Evaluation of ECMWF model to predict daily and monthly solar radiation over Indonesia region

Y Sianturi, A Sopaheluwakan and K A Sartika

Indonesian Agency for Meteorology, Climatology, and Geophysics, Jalan Angkasa I No.2, Jakarta Pusat, Indonesia

yculina@gmail.com

Abstract. Solar radiation forecast is a pivotal information needed in the operational activity of large-scale solar energy production. In this study, the reliability of SSRD (surface solar radiation downward) forecast from the 51 ensemble members in the ECMWF (European Centre for Medium Range Forecast) long-range forecast to predict daily and monthly radiation in 5 climatological stations in Indonesia is evaluated. The global horizontal irradiance (GHI) data from the solar radiation observation network from January 2018 – December 2020 are used in the quantitative evaluation of the SSRD forecast. Post-processing methods are applied to the model output, namely the bilinear interpolation method and the empirical quantile mapping to reduce consistent biases in the model output. The evaluation was carried out for different cloud covers based on the calculation of clearness index (k_t). The cloud condition affects the performance of the model, where the highest correlation value is achieved during sunny days (0.18 – 0.65) and the lowest correlation happens in overcast days (0.05 – 0.35). Models also tend to underestimate radiation when the sky is clear and overestimate it in cloudy days, based on negative MBE values during clear days (-0.47 kWh/m2 – -1.29 kWh/m2). The spatial averaging method did not necessarily improve the accuracy of the forecast, but the empirical quantile mapping method provides better accuracy, which is indicated by α values (mean error ratio) lower than 1 in most stations. Information about the influence of cloud cover on model performance can be used in future application of the model output and the bias correction process carried out in this study can be applied to reduce bias in the model.

1. Introduction

Given the increasing demand for renewable energy, the need for an accurate forecast of meteorological parameters is also rising. In solar energy projects, one of the most important things needed to improve the efficiency in the solar energy production activities is a reliable solar radiation forecast. Compared to other energy sources, the energy input in solar power plants tends to fluctuate more, and it might cause potential externalities and inefficiencies in the power system. Solar radiation forecasts on different time scales can be utilized in making the operational plan, to reduce the possibility of these potential losses. With the increasing prospect of solar energy as one of the promising renewable energy sources in Indonesia, the information of solar radiation forecasts to support the development of solar energy projects in this region will be even more important. In solar energy operational activities, short-term radiation forecast – or also known as nowcasting – is needed in short-term operational planning, while longer-term forecast, from daily to monthly, is also needed to make long-term planning [1].
Solar radiation forecasts are generally produced using different tools and methods, from ground-observed meteorological variables, satellite data, to numerical weather prediction (NWP) model output. Statistical models that utilized station data and satellite observations show good performance in predicting short-term irradiance, for different temporal ranges from minutes to hours [2–5]. For a longer-term forecast, however, the output of numerical models were found to outperform the forecast generated using statistical models and satellite data. The evaluation of NWP model output are done in different regions in the world with different forecast horizon and intervals, ranges from hourly forecast to monthly accumulation forecast. Perez [2] used the hourly solar forecast from the National Oceanic and Atmospheric Administration’s (NOAA) global forecasting system (GFS) model up to 6 days ahead to predict the short and medium term solar radiation forecast. Perez et al [6] also compared global horizontal irradiance (GHI) forecasts from several numerical models, including ECMWF and GFS-based Weather Research and Forecasting (WRF) model for US, Europe, and Canada region. In those studies, the solar radiation forecast provided by ECMWF model tends to outperform the persistence model and other NWP models in most locations. The ECMWF model is one of the models with best performance world-wide [7] and currently is used as the main numerical model to predict most of the meteorological parameters, including the solar irradiance, by the national meteorological institution in Indonesia, the Agency for Meteorology, Climatology, and Geophysics (BMKG).

One of the earliest evaluations of ECMWF’s solar irradiance forecast is documented in Bofinger [8], that used the improved solar radiation forecast from ECMWF to predict hourly to daily solar energy output in Germany with encouraging results. In a more recent study, Thorarinsdottir et al. [9] used the ensemble forecast from ECMWF that consists of 51 members to make probabilistic estimation of solar energy from hours to several days ahead. The ECMWF models were also used to forecast direct irradiance [10–12], even though forecasting the direct component of solar radiation with numerical models was found to be less reliable compared to forecasting global radiation, as direct radiation forecast is more sensitive to the ability of the model to simulate cloud cover. In Indonesia region, the evaluation of the surface solar radiation downward from ECMWF’s reanalysis product that combines observation data and past weather forecasts, ERA5 reanalysis, was done in Sianturi [13], and the biases in the estimates were assumed to originate from the less well-represented cloud processes in the model. For a region with strong cloud formation as Indonesia, this might contribute to a higher uncertainty in solar radiation forecasting.

Post-processing methods are often applied to the model output to remove continuous and systemic bias in the model. Model output statistics and inverse distance weighting (IDW) are done in Bofinger [8] to overcome the coarse grid resolution of numerical weather model. Other post processing methods, such as artificial neural network and spatial averaging method also used to reduce systematic bias in the model in Lauret [14]. Some studies also used the information of model errors in different cloud cover and solar zenith angles to bias correct the model output and it significantly reduce the mean error in the model [11][15]. In those studies, it was shown that the post-processing methods contribute to improve the accuracy of the forecasts, even with different degrees of improvement. The spatial averaging method, for instance, was better when it combines a smaller number of grid than averages over a greater area [14–15].

Studies on solar radiation forecasts in the Indonesia region are scarce, as solar radiation observation network and solar radiation forecast products in Indonesia are still very limited. This study aims to assess the reliability of ECMWF’s solar radiation forecast in Indonesia region and evaluate the application of a few post-processing methods to the model output. Evaluations based on cloud cover profiles, utilizing the calculated clearness index values, will also be done to identify the influence of clouds to forecast accuracy. Two post-processing methods are proposed within this study, a spatial averaging method namely bilinear interpolation method and bias-correction method using the CDFs of both observation data and model output, the empirical quantile mapping method. The paper will be organized as follow: data and methods used are described in Section 2, Section 3 includes the results and discussions of the results, and the conclusion will be given in Section 4.
2. Data and Methods

2.1 ECMWF’s Surface Solar Radiation Downward
The radiation model used in the ECMWF forecast was described in Trocolli and Morcrette [11]. In the ECMWF forecast, the GHI value is denoted as the surface solar radiation downward (SSRD). The variable is constructed as integrated values of global irradiance in the unit of J/m$^2$ every 6-hour period, from the beginning of the initial forecast up to 7 months ahead. In this evaluation study, the accumulated irradiance values will be converted into daily and monthly accumulation of solar radiation. The forecast used in this study is the reforecast, which means the corrected version of the running forecast by removing the continuous, persistent errors from the model. There are 51 members in this NWP product and in some parts of the evaluation, the ensemble mean of all the members is treated as the forecast value.

2.2 Ground Observation of Solar Radiation
The data from solar radiation observation network is used in this study to evaluate the SSRD forecast. The solar radiation observation network that is managed by BMKG, consists of 27 climatological stations in Indonesia region, which every station measures several types of radiation including global (GHI), diffuse (diffuse horizontal irradiance/ DHI), and direct radiation (direct normal irradiance/ DNI). Due to the length of data record and data quality, only five stations are used in this evaluation study. Prior to the analysis, a quality control procedure is applied to the GHI data based on the physically possible limits and extremely rare limits, as was described in Sianturi [13]. The automatic observation is done in 10-minutes interval, and from the instantaneous values, the hourly value is obtained by averaging all the observed values in that hour. The daily value is then calculated by integrating all the hourly averages in one day to generate a quantity in kWh/m$^2$. The same integrating method is also applied to the ECMWF model output, to get the daily and monthly forecast of solar radiation.

![Figure 1. Station locations.](image)

2.3 Evaluation Metrics
The evaluation of the radiation forecast will be done in daily and monthly basis. To quantify the evaluation results, other than data visualization, some statistical metrics will also be used in the analysis. The root means square error (RMSE), mean absolute error (MAE), and mean bias error (MBE) values will be used to assess the accuracy of the forecast. The Pearson correlation coefficient will be used to measure the strength of the relationship between the observed and modelled radiation. Two new metrics are defined to assess the ability of the bias-correction method to reduce biases in the model output, that are denoted by $\alpha$ and $\beta$, that represent mean error ratio and standard deviation ratio:
\[
\alpha = \frac{x_{corr} - Obs}{|x_{mod} - Obs|} \\
\beta = \frac{\sigma_{corr}}{\sigma_{mod}}
\]  

Each variable indicates the following quantity: \(x_{corr}\) is the ensemble average of the corrected forecast, \(x_{mod}\) is the ensemble average of uncorrected forecast, \(Obs\) is the observed radiation, \(\sigma_{corr}\) is the standard deviation of the corrected forecast (spread of the 51 members), and \(\sigma_{mod}\) is the standard deviation of the uncorrected forecast. Improvement that is provided by the bias correction method will be indicated by the value of \(\alpha\) and \(\beta\) to be less than 1 (lower error and lower spread in the corrected forecast).

The assessment of the impact of cloudiness on the solar radiation forecast for Indonesia region is crucial, considering the fact that strong cloud formation is one of the main features of Indonesia’s climatic condition. Cloud cover condition is often used to determine the model performance. One of the most used metrics is the clearness index, because not only representing the cloud cover, but also the aerosol concentration in the atmosphere. In the first part of the evaluation, the error based on the sky condition will be assessed, to see how cloudiness affect the forecast performance. The sky condition will be classified using the daily clearness index \((k_t)\). Clearness index is the fraction of the actual solar radiation that reach the earth surface over the solar radiation that would be received in a completely clear sky [16]. The daily clearness index in each station is calculated using these formula:

\[
k_t = \frac{S}{S_0}
\]

\[
S_0 = \frac{24}{\pi} I_{sc} \left(1 + 0.033 \cos \frac{360D}{365} \right) \left(\cos \varphi \cos \delta \sin \omega_s + \frac{2\pi \omega_s}{360} \sin \varphi \sin \delta \right)
\]

\[
\omega_s = \cos^{-1}(\tan \varphi \tan \delta)
\]

\[
\delta = 23.45 \left(360 \frac{284 + D}{365} \right)
\]

\(k_t\) = daily clearness index  
\(S\) = observed daily solar radiation (in Wh/m²)  
\(S_0\) = daily solar radiation at the top of the atmosphere (in Wh/m²)  
\(I_{sc}\) = solar constant (1367 Wh/m²)  
\(D\) = order of days in a year (Jan 1 = 1, Jan 2 = 2)  
\(\varphi\) = solar declination angle  
\(\omega_s\) = local hour angle

In this study, the clearness index will be divided into three categories: the clear sky condition \((k_t > 0.65)\), cloudy condition \((0.4 < k_t < 0.65)\), and overcast condition \((k_t < 0.4)\).

2.4 Post-processing Methods
Two kinds of post-processing methods are used to remove systematic biases in the forecast. The first method, bilinear interpolation, is the process of resampling the model output using the values of the four closest grids using respective distances act as weights [15]. With \(P\) as the value in the station, and \(Q_{ij}\)
as the values in the four closest grids with \((x_i, y_i)\) as the respective coordinate pairs. The formula can be written as follow:

\[
P = \frac{1}{(x_2 - x_1)(y_2 - y_1)} \left[ x_2 - x \frac{Q_{11} Q_{12} (y_2 - y)}{Q_{21} Q_{22} (y - y_1)} \right]
\]

The second post-processing method is the empirical quantile mapping, which utilizes the information of the probability distribution of observation data and model output (hindcast) in the training period to construct a transfer function that can be used to bias-correct the forecast based on its probability distribution [17]. In this study, the transfer function is designed for each month of the year, to modify the forecasts in the respective month in the future period. Here, the observed daily solar radiation and model output in the training period (January – December 2018) will be used to construct transfer functions to correct the forecasts that are generated in January – July 2019.

![Figure 2. Illustration of the bias correction method. The solid black line indicates the correction factor between the ground-observed solar radiation (blue CDF) and model output (red CDF), at the cumulative probability value that is shown by the dashed black arrow.](image)

3. Results and Discussions

3.1 Forecast biases under different cloud covers

The clearness index in the five stations is calculated for January 2018 – December 2019 period. The maximum value of clearness index in all five stations ranges from 0.66 – 0.72 that indicates the general cloudy condition over these regions. In Jambi and Palembang stations, there is not even a single day where clearness index value exceeds the threshold for clear sky condition (0.65), as can be seen in Table 1. In this part, the forecast value used is the ensemble average of daily accumulated solar radiation from all 51 members. In all stations, the correlation value decreases consistently with increasing cloud cover, which correlation between model output and observed radiation is strongest when the sky is clear (0.18 – 0.65) and weakest during overcast condition (0.05 – 0.35). The forecast error, on the other hand, tends to be lower during cloudy days compared to clear and overcast days.

From the MBE values, it can be seen that the model tends to underestimate the daily radiation when the cloud cover is minimum, as negative values of MBE are produced in the evaluation during clear days ((-0.47) – (-1.29)). This trend is in line with the results of the evaluation of the ECMWF reanalysis data which was built using the same model in the Indonesian region in Sianturi [13]. In that evaluation, the reanalysis dataset underestimate radiation on clear sky condition and overestimate the radiation in
dark days. This is an important note in the application of SSRD ECMWF’s forecasts in the Indonesian region, given the relatively strong cloud formation and the fewer number of sunny days. In future implementation of the forecast in the solar energy sector, it should be kept in mind that there is a possibility that the forecast confidence level might be lower in the wet months.

### Table 1. Evaluation results under different cloud covers.

| Station   | Sky Condition | Correlation Coefficient | RMSE (kWh/m²) | MAE (kWh/m²) | MBE (kWh/m²) |
|-----------|---------------|--------------------------|---------------|--------------|--------------|
| Banjarbaru| Clear         | 0.65                     | 1.18          | 0.97         | -0.97        |
|           | Cloudy        | 0.39                     | 0.77          | 0.65         | 0.04         |
|           | Overcast      | 0.28                     | 1.83          | 1.62         | 1.61         |
| Bengkulu  | Clear         | 0.18                     | 1.43          | 1.29         | -1.29        |
|           | Cloudy        | 0.10                     | 0.76          | 0.61         | -1.15        |
|           | Overcast      | -1.13                    | 2.01          | 1.82         | 1.81         |
| Jambi     | Clear         | -                        | -             | -            | -            |
|           | Cloudy        | 0.26                     | 0.76          | 0.62         | 0.01         |
|           | Overcast      | 0.05                     | 1.73          | 1.47         | 1.44         |
| Jembrana  | Clear         | 0.65                     | 0.69          | 0.51         | -0.47        |
|           | Cloudy        | 0.60                     | 0.85          | 0.68         | 0.46         |
|           | Overcast      | 0.35                     | 2.40          | 2.28         | 2.28         |
| Palembang | Clear         | -                        | -             | -            | -            |
|           | Cloudy        | 0.30                     | 0.79          | 0.63         | 0.21         |
|           | Overcast      | 0.30                     | 1.69          | 1.54         | 1.54         |

#### 3.2 Forecast biases under different cloud covers

In this paper, all evaluation metrics are calculated for both the uncorrected ensemble average and the spatially-interpolated values, and some of the results are shown in Figure 3. The original value that is used as comparison is the value generated in the closest grid from the station. With the application of spatial interpolation method, it is expected that the biases caused by the topographical variations within the grid can be reduced. Figure 3 reveals that the interpolation method did not provide consistent improvement to the uncorrected forecast, as most of the correlation values and error values do not significantly vary between the original and interpolated values. In almost all stations, there is always at least one month when the interpolated forecast shows higher error or lower correlation, even though there is no consistent pattern regarding the months or seasons with lower evaluation results.

The second post-processing method proposed is the empirical quantile mapping, using the observed daily solar radiation, and the SSRD forecast in 2018 used as training data to bias-correct the solar radiation forecast in January – July 2019 period. The comparison of uncorrected and corrected forecast in Jambi Station is shown in Figure 4. The daily accumulated solar radiation forecast from each of the ensemble member in the January – July 2019 period are accumulated into monthly values, and the spread of the monthly values indicates the ‘precision’ of the forecast. The ground-observed daily solar radiation is also accumulated into monthly value and depicted as yellow lines in Figure 4. From Figure 4, in general, it can be seen that the value of the corrected forecast (right figures) is relatively closer to the observed value and the spread of the forecast seems to be smaller than the uncorrected forecast (left figures). The accuracy of the forecast, however, does not necessarily appear to depend on the lead times. This tendency has also been seen in Thorarinsdottir [9], where the bias in the model (for solar energy
forecast) is consistent for all lead times and the performance is slightly better in clear sky condition. Similar pattern is seen in almost all stations, and to find out whether the bias correction process produces significant improvements, a quantitative assessment is needed.

Figure 3. Correlation values between surface observation and solar radiation forecast in (a) one closest grid (uninterpolated) and (b) four closest grids (bilinearly interpolated). RMSE values generated from the forecast in (c) one closest grid (uninterpolated) and (d) four closest grids (bilinearly interpolated). The y-axis indicates the initial running time of the forecast.
Figure 4. The boxplot of monthly accumulation of solar radiation forecast in Jambi station. Left graphs depict uncorrected forecast and right graphs are the corrected forecast.
The quantitative assessment of the accuracy and precision of the corrected forecast, that was carried out by the calculation of $\alpha$ and $\beta$ values, is summarized in Figure 5. Accuracy improvement is indicated by an $\alpha$ value that is less than one. An increase in precision or a decrease in the spread of the forecast can be seen from a $\beta$ value that is smaller than one. In Figure 5, despite some outliers, the $\alpha$ value in all stations is lower than 1 for most initial running time and lead time. This indicates the potential for using this bias correction technique to improve the accuracy of solar radiation forecasts. Meanwhile, when viewed from the $\beta$ value, a large proportion of $\beta$ values exceed 1, that implies that the bias correction technique did not necessarily reduce the spread of the forecast. The ensemble spread provides more useful information than the ensemble average, as the ensemble value distribution can be used to provide forecast probabilities, which might be more applicable in the solar energy sector [9]. Other bias-correction techniques, such as artificial neural network or statistical downscaling, might be needed to generate forecast values with a higher degree of precision, although further evaluation should be done to find out which post-processing method best reduces biases in the model.
4. Conclusions
The evaluation of one of ECMWF products, the surface solar radiation downward or SSRD, has been done using the ground-observed radiation in some climatological stations in Indonesia region. Evaluation is carried out based on different cloud conditions, and two post-processing methods are applied to the estimated value to see whether added value can be given to the forecasts. The evaluation results show that a larger correlation value is seen on days with clear sky conditions, and an increase in the amount of cloud cover is associated with a decrease in the correlation value. However, this tendency is not shown by other error metrics, as the error value is the lowest in cloudy days. In sunny days, the model tends to underestimate the actual solar radiation, meanwhile in dark days, the forecast tends to be higher than the observed values. This is thought to stem from the fact that the model is less skilful to represent cloud-related processes.

The results of the evaluation of the two post-processing methods show that the spatial averaging method, in this case bilinear interpolation, does not provide a significant increase in the accuracy of the forecast value. On the other hand, the forecasts corrected using the empirical quantile mapping method show higher accuracy than the uncorrected forecast, although this bias correction method still cannot increase the level of precision of the ensemble forecast. Even though this evaluation is still carried out in a limited manner, as it is conducted at only five stations in the Indonesia region, this research is expected to provide preliminary information needed in the use of ECMWF products for prediction of solar radiation in the Indonesia region. The development of applied climate services for the solar energy sector can be done by utilizing information related to model performance and bias correction processes, as well as by combining sectoral information from potential users.

References
[1] Koo C, Hong T, Lee M, Park H 2013 Estimation of the Monthly Average Daily Solar Radiation using Geographic Information System and Advanced Case-Based Reasoning *Environmental Science & Technology* **47** 4829-4839
[2] Perez R, Kivalov S, Schlemmer J, Hemker K, Renné D, Hoff T 2010 Validation of short and medium term operational solar radiation forecasts in the US *Solar Energy* **84** 2161-2172.
[3] Marquez R, Guerguiev V, Coimbra C. 2012 Forecasting of Global Horizontal Irradiance Using Sky Cover Indices *Journal of Solar Energy Engineering* **135** doi:10.1115/1.4007497.
[4] Bernecker D, Riess C, Angelopoulos E, Hornegger J. 2014 Continuous short-term irradiance forecasts using sky images *Solar Energy* **110** 303-315.
[5] Lopes F, Silva H, Salgado R, Cavaco A, Canhoto P, Collares-Pereira M. 2018 Short-term forecasts of GHI and DNI for solar energy systems operation: assessment of the ECMWF integrated forecasting system in southern Portugal *Solar Energy* **170** 14-30.
[6] Perez R, Lorenz E, Pellant S, Beauharnois M, Van Knowe G, Hemker K et al. 2013 Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe *Solar Energy* **94** 305-326.
[7] Haiden T, Janousek M, Bidlot J. 2016 *Evaluation of ECMWF forecasts, including the 2016 resolution upgrade.*
[8] Bofinger S, Heilscher G. 2006 Solar electricity forecast – approaches and first results In: 20th Europ. PV conf.
[9] Thorarinsdottir, T., Løland, A. and Lenkoski, A., 2019 Probabilistic forecasting of temporal trajectories of regional power production-part 2: Photovoltaic Solar *arXiv preprint arXiv:1903.01186*.
[10] Lara-Fanego V, Ruiz-Arias J, Pozo-Vázquez D, Santos-Alamillos F, Tovar-Pescador J. 2012 Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain) *Solar Energy* **86** 2200-2217.
[11] Troccoli A, Morcrette J. Skill of Direct Solar Radiation Predicted by the ECMWF Global Atmospheric Model over Australia 2014 *Journal of Applied Meteorology and Climatology* **53** 2571-2588.
[12] Lopes, Conceição, Silva, Fasquelle, Salgado, Canhoto et al. Short-Term Forecasts of DNI from an Integrated Forecasting System (ECMWF) for Optimized Operational Strategies of a Central Receiver System 2019 Energies 12 1368.

[13] Sianturi Y, Marjuki M, Sartika K. 2019 Evaluation of ERA5 and MERRA2 reanalyses to Estimate solar irradiance using ground observations over Indonesia region In: International Energy Conference ASTECHNOVA 2019. [1] Lauret P, Lorenz E, David M. 2016 Solar Forecasting in a Challenging Insular Context Atmosphere 7 18.

[15] Lorenz E, Hurka J, Heinemann D, Beyer H. 2009 Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2 2-10.

[16] Liu, B.Y.H., Jordan, R.C. 1960 The interrelationship and characteristic distribution of direct, diffuse and total solar radiation Solar Energy 4 1–19.

[17] Themeßl, M. J., Gobiet, A., & Heinrich, G. 2012 Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal Climatic Change 112 449–468.