around us can communicate with each other via the internet. Fig. 1. depicts the IoT architecture which consists of 3 layers, namely perception layer, network layer, and application layer. This perception layer is in the form of sensors, actuators that can recognize or influenced by environmental conditions and are connected to the internet. The network layer is the part that

does communication between 'things'. Then the benefits received by humans are called application layers [1].

Compared to traditional industrial communications using internet to illustrate [2][3], IoT empowers 'things' to communicate and coordinate with one another. Because the number of 'things' is very large and varied, then it is necessary to have integrated access that allows communication to occur. This is a major problem in the IoT field [1]. Because of this important role, IoT has also been defined as one pillar in Industry 4.0 [4] [5].

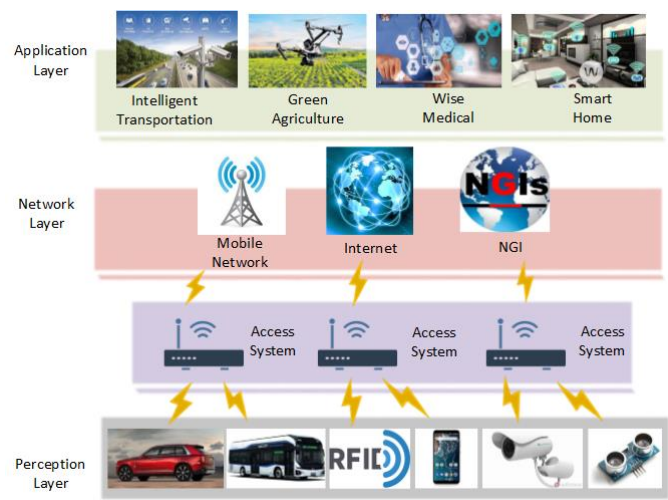


Fig. 1. General IoT architecture [1]

IoT has extensively used in many cases such as flood early warning detection [6], vehicle tracker [7], ambient environmental quality monitoring [8], water quality monitoring [9], and many others field. One system that really needs monitoring online is an energy system such as Solar Power Plants. Solar Power Plants need to be applied by IoT so that it can be known as soon as possible the results of the production of energy, current, voltage and others arrived. They are used for monitoring, management and prediction purposes. The acquisition data using the IoT system is then processed for future analysis purposes.

One case to be concerned, in the tropics the solar cell system has a low reliability compared to clemency region [10]. Therefore, it needs good management so that the solar power system can overcome this reliability problem. One of them is by predicting the decrease in the ratio of solar cell

Manuscript received November 11, 2019.

Steven Bandong is a research associate of Artificial Intelligence, Control and Automation Laboratory, Faculty of Industrial Technology, Institut Teknologi Bandung (e-mail: bandong.steven@gmail.com).

Endra Joelianto is with the Instrumentation and Control Research Group, Institut Teknologi Bandung (e-mail: ejoel@tf.itb.ac.id).

Edi Leksono is with the Engineering Physics Research Group, Institut Teknologi Bandung (e-mail: edi@tf.itb.ac.id).

Ayu Purwarianti is with the Informatics Research Group, Institut Teknologi Bandung (e-mail: ayu@stei.itb.ac.id).

Irsyad N. Haq is with the Engineering Physics Research Group, Institut Teknologi Bandung (e-mail: inhprop@gmail.com).

performance. The result of this prediction will give consideration and proposals for the needs of maintaining or replacing the system for increased reliability. In this paper, the prediction is done using Autoregressive Integrated Moving Average (ARIMA) time series method which has been used in many areas of research [11][12]. Its accuracy is then compared to Support Vector Machine (SVM) and Multiple Linear Regression (MLR) [13]-[17]. In other words, this paper will contribute on the application layer at Fig.1.

II. METHODS

The research is implemented by utilizing data collected from 2015-2018 by the IoT system of the solar power plant at Energy Management Laboratory, Institut Teknologi Bandung. The IoT architecture used to collect data is depicted in Fig. 2 [18]. Data of Battery system, photovoltaic array, and electricity load are collected by the local module or by IoT means and transferred to HMI for the display. They are also stored in the cloud system and can be used for further analysis. These data are then processed through 4 stages as shown in Fig. 3 to make performance ratio prediction.

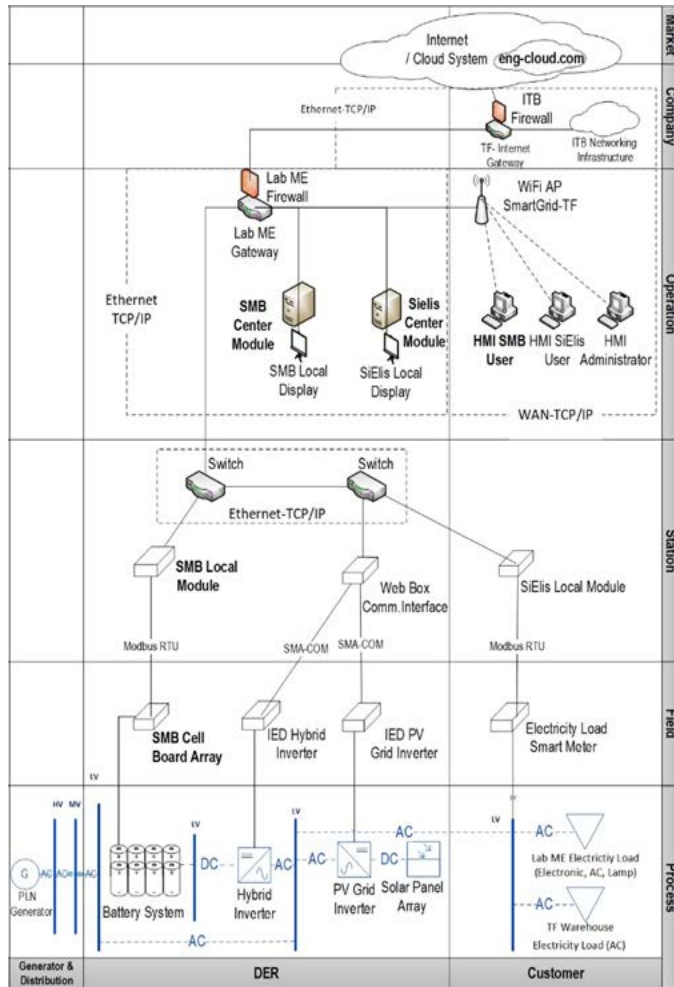


Fig. 2. IoT Architecture at the Energy Management Laboratory [18]

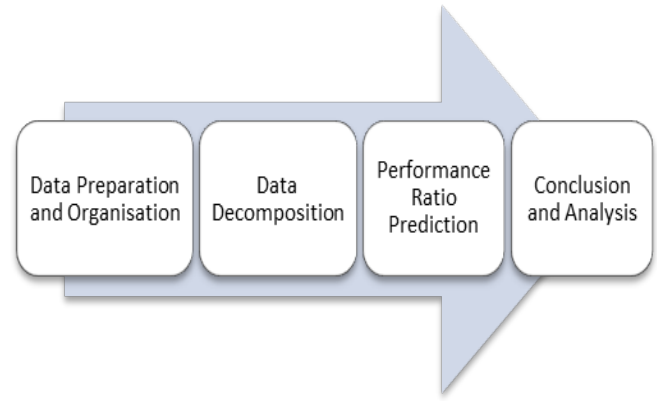


Fig. 3. Stages of research

A. Performance Ratio

The performance of a solar cell system over time can be assessed using the performance ratio. If the performance ratio of a solar cell system decreases then it means higher costs for energy use. The overall loss experienced by a system due to temperature, inefficiency, and component failure can be determined by Performance Ratio (**PR**) analysis.

$$PR = \frac{\sum_i EN_{AC_i}}{\sum_i \left[P_{STC} \left(\frac{G_{POA_i}}{G_{STC}} \right) \right]} \quad (1)$$

EN_{AC} is AC (Alternating Current) power that generated from solar cell system (W), P_{STC} is the total of modules installed power (W), G_{POA} is Irradiance from sun that hits the module (W/m^2) and G_{STC} is Irradiance under standard test conditions ($10000 W/m^2$) [19]. When calculating the performance ratio using (1), there were some missing values of electrical variable such as voltage, current and AC (Alternating Current) energy produced from the PV system. To overcome, the estimation of performance ratio is done according a method in [20] by using environmental data and applying Principal Component Analysis to extract the important feature from data and Support Vector Machine as machine learning to make prediction (PCA-SVM).

B. Performance Ratio Prediction

The calculation is made in one-step and multi-step. One-step predictions are a prediction for one period in the future. The periods are varied into one day, one week, two weeks, three weeks and one month, so that the prediction models are obtained up to one month or more. Multi-step predictions are predictions for several periods in the future so that once a prediction is made, the obtained results are several performance ratio data in the future according to the steps given in that period. For example, in multi-step predictions for one week, the number of steps used is four, so the predictions obtained up to the next month. One-step predictions have been used in [21][22] and multi-step predictions have been applied in [23][24]. The stages of prediction can be seen in Fig. 4.

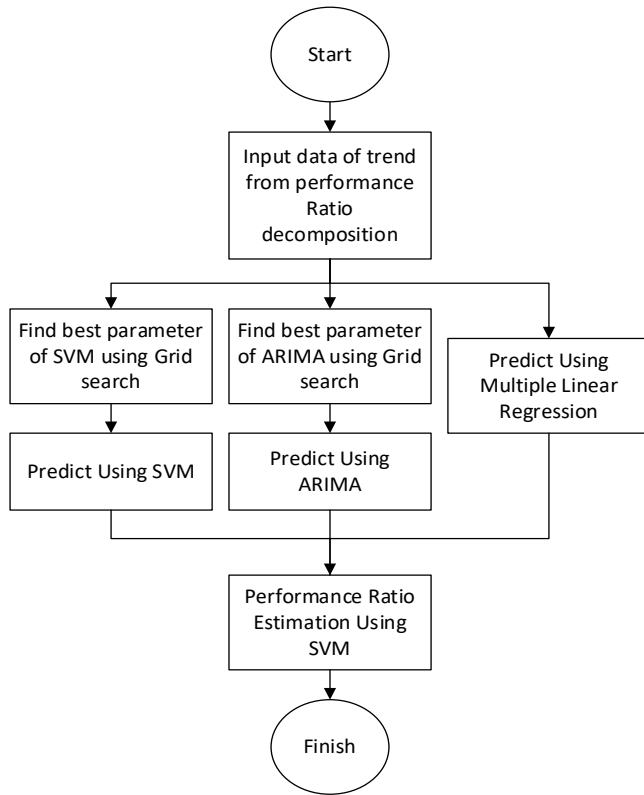


Fig. 4. Predictions flowchart

In this paper, the ARIMA time series prediction method is used to predict performance ratios in one-step and multi-step.

1. ARIMA

The ARIMA model is generally used because it is quite robust and easy to implement. ARIMA (p, d, q) can be explained as follows:

- p is the degree of autoregression (AR) of the model, and the number of the previous measurement series used for prediction.
- d is the differential degree (I) to make the model stationary.
- q is the degree of moving average (MA) of the model and the number of previous prediction errors that used.

Mathematically, the ARIMA (p, d, q) model can be expressed as:

$$y_t = c_t + \sum_{i=1}^p \phi_m y_{t-i} + \sum_{j=1}^q \theta_n e_{t-j} \quad (2)$$

where c_t is a constant and represents the average time series value in question, ϕ_m is the m -th coefficient of the autoregression parameter, θ_n is the n -th coefficient of the moving average parameter, y_t is the result of prediction on t . The notation y_{t-i} is the value of y in $(t-i)$ and e_{t-j} is the

prediction error when compared to the actual value in $(t-j)$ [25].

Machines learning, in this study SVM and multiple linear regression, are implemented as comparisons to ARIMA.

2. Support Vector Machine (SVM)

According to [26], suppose $D(x, y)$ is a given set of n input and output data pairs that need to be processed. It can be written as $D(x, y) = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \in R^n \times R$. The sets of these paired input and output data are displayed as $X = \{x_1, x_2, \dots, x_n\}$, $Y = \{y_1, y_2, \dots, y_n\}$ respectively. We assume $F = \{f \mid f(x) = \omega^T \cdot x + b, \omega \in R^n\}$ is the regression function. The main goal of regression using SVM is to achieve the optimal value ω . This value is then used to form an optimum hyperplane in the case of data that can be linearly modelled. In the nonlinear case, a non-linear mapping to a higher dimension is required so that linear regression can be carried out on this dimension. The data can most likely be separated linearly at these higher dimensions. This optimization task then can be viewed as a minimization of:

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (3)$$

which fulfills:

$$y_i(\omega \cdot \phi(x_i) + b) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0, i = 1, 2, 3, \dots, C > 0 \quad (4)$$

Consecutively, ε , C and (ξ_i, ξ_i^*) define the precision parameter, the constant, and slack variables that determine the balance between the flatness of f and tolerance of ε . To be able to predict non-linear data, the kernel is applied to SVM. The kernel $K(x_i, x_j)$ such as Linear, Sigmoid, Radial Basis Function (RBF) and Polynomial are commonly used to replace the dot product $\phi(x_i)$ in higher dimensions [26]. In this paper, RBF is used to carry out this task. In some combinations of parameters, RBF is considered to be linear and sigmoid, resulting in higher accuracy [27].

3. Multiple Linear Regression (MLR)

MLR is commonly used to compare the influence of several independent variables on the dependent variable. The relationship between the independent variables and the predicted variables is assumed to be linear. Although simple but this method has been used widely and gives quite good results. Multiple Linear Regression is expressed by the following equation [28]:

$$Y = a + b_1 X_1 + b_2 X_2 + \dots + b_n X_n \quad (5)$$

Y is the target or dependent variable, a is a constant, $b_1, b_2, \dots, b_n$ are regression coefficient, and $X_1, X_2, \dots, X_n$ are independent variables

III. RESULTS AND DISCUSSIONS

A. Data Decomposition

Data decomposition is done to estimate the trend value in the data. This is done to remove noise and seasonal data from the data. In addition, trend data are more predictable and

produce high accuracy. Trend data also represent the movement of the average performance ratio because this is the moving average value. In this process, the moving average is made every three data so that the trend data represent the original data properly. The decomposition results can be seen in Fig. 5.

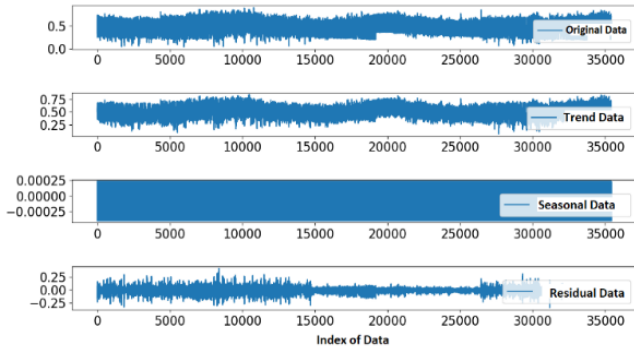


Fig. 5. Decomposition of Performance Ratio Data

B. Performance Ratio Prediction

Predictions are made using ARIMA which the prediction results accuracy is then compared with SVM and MLR. In order to obtain optimal prediction results, optimization of the parameters p, d, q is done by applying Grid Search [29]. For the comparison, SVM parameter optimization is done except multiple linear regression (MLR) because it does not have parameters that need to be optimized. One example of optimization of ARIMA and SVM parameters in the 2 week prediction period can be seen in Fig. 6 and 7. The optimal ARIMA parameters p, d, q are 0,1,2. The obtained optimal parameters SVM are $C = 82$, $\gamma = 0.25$ and $\varepsilon = 0.0008$.

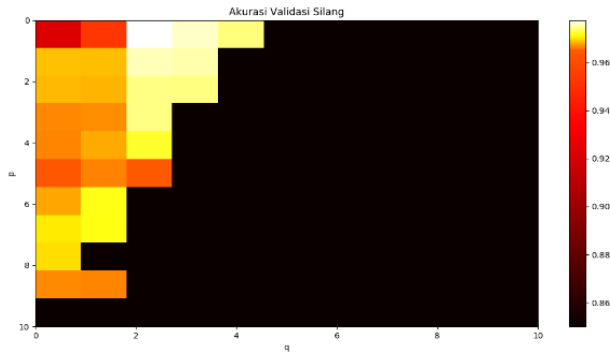


Fig. 6. ARIMA parameters optimization

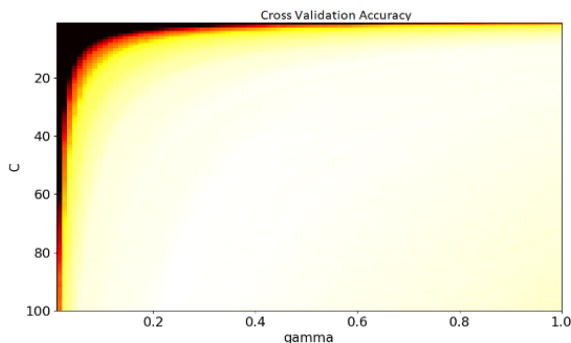
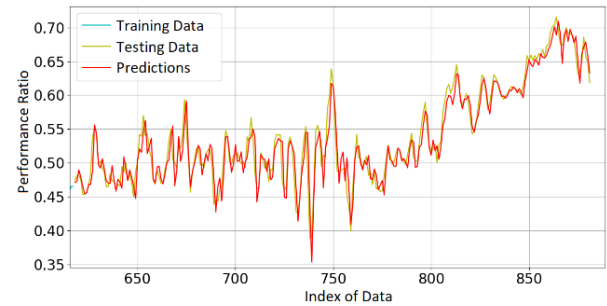


Fig. 7. SVM parameter optimization

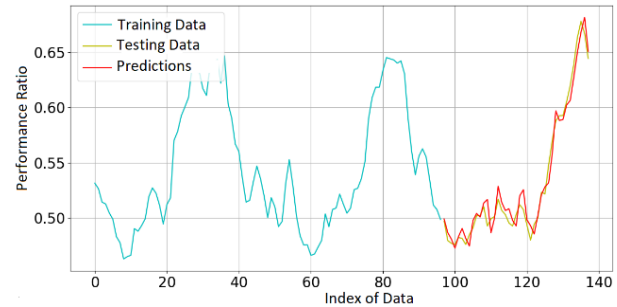
C. One-Step Prediction

One-step predictions were performed using ARIMA and Fig. 8 shows the results. ARIMA shows good prediction results because the prediction results follow the patterns in the test data very well in all prediction periods. This is indicated by the high R^2 and r values of more than 0.93 as shown in Table 1. R^2 and r show how well the prediction results follow the patterns contained in the test data. The closer the value to 1 indicates the better the ability of prediction results to follow the test data pattern. In addition, the RMSE value for each prediction period is also very small at a maximum of 0.02. This indicates that one-step ARIMA has been able to predict the value of performance ratio.

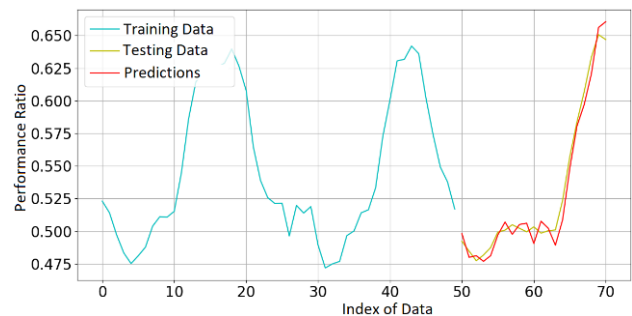
ARIMA also gives more accurate predictions than SVM and MLR. ARIMA RMSE value is smaller than SVM and MLR in each prediction period so that the ARIMA prediction results are closer to the true value. The value of R^2 and r of ARIMA is also the closest to 1 compared to SVM and MLR so that the results of ARIMA predictions are better in following the pattern of test data.



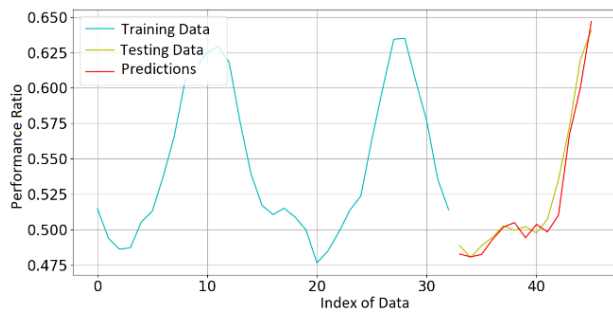
a. One day prediction



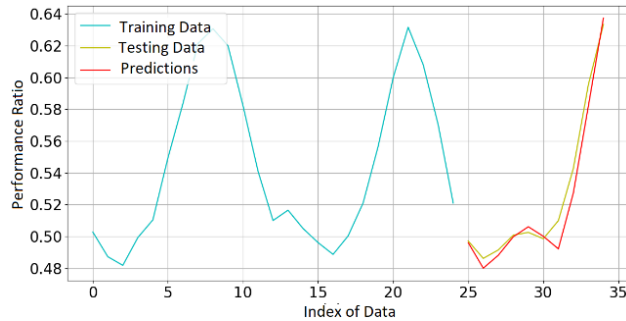
b. One week prediction



c. Two weeks prediction



d. Three weeks prediction



e. One-month prediction

Fig. 8. Prediction of the performance ratio using ARIMA one-step

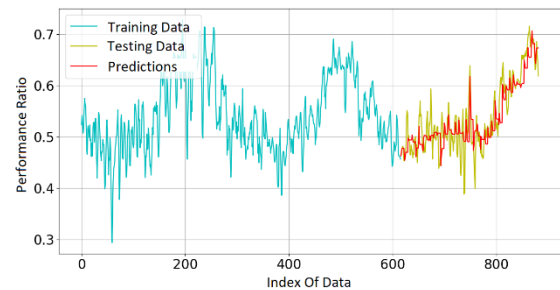
TABLE I
COMPARISON OF ONE-STEP PREDICTION ACCURACY

Methods	Accuracy	One-step prediction				
		1 day	1 week	2 weeks	3 weeks	4 weeks
ARIMA	RMSE	0.018	0.009	0.008	0.01	0.009
	R^2	0.93	0.97	0.98	0.96	0.96
	r	0.96	0.99	0.99	0.98	0.99
SVM	RMSE	0.023	0.011	0.009	0.01	0.017
	R^2	0.89	0.96	0.97	0.94	0.87
	r	0.95	0.98	0.98	0.97	0.94
MLR	RMSE	0.024	0.011	0.009	0.01	0.02
	R^2	0.88	0.96	0.97	0.94	0.86
	r	0.94	0.98	0.98	0.97	0.95

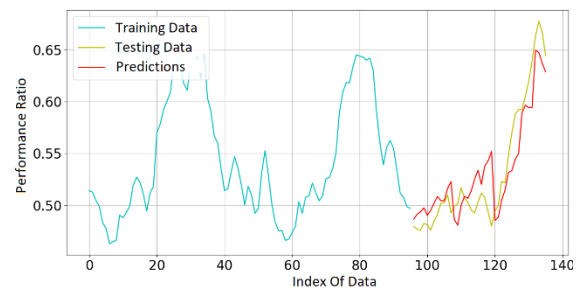
D. Multi-Step Prediction

Daily data prediction is made into 7 steps ahead to predict up to one week in the future. The predictions are also made in the next 4 steps using weekly data to obtain prediction results up to one month in the future. Fig. 9 shows the prediction results. Table 2 describes the prediction accuracy using ARIMA, SVM and MLR. ARIMA shows a fairly good prediction result because the predictions results follow the pattern and quite precise to the test data as seen in Fig.9. This is quantified by the R^2 , r and RMSE values. In the prediction of 1 day with 7 steps and 1 week with 4 steps, the value of R^2 and r of ARIMA is very close to 1 which means that the pattern in the test data can be well followed by the prediction results. In addition, the RMSE value in these two prediction periods is quite small, which means the prediction results are very close to the test data. So, it can be concluded that ARIMA can give good multi-step prediction of performance ratio data.

Multi-step predictions are also made using SVM and MLR. The accuracies of the predictions are then compared. ARIMA is obtained showing more accurate prediction results. This can be seen from the ARIMA r and R^2 values which are higher than those produced by SVM and MLR. This means that the ARIMA prediction results are more able to follow the pattern in the test data. In addition, ARIMA RMSE value is smaller which means the prediction results are closer to the true value.



(a) Daily prediction with 7 steps



(b) Weekly prediction with 4 steps

Fig. 9. Predictions of performance ratio using multi-step ARIMA

TABLE II
COMPARISON OF MULTI-STEP PREDICTION ACCURACY

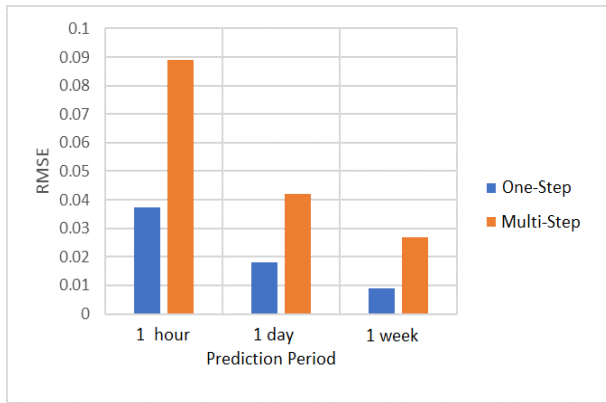
Methods	Accuracy	Multi-step prediction	
		1 day 7 steps	1 week 4 steps
ARIMA	RMSE	0.018	0.009
	R^2	0.93	0.97
	r	0.96	0.99
SVM	RMSE	0.023	0.011
	R^2	0.89	0.96
	r	0.95	0.98
MLR	RMSE	0.024	0.011
	R^2	0.88	0.96
	r	0.94	0.98

Predictions were made in several time ranges, namely predictions for one day, one week, two weeks, three weeks, and four weeks. To optimize SVM hyperparameter, grid search is implemented. For example, for the two weeks prediction, the accuracy grid is obtained as shown in Fig.7 after testing all possible combinations of parameter value ranges.

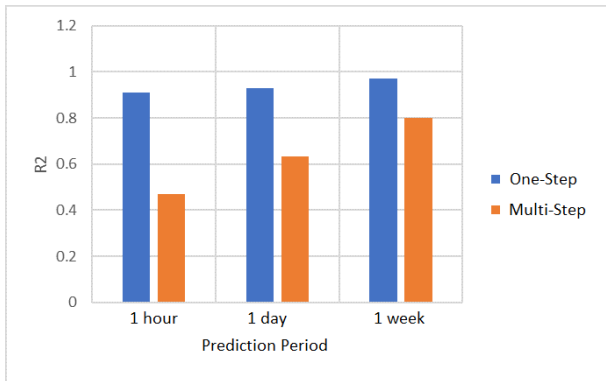
E. Comparison of One-Step and Multi-Step Prediction Results

The results of one-step and multi-step predictions are then compared in Fig. 10. One-step prediction results show a smaller RMSE value and a higher R^2 . This is due to the accumulation of errors in multi-step method. This method uses

the results of the previous prediction to make the next prediction so that the prediction error will increase with the number of steps made. However, multi-step R^2 is getting closer to one-step because the increasing base period means the pattern that must be followed decreases so that it will be easier to follow. However, the multi-step method has advantages because its prediction results for a month also show the changes on every day. This is different from one-step which only displays prediction results in the next month alone. Overall, based on RMSE and R^2 , it was found that the one-step method provides better prediction results.



a. RMSE Comparison



b. R^2 Comparison

Fig. 10. Comparison of One-Step and Multi-Step Prediction Results

The results of this prediction can be expressed in equations especially for the ARIMA and MLR methods. For example, in predictions for the next one week, ARIMA equations that refer to (2) can be expressed as:

$$y_t = 0.5410 + 1.4771y_{t-1} - 0.1199y_{t-2} - 0.3807y_{t-3} + 0.0099e_{t-1} - 0.0101e_{t-2} - 0.9998e_{t-3} \quad (6)$$

Using the MLR method for predictions one week ahead based on (5), the following equation is obtained:

$$y_t = 0.0286 + 1.2536y_{t-1} - 0.1457y_{t-2} - 0.1599y_{t-3} \quad (7)$$

From these two equations, the value of the performance ratio of the previous one week is the most influential compared to the previous second and third values. This is

evident from the coefficient of y_{t-1} which is greater than the others. ARIMA clearly has advantages over MLR because in its predictions it also takes into account the error of the results of the previous prediction.

IV. CONCLUSION

A method to accurately predict the ratio of solar cell performance was considered in the paper. Predictions were made by estimating the trend value from the moving average decomposition data. After that, a time series prediction model was made using ARIMA on one-step and multi-step. One-step and multi-step prediction gave high accuracy which shown by small RMSE. The results of prediction were also capable to follow the pattern of original data shown by both R^2 and r close to one. One-step prediction gave better prediction accuracy than multi-step. ARIMA resulted in better prediction accuracy than SVM and MLR in both one-step and multi-step prediction modes. These were marked by smaller RMSE value and higher R^2 .

REFERENCES

- [1] S. Wang, Y. Hou, F. Gao, and X. Ji, "A novel IoT access architecture for vehicle monitoring system," in *IEEE 3rd World Forum on Internet of Things (WF-IoT)*, pp. 639-642, 2016.
- [2] E. Joelianto and Hosana, "Performance of an industrial data communication protocol on ethernet network," in *2008 5th IFIP International Conference on Wireless and Optical Communications Networks (WOCN'08)*, IEEE, Surabaya, Indonesia, 2008, pp. 1-5.
- [3] E. Joelianto and Hosana, "Loop-back action latency performance of an industrial data communication protocol on a PLC ethernet network," *Internetworking Indonesia Journal*, vol. 1(1), pp.11-18, 2009.
- [4] V. Pilloni, "How data will transform industrial processes: Crowdsensing, crowdsourcing and big data as pillars of industry 4.0," *Future Internet*, vol. 10(3), p.24, 2018.
- [5] V.L. Silva, J.L. Kovaleski, and R.N. Pagani, "Technology Transfer and Human Capital in the Industrial 4.0 Scenario: A Theoretical Study," *Future Studies Research Journal: Trends and Strategies*, vol. 11(1), pp. 102-122, 2019.
- [6] J.W. Simatupang and F. Naufal, "Flood Early Warning Detection System Prototype Based on IoT Network," *Internetworking Indonesia Journal*, vol. 11(1), pp 17-22, 2019.
- [7] S. Liawatimena and J. Linggarjati, "Vehicle Tracker with a GPS and Accelerometer Sensor System in Jakarta," *Internetworking Indonesia Journal*, vol. 9(2), pp. 10-15, 2017.
- [8] A. Djajadi and M. Wijanarko, "Ambient Environmental Quality Monitoring Using IoT Sensor Network," *Internetworking Indonesia Journal* vol. 8(1), pp.41-47, 2016.
- [9] T. Yuwono, L. Hakim, and I. Ardi, Umar, "The Application of Internet of Things System for Water Quality Monitoring," *Internetworking Indonesia Journal*, vol. 8(1), pp. 49-53, 2016.
- [10] O.O. Osarumen, H. Emeka, Amalu, N.N. Ekere, and P.O. Olagbegi, "A review of photovoltaic module technologies for increased performance in tropical climate," *Renewable and Sustainable Energy Reviews*, vol. 75, pp.1225-1238, 2017.
- [11] M. Rhanoui, Y. Siham, M. Mounia, and M. Hajar, "Forecasting Financial Budget Time Series: ARIMA Random Walk vs LSTM Neural Network," *IAES International Journal of Artificial Intelligence*, vol. 8(4), 2019.
- [12] M. Matyjaszek, P.F. Riesgo, A. Krzemień, K. Wodarski, and G.F. Fidalgo, "Forecasting coking coal prices by means of ARIMA models and neural networks, considering the transgenic time series theory," *Resources Policy*, vol. 61, pp 283-292, 2019.
- [13] S. Barhmi, and O.E. Fatni, "Hourly wind speed forecasting based on Support Vector Machine and Artificial Neural Networks," *IAES International Journal of Artificial Intelligence*, vol. 8(3), p. 286, 2019.

- [14] J. Li, J. K. Ward, J. Tong, L. Collins, and G. Platt, "Machine learning for solar irradiance forecasting of photovoltaic system," *Renewable Energy*, vol. 90, pp. 542–553, 2016.
- [15] JS. Chou and DS. Tran, "Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders," *Energy*, vol. 165 pp. 709-726, 2018.
- [16] M. Bouzardoum, A. Mellit, and A.M. Pavan, "A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant," *Solar Energy*, vol. 98, pp. 226-235, 2013.
- [17] M.Q. Raza, M. Nadarajah, and C. Ekanayake, "On recent advances in PV output power forecast," *Solar Energy*, vol. 136, pp.125-144, 2016.
- [18] I.N. Haq, "Development of Smart Battery Management System Architecture Model for Energy Storage Condition Monitoring and Improvement," Engineering Physics, Doctoral Dissertation, Institut Teknologi Bandung, 2019.
- [19] T. Dierauf, A. Growitz, S. Kurtz, J.L. Cruz, E. Riley, and C. Hansen, "Weather-Corrected Performance Ratio," Energy Lab.(NREL), Golden, CO (United States), 2013 Apr 1.
- [20] S. Bandong, E. Leksono, A. Purwarianti, and E. Joelianto, "Performance Ratio Estimation and Prediction of Solar Power Plants Using Machine Learning to Improve Energy Reliability," In *2019 6th International Conference on Instrumentation, Control, and Automation (ICA)*, pp. 36-41. IEEE, 2019.
- [21] B.S. Yang, M.S. Oh, and A.C.C. Tan, "Machine condition prognosis based on regression trees and one-step-ahead prediction," *Mechanical Systems and Signal Processing*, vol. 22(5), pp. 1179-1193, 2008.
- [22] G. Singh, M. R. Mukerjee, and N. Sharma, "Multinode Load Forecasting Using Timeseries," *IFAC Proceedings* vol. 12(5), pp. 288-294, 1979.
- [23] R. Ye and D. Qun, "MultiTL-KELM: A multi-task learning algorithm for multi-step-ahead time series prediction," *Applied Soft Computing*, vol. 79, pp. 227-253, 2019.
- [24] J. Wang and L. Yaning, "Multi-step ahead wind speed prediction based on optimal feature extraction, long short term memory neural network and error correction strategy," *Applied Energy*, vol. 230, pp. 429-443, 2018.
- [25] S.N. Singh and A. Mohapatra, "Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting," *Renewable Energy*, vol. 136, pp. 758-768, 2019.
- [26] H.S. Jang, K.Y. Bae, H.S. Park, and D.K. Sung, "Solar power prediction based on satellite images and support vector machine," *IEEE Transactions on Sustainable Energy*, vol. 7(3), pp. 1255-1263, 2016.
- [27] C.W. Hsu and C.J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE transactions on Neural Networks*, vol.13(2), pp. 415-425, 2002.
- [28] Y. Lee, C. Jung, and S. Kim, "Spatial distribution of soil moisture estimates using a multiple linear regression model and Korean geostationary satellite (COMS) data," *Agricultural water management*, vol. 213, pp. 580-593, 2019.
- [29] I. Syarif, A. Prugel-Bennett, and G. Wills, "SVM parameter optimization using grid search and genetic algorithm to improve classification performance," *Telkomnika*, vol. 14(4), p. 1502, 2016.