REGULAR PAPER

Probability prediction of solar irradiance in the tropic using ensemble forecasting

To cite this article: Daiki Harada *et al* 2023 *Jpn. J. Appl. Phys.* **62** SK1057

Japanese Journal of Applied Physics **62**, SK1057 (2023) **REGULAR PAPER**
https://doi.org/10.35848/1347-4065/acdf68

Probability prediction of solar irradiance in the tropic using ensemble forecasting

Daiki Harada[1], Naoki Moriai[2], Perawut Chinnavornrungsee[3], Songkiate Kittisontirak[3], Nuwong Chollacoop[3], Sasiwimon Songtrai[4], Kobsak Sriprapha[4], Jun Yoshino[2], and Tomonao Kobayashi[2*]

[1] *JR Central Consultants Company, Japan*
[2] *Gifu University, Japan*
[3] *National Energy Technology Center (ENTEC), Thailand*
[4] *National Electronics and Computer Technology Center (NECTEC), Thailand*

[*] E-mail: kobayashi.tomonao.e2@f.gifu-u.ac.jp

Received December 26, 2022; revised June 2, 2023; accepted June 19, 2023; published online July 7, 2023

As photovoltaic (PV) power generation systems become more widespread, the instability of electric power grids with PV connection is becoming an issue. For appropriate management of the grids, probability prediction of solar irradiance is proposed. The lagged average forecasting method is used for ensemble forecasting. The 72 h ahead forecasting of solar irradiance is operated in Thailand once a day, and it contains intraday, next-day, and 2-day ahead forecasts. Ensemble forecasting has three ensemble members. The accuracy of intraday forecasting is higher than that of the other members, and it is employed as the most probable value of the forecast. The relation between spreads and forecasting errors is analyzed. From the result, the confidence intervals of the predictions are derived for an arbitrary confidence level. The probability prediction is performed with the most probable value and the confidence intervals. The interval changes its width due to spread changes and captures the observation in it.

1. Introduction

A huge amount of photovoltaic (PV) systems are widely used and connected to electric power grids these days. PV systems become one of the major sources for grids in some countries. In Germany, PV systems provided 6.5% of electricity consumption nationwide in 2013[1] and the rate increased to 8.6% in 2019.[2] PV penetration in national electricity supply is more than 8.0% not only in Germany but also in five other countries, e.g. Honduras (14.8%) and Israel (8.7%), in 2019.[2] Under the circumstances, the supply of electricity by PVs of regularly covered about one-third of the noon peak demand on sunny summer days in Germany in 2013.[1] However, their output is not stable due to changes in weather and fluctuations in solar irradiance, causing the risk of instability of the electric power supply from the grids.

One of the ways to reduce the risk and ensure the efficient management of grids with PV systems is the prediction of solar irradiance related to the PV output, which is employed for the management of the grids. Diagne et al.[3] reviewed the reliability of solar irradiance forecasting with statistical models, cloud imagery, or numerical weather prediction (NWP) models, and showed the advantage of NWPs. Heinemann et al.[4] also investigated two kinds of forecasts; image processing for cloud development for very short-term forecasts and NWP for up to 2 days, and indicated the validity of a combination of the NWP and post-processing for solar irradiance forecasting. Lorentz et al.[5] employed weather forecasting provided by the European Center for Medium-Range Weather Forecasts (ECMWF) for solar irradiance and PV power predictions, and used them for electric grid management in Germany. Their forecasting period is 3 days. They also discussed the effect of area size for reducing the forecasting error. Lara-Fanego et al.[6] forecasted solar irradiance by using NWP in southern Spain. They evaluated not only global horizontal irradiance but also direct normal irradiance by introducing physical post-processing of the NWP. Shimada et al.[7] also performed solar irradiance forecasting in Japan with the same NWP as that used by Lara-Fanego et al.[6] and obtained results with lower accuracy than Lara-Fanego et al. As the reason for the low accuracy, they pointed out that Japanese weather conditions, which are less sunny and more cloudy compared with southern Spain, change frequently, making it difficult to conduct forecasting. Aryaputera et al.[8] also used NWP to forecast the irradiance in the Southeast Asia region, Singapore, which is near the target area, Thailand, in this study. They also proposed stochastic methods, and improved the forecasting accuracy by combining the results from the NWP and the methods.

We have constructed a weather and solar irradiance forecasting system for the prediction of PV generation and its fluctuations in Thailand.[9] The Thailand national government promotes the installation of large-scale PV power plants to make PVs a major electric power source in Thailand to shift to renewable energy to create a sustainable society, and EA Solar Phitsanulok plants with 133.92 MWp and EA Solar Lampang plants with 128.39 MWp[10] are now operating. The weather and the solar irradiance forecasting system are operated for the management of the electric power grid to which the large PV power plants are connected. For its proper management, highly accurate forecasting is required.

The weather in Thailand is divided into three seasons. First is the rainy season from mid-May to October. Air is wet and warm, and squalls occur frequently during this season. Winter is from November to mid-March, and the weather is dry with mild temperatures. Summer is from mid-March to mid-May, and the weather is very hot. It is said that weather forecasting in the tropics is relatively more difficult than in the temperate (middle) latitude zone, like Japan.[11,12] Clouds and rainfalls in temperate latitudes are due to the frontal development of air masses, and their movement is gradual. Therefore, forecasting in the latitudes is easy and very reliable. On the other hand, in the tropic, convectional clouds and rainfalls occur due to thermally induced low-pressure systems. Their movement is vertical and fast, and they make forecasting the weather phenomenon in the tropics difficult. Cumulus and cumulonimbus are common in the tropics, and their behavior, especially their generation, is complicated and difficult to forecast. This causes difficulty in solar irradiance forecasting and the prediction of PV generations in the tropics.

There are several ways to improve the accuracy of weather and solar irradiance forecasting with NWPs, e.g. model optimization and post-processing. Shenoy et al.[13] optimized an NWP to forecast tropical cyclones over the Bay of Bengal. Yoon et al.[14] also optimized it for sea breeze prediction over the northeastern coast of South Korea. They performed sensitive analyses of the physical parameterization options used in the NWP, and found the optimum combination of schemes of the options for their purpose. Arasa et al.[15] and Verbois et al.[16] also employed the same analysis and improved the weather forecasting in southern Spain and Singapore, respectively. The authors focused on solar irradiance forecasting in Thailand and found the optimal schemes with the same analysis as them.[17] The representation of the short-period fluctuation of the irradiance is improved with optimization. However, the forecasting error caused by the NWP itself remains due to imperfections in the model.

Post-processing can be applied to adjust NWP outputs. National meteorological centers worldwide commonly employ processing to correct NWP outputs and produce the results for public distribution, which are referred to as guidance. The Japanese Meteorological Agency employs appropriate post-processing methods for each meteorological parameter computed with NWP; multiple linear regression, logistic regression, neural networks, Kalman filters, and others.[18] Yang et al.[19] applied Bayesian and simple linear regressions as post-processing of NWP forecasting and improved the wind speed prediction of storms in the northeastern United States of America. Both Diagne et al.[20] and Rincón et al.[21] employed Kalman filters for post-processing for solar irradiance forecasting at Reunion in the Indian Ocean and Catalonia in Spain, respectively. The authors develop a nonlinear Kalman filter as a post-processing of NWP to adjust the solar irradiance forecasted with the NWP.[22] Recently, machine-learning techniques have also been employed for not only solar irradiance forecasting but also for the prediction of PV output. Hossain and Mahmood[23] used a long short-term memory neural network for the prediction of PV generation. Zhang and Zou[24] used machine learning with historical PV power generation data and meteorological forecasting.

Many researchers and operators are working to improve the accuracy of forecasting. However, there is a limit to the improvement of forecasting accuracy through model improvement, etc. due to the uncertainty in the representation of

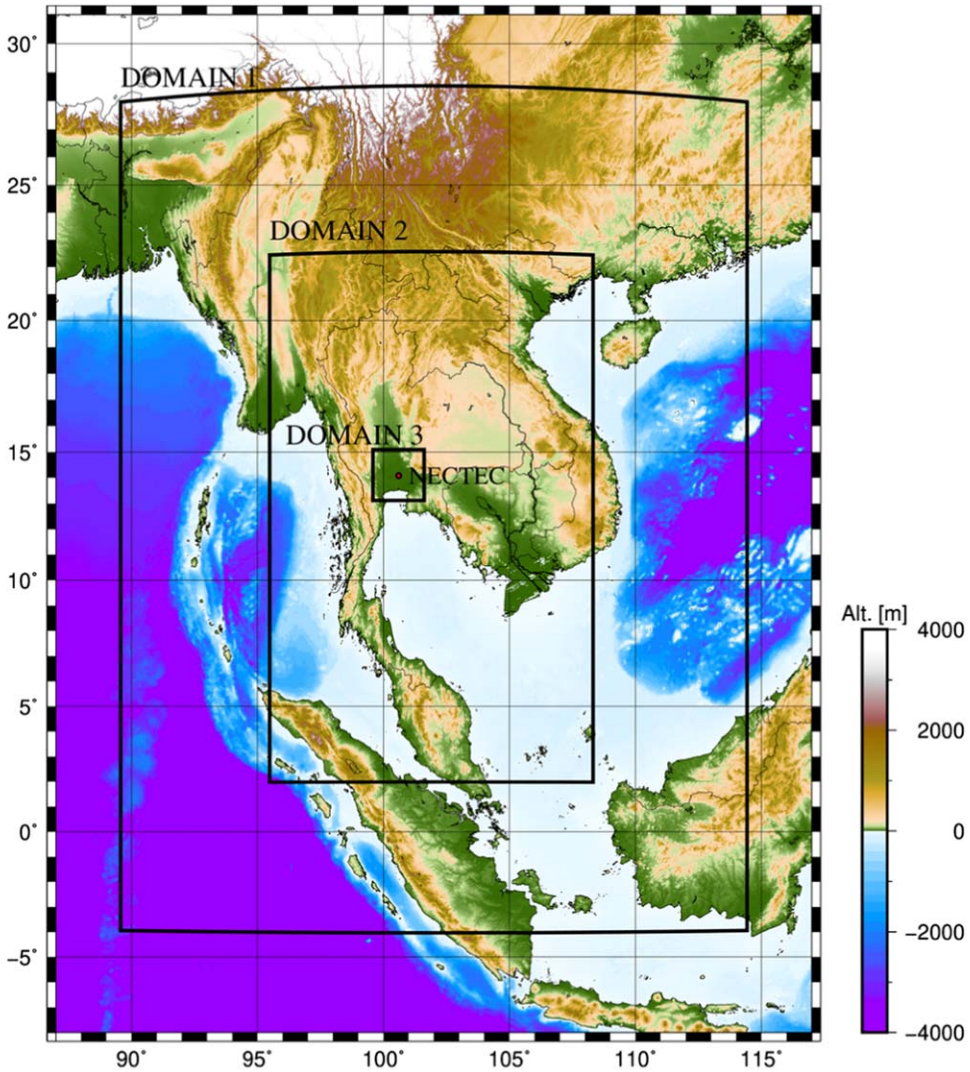


Fig. 1. Topography of computational domains for simulating solar irradiance with the meteorological model WRF. The observation site, NECTEC, is located at the center of Domain 3.

Table I. Computational conditions of the meteorological model WRF for forecasting solar irradiance in Thailand.

Period	Start: 12:00 UTC (19:00 LST) 72 h forecasting in computation
Input data	NCEP GFS-0.25 (1 hourly, 0.25° × 0.25°)
Nesting	2-way nesting
Domain (horizontal resolution, number of grids)	Domain 1 (18 km, 135 × 179 grids) Domain 2 (6 km, 217 × 340 grids) Domain 3 (2 km, 100 × 100 grids)
Vertical layer	50 levels (from surface to 100 hPa)
Physics parameterization options	SBU-Lin microphysics Goddard longwave radiation scheme Dudhia shortwave radiation scheme KF cumulus parameterization MM5 surface layer scheme YSU planetary boundary layer scheme
FDDA option	Enable (Domain 1 only)

the models and weather forecasting products. As an alternative method for improving the accuracy, probability prediction of the irradiance with ensemble forecasting[25] is proposed. Today, several national meteorological centers, e.g. the ECMWF,[26] National Centers for Environmental Prediction (NCEP)[27] in the USA, and the Japan Meteorological Agency (JMA),[28] use it in their daily operations. Instead of single computation for forecasting, several computations with different setups or model formulations are performed in ensemble forecasting. Each computation is called an ensemble member. The forecasting and its probability are evaluated from the set of members. This ensemble forecasting has been used for weather forecasting and is now also used for solar irradiance forecasting. Liu et al.[29] performed ensemble forecasting of the irradiance in Japan. They predicted both the solar irradiance and its reliability, and the confidence interval of the forecasting captures the observed one. The method they used is a traditional one in meteorology. On the other hand, Singla et al.[30] employed a machine-learning technique for the ensemble forecasting of solar irradiance.

This study uses ensemble forecasting to predict solar irradiance in Thailand. The applicability of the ensemble forecasting to large solar radiation fluctuations in Thailand will be examined.

2. Experimental methods

2.1. Weather Research and Forecasting model

The Weather Research and Forecasting (WRF) model[31] is applied in this study for forecasting solar irradiance. This model is a physical meteorological model with fully compressible non-hydrostatic equations developed by the National Centers for Atmospheric Research (NCAR) and the NCEP, and is used not only for real-time numerical weather forecasting (e.g. Salvação and Guedes Soares[32]) but also for research under idealized conditions, data assimilation, and others (e.g. Pryor et al.[33]). It contains many physical parameterization options for meteorological microprocesses, e.g. planetary boundary layer physics and cumulus parameterization, and users select schemes of the options for their specific purpose. It can simulate two-way nesting and be applied across scales ranging from meters to thousands of kilometers. It is an open-source software and is used widely in the world.

2.2. Prediction of the solar irradiance forecasting system in Thailand and its improvement

A series of our studies is for appropriate solar irradiance forecasting in Thailand. First, a solar irradiance forecasting system was constructed with WRF. Next, we improved the WRF itself by optimization of the parameterization options. Third, post-processing was installed for adjusting WRF computational results. In the end, we try probability forecasting of the solar irradiance.

For the prediction of PV generation for strategic electric power grid management in Thailand, the solar irradiance forecasting system was constructed based on WRF and has been operated.[9]

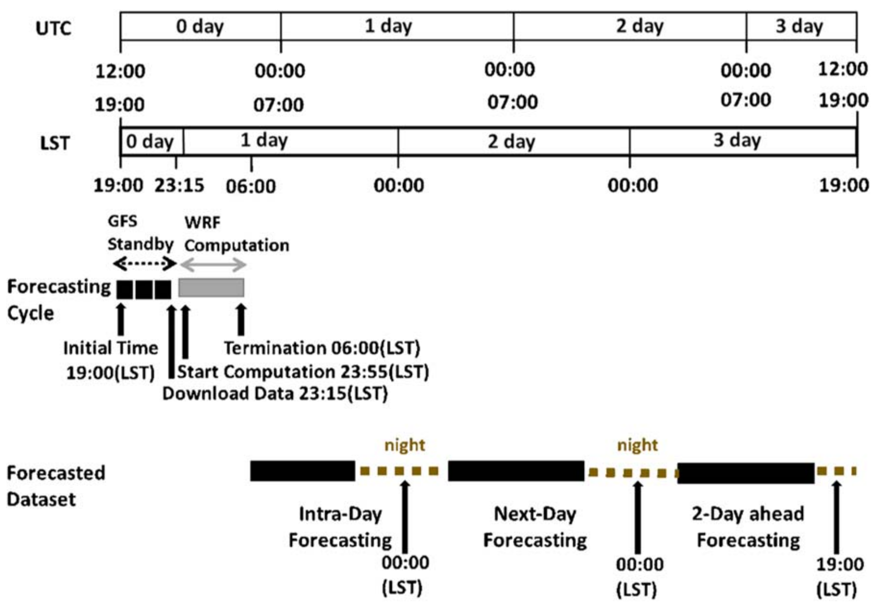


Fig. 2. Forecasting cycle of weather and solar irradiance with WRF and forecasted dataset period in the computed result.

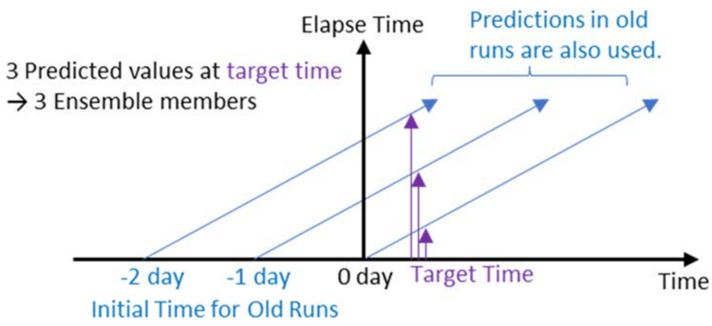


Fig. 3. Diagram of the LAF method. The vertical axis indicates the elapsed time of each forecasting computation from its initial time. The diagonal arrows pointing up on the right indicate each forecasting computation.

In the tropics, Thailand, which is the target area in this study, cumulus and cumulonimbus are common, and it is difficult to simulate them with NWPs, including WRF, as explained in the previous section. We performed a sensitivity analysis for the physical parameterization options of WRF and derived the optimal schemes for their options.[17] In the previous study, we employed the optimal schemes: SBU-Lin,[34] Goddard,[35,36] Dudhia,[37] KF,[38] MM5, and YSU[39] are employed for the options of microphysics, longwave radiation, shortwave radiation, cumulus, surface layer, and planetary boundary layer, respectively. The WRF is sophisticated and by introducing the optimal schemes, the forecasting accuracy is improved.

A complete simulation with WRF itself is not feasible because of its imperfections. NWPs including WRF usually have a specific tendency, including in their forecasting results. We developed the post-processing technique with a Kalman filter to remove the specific tendency of WRF and adjusted the solar irradiance forecasted with it.[22] In the previous study with the Kalman filter, we found that the WRF tended to overestimate the irradiance, especially at moderate irradiance intensities, 200–600 W m^{-2}. The filter we used is a nonlinear Kalman filter, and it removed the overestimation and adjusted the irradiance WRF computed.

In solar irradiance forecasting with the sophisticated WRF and the Kalman filter, forecasting errors still remain due to the insufficiency of our method developed in a series of our studies. Then, we employed probability forecasting and forecasted not only the irradiance but also its reliability. The probability forecasting used in the series of our studies is explained in this article.

2.3. Computational domain and target area

Figure 1 shows the topography of Thailand and the computational domains in this study. It is three-level nesting; the coarse domain is Domain 1, and its child domain is Domain 2. The parent domain of Domain 3 is Domain 2. The horizontal resolutions of Domains 1–3 are 18, 6, and 2 km, respectively. Domain 2 covers all of Thailand. At the center of Domain 3, the global horizontal solar irradiances and some meteorological parameters are observed with a pyranometer and a weather station as references of the WRF forecasting. The observation point is the National Electronics and Computer Technology Center (NECTEC).

2.4. Computational conditions

The computational settings of weather forecasting with WRF are summarized in Table I. The analysis data and forecasted data 72 h ahead of the NCEP operational Global Forecast System (GFS) 0.25° grids (horizontal resolution 0.25° × 0.25°, 1 h interval)[40] are used as the initial and boundary conditions for model simulation. Figure 2 shows the solar irradiance forecasting cycle in daily operation in this study.[9] The computation of forecasting is performed once a day. The initial time of the forecasting is 19:00 LST (12:00 UTC). Here, Thailand Local Standard Time (LST) is 7 h ahead of Universal Coordinated Time (UTC). The GFS data[40] is downloaded about 4 h after the initial time, and the forecasting computation of the weather and the solar irradiance with WRF is initialized. Soon after the initialization, WRF starts the computation of the forecasting. Finally, the computation is terminated, and the results are displayed in the morning, at 6:00 LST, of the next day in LST. The forecasting period is from the initial time to 72 h ahead, the computational period consists of three time periods during the day, i.e. intraday, next-day, and 2-day ahead from the termination of computation.

The target period for the forecasting in this study is from 2019 October 1 to 2020 September 30, the entire 1 yr, and the forecasted solar irradiance at the observation point is output every 10 min.

2.5. Ensemble forecasting of solar irradiance

As explained in the previous sections, we improved WRF by introducing the optimal schemes and employing a Kalman filter as post-processing to evaluate solar irradiance with high accuracy. However, forecasting errors still remained; then, we introduce an ensemble forecasting of the solar irradiance for its probabilistic prediction.

Ensemble forecasting is used for the probability prediction of numerical weather forecasting. It consists of a set of forecasts. The set aims to give the most probable forecast and the range of possible states in the forecast. Several computations with different simulation setups or model formulations are performed to prepare the set. Each computation is called an ensemble member. The forecasting and its probability are evaluated from the set of members. There are several ways to obtain members in ensemble forecasting. In this study, the lagged average forecasting (LAF) method[41] is employed for the ensemble forecasting. Weather forecasting is computationally expensive, and several forecasts are required to prepare the set of members in ensemble forecasting. This is the main reason for the difficulty in ensemble forecasting. In daily weather forecasting, the days covered by the forecasting overlap, for example, intraday forecasting operated on the same day and next-day forecasting 1 day before. LAF uses overlapped forecasting as the members of the ensemble forecasting. With this method, the computational load of ensemble forecasting is the same as the daily operation of single forecasting. Between each daily weather forecast, there are time lags in the computational starting time. This means that the ensemble members of LAF have time lags at the start of each other.

We constructed the solar irradiance forecasting system in Thailand and are operating it.[9] It forecasts the irradiance once a day, and its forecasting period is from the initial time to 72 h ahead, the computational period consists of three time periods during the day, i.e. intraday, next day, and 2-day ahead, as explained in the previous section. Figure 3 shows a diagram of the LAF used in this study. The horizontal axis indicates the time, and the vertical axis indicates the elapsed time of each forecasting computation from its initial time. Each computation proceeds along the diagonal arrows from the initial time on the horizontal axis as time elapses. There are three ensemble members at the target time, intraday forecasting operated on the same day (0 day), next-day forecasting on 1 day before (−1 day), and 2-day ahead forecasting on the 2 days before (−2 day) as shown in the figure. The most probable value and its prediction interval in solar irradiance forecasting are evaluated by the members.

2.6. Observation data

The global horizontal solar irradiance and meteorological parameters, e.g. wind speed and ambient temperature, are observed at the center of Domain 3 as shown in Fig. 1 as references of the forecasting, as explained before. The sampling interval of the observation is 1 min. As the WRF output is a 10 min interval, 10 min average observational data is used in this study.

2.7. Statistical error indices

There are several statistical indices, e.g. rms error (RMSE) and mean absolute error (MAE). In this study, we employ RMSE and MAE for the forecasting error to the observation.

The RMSE and MAE are defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}, \tag{1}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i|, \tag{2}$$

where x_i and y_i are simulated and observed diurnal irradiances at the same time, respectively. Their time intervals are 10 min in this study. Nighttime irradiances are zeros, and they are excluded from the evaluation. n is the number of data x_i or y_i.

3. Forecasting results and discussion

3.1. Forecasting accuracy of ensemble members

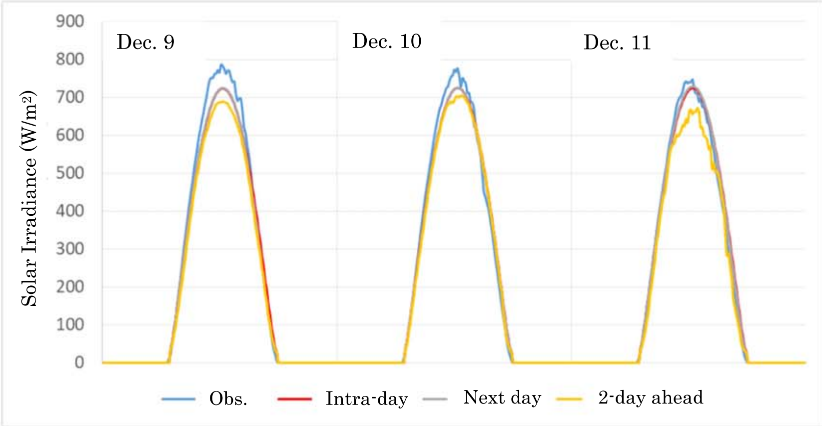
The forecasting period for the daily operation of the solar irradiance forecasting is 72 h ahead, the computational period consists of three time periods during the day, i.e. intraday, next day, and 2-day ahead, as explained before. The LAF method is used for the ensemble forecasting in this study, and the forecasted irradiance for each day is an ensemble member. Here the accuracy of each ensemble member is discussed.

The statistical error indices, RMSE and MAE, of ensemble members throughout the year and in each season are listed in Table I. The RMSE and MAE are small in winter compared with the other seasons. The weather in winter is mild and dry, and there are many fine days. Therefore, WRF represents the weather and solar irradiance during the season with high accuracy. On the other hand, there are many cloudy and rainy days during the summer and rainy seasons, and the irradiance fluctuates due to shading by clouds. This makes it difficult for WRF to forecast the irradiance during the seasons. Both the RMSE and MAE of the intraday forecasting are the smallest, and that for the 2-day ahead forecasting is the largest evidently in the summer and rainy seasons. This is caused by the difficulty of forecasting during the season. The differences in the RMSEs and MAEs between different forecast time periods in winter are very small, thereby keeping the differences in annual errors low. Usually most of the time, earlier forecasting is considerably less efficient than later forecasting. When the most probable irradiances are evaluated, the different members cannot be treated fairly because of different forecasting errors in LAF, as shown in Table II. Ebisuzaki and Kalnay[42] proposed the scaled lagged average forecast (SLAF) method, which uses different weights for different members when the most probable irradiances are evaluated. In this study, the irradiance of the intraday forecasting is used as the most probable one in the ensemble forecasting because its accuracy throughout the year is the highest, as shown in Table II.

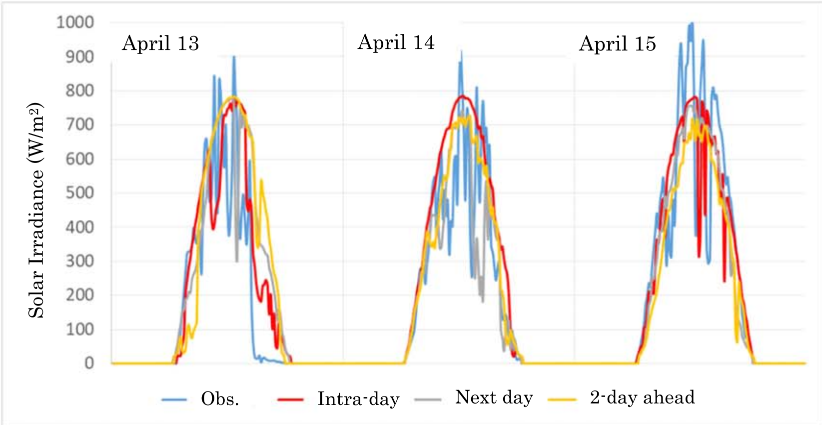
Table II. Forecasting accuracies, RMSE, and MAE of each ensemble member.

Ensemble member	Period (season)			
	Year	Winter	Summer	Rainy
(a) RMSE (W m^{-2})				
Intraday forecasting	174.9	118.7	193.0	206.8
Next-day forecasting	175.3	116.1	199.8	206.2
2-day ahead forecasting	183.3	118.0	213.8	215.3
(b) MAE (W m^{-2})				
Intraday forecasting	112.2	69.8	121.4	147.1
Next-day forecasting	114.5	70.2	128.4	149.0
2-day ahead forecasting	120.7	73.0	139.2	156.4

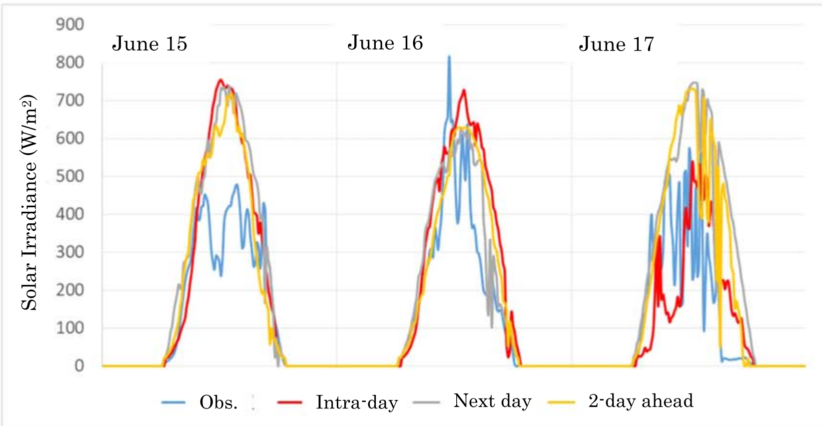
Figure 4 shows the forecasted daily solar irradiances of the ensemble members. Three members, intraday, next-day, and 2-day ahead forecasting are indicated with red, gray, and yellow lines, respectively. The observation is also plotted with blue lines in the figure. Since solar irradiance fluctuation varies greatly with the seasons in Thailand, it is shown in the figure for each season. In winter, from November to mid-March, there are almost no clouds, and the solar irradiance fluctuations due to clouds are not found in the observation (blue line) shown in Fig. 4(a). The accuracy of the solar irradiance forecasting is high during this season as indicated in Table II, and the irradiance of each member traces well to the observation as shown in this figure. In the summer, from mid-March to mid-May, the observed solar irradiance (blue line) fluctuates widely due to clouds as shown in Fig. 4(b). From the characteristics of the fluctuation, cumulus and cumulonimbus are major clouds during this season in Thailand. The irradiances of ensemble members also fluctuate but the intensities of the fluctuation are small evidently compared with the observation as shown in Fig. 4(b). Miyamoto et al.[43] performed weather forecasting with an



(a) Winter (from November to mid-March)



(b) Summer (from mid-March to mid-May)



(c) Rainy season (from mid-May to October)

Fig. 4. Time series of forecasted daily solar irradiances of ensemble members.

NWP model and investigated the reproducibility of the cumulus and cumulonimbus. They suggested that forecasting with a horizontal resolution of 2 km is appropriate for the reproduction of the clouds. However, they also reported that the cloud reproducibility of the NWP model is about 5 times higher than the horizontal resolution. The finest resolution used in that study is 2 km, just the same as their numerical experiment. According to their result, the smallest cloud that the NWP model can reproduce is about 10 km in this study. The scale of the clouds or nonuniformity of clouds that cause the fluctuation of the observed irradiance shown in Fig. 4(b) may be smaller than the representation limits of NWPs, 10 km and they are not represented well in the WRF forecasting. During the rainy season, from mid-May to October, the observed irradiance is smaller and fluctuates less than that in the summer as shown in Figs. 4(b) and 4(c). The characteristics of the irradiance shown in Fig. 4(c) indicate that thick clouds cover the sky during this season and shade the irradiance more than in the summer. On 15 June, the observed irradiance is shaded with thick clouds as in shown Fig. 4(c); however, all the ensemble members do not forecast the clouds and their shading. On 17 June, two members, next-day and 2-day ahead forecasting, show incorrect results, similar to those on 15 June, but the intraday

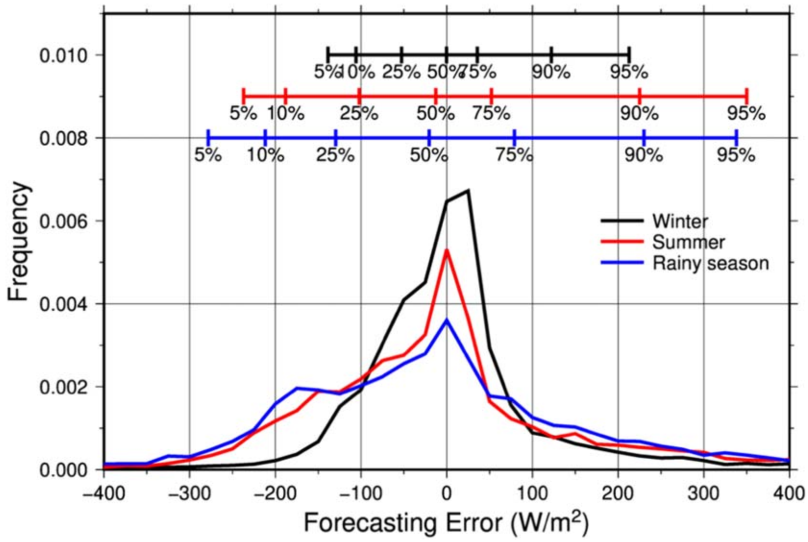


Fig. 5. Histograms of forecasting error in intraday forecasting. Their quantiles are indicated at the top of the figure.

forecasting improves the incorrect forecasting, as shown in the figure.

Figure 5 shows histograms of forecasting errors in each season. Their quantiles are also indicated at the top of the figure. Intraday forecasting, which is the most probable one in this study, is used in this figure. Negative values of the errors mean the underestimation of the forecasting. The error in winter is smaller than in the other seasons, and its distribution concentrates around the origin. The absolute error of the quantile 95% is larger than the one with 5%. This means that the overestimation of forecasting is larger than the underestimation. The distributions during the summer and rainy seasons spread wider than the ones in winter, and they indicate larger forecasting errors in the seasons. The medium values of the seasons, whose quantiles are 50%, are a slightly negative underestimation in the forecasting. The positive side of the distributions spread wider than the negative in the seasons, similar to in winter. As shown in the figure, the distributions of forecasting errors deform slightly from symmetric.

3.2. Relation between spread and forecasting error and evaluation of confidence interval

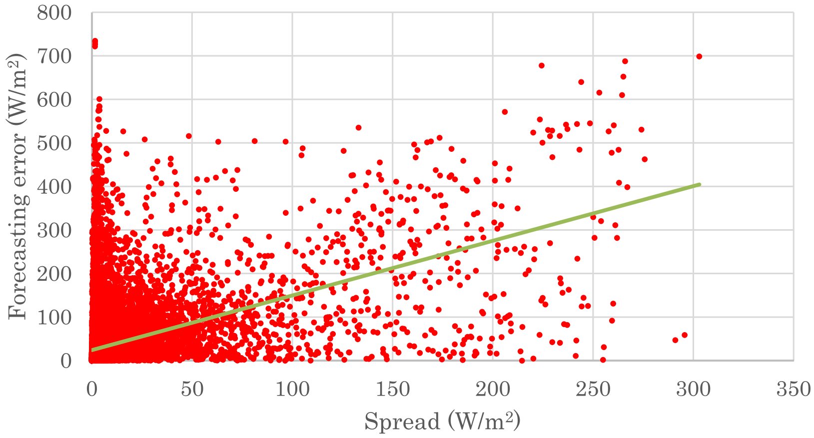
As an index of forecasting confidence level, the spread is used in ensemble forecasting. The standard deviation of irradiances of the ensemble members in arbitrary time is defined as spread. It is expressed as the following equation:

$$\text{spread} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{3}$$

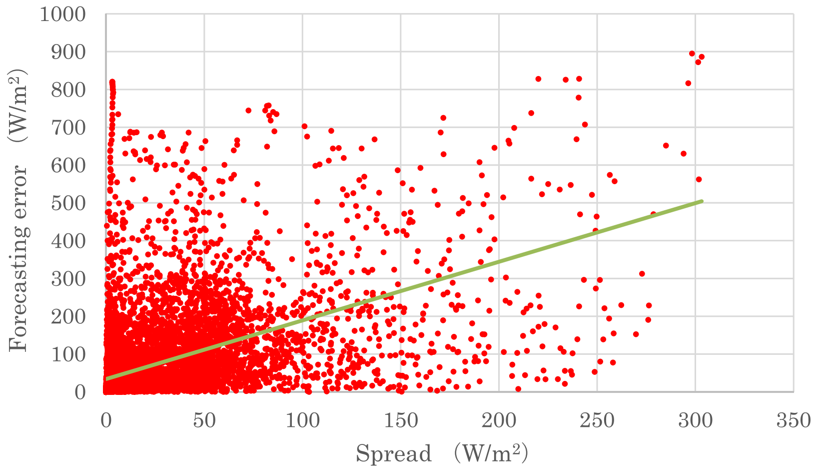
where x_i is simulated irradiance at an arbitrary time, and $\bar{x}$ is the average of x. n is the number of data x_i. Since there are three ensemble members in this study, n is 3. The spread corresponds to the variation among ensemble members as the equation indicates. When the spread is small, the confidence level of the forecasting is high.

Figure 6 shows the relation between the spreads and absolute forecasting errors. The lines in the figure indicate linear regression lines. The spreads and the errors are evaluated every time step (10 min interval) of WRF output and plotted on the figure. The absolute forecasting error is calculated as the absolute difference between the most probable irradiance which is of the intraday forecasting, and the observation. There are three diagrams in the figure for each season in Thailand. In each diagram, there is more plotted data as the spread is lower. It is also found from the regression lines in the figure that the forecasting error becomes larger as the spread becomes bigger. In the winter, the plotted data in Fig. 5(a) are concentrated at the bottom, and indicate good forecasting accuracy during this season. On the other hand, in the summer and rainy seasons shown in Figs. 6(b) and 6(c), there are plotted data with large forecasting errors, especially in areas with large spreads. This indicates the difficulty of solar irradiance forecasting during these seasons. Figure 7 shows the frequency of the forecasting errors for the spreads between (a) 25 and 50 W m^{-2}, (b) 100 and 125 W m^{-2}, and (c) 200 and 225 W m^{-2} during the rainy season. Different from ordinal frequency charts, the vertical axes indicate the forecasting errors for the comparison with Fig. 6(c), and the frequencies are denoted in the horizontal axes. Its distribution profiles are rough due to the small number of data, but it is found from the figure that there are many data with small errors and the number of data reduces as the error becomes bigger. This characteristic is also found in the data with a similar spread in other seasons.

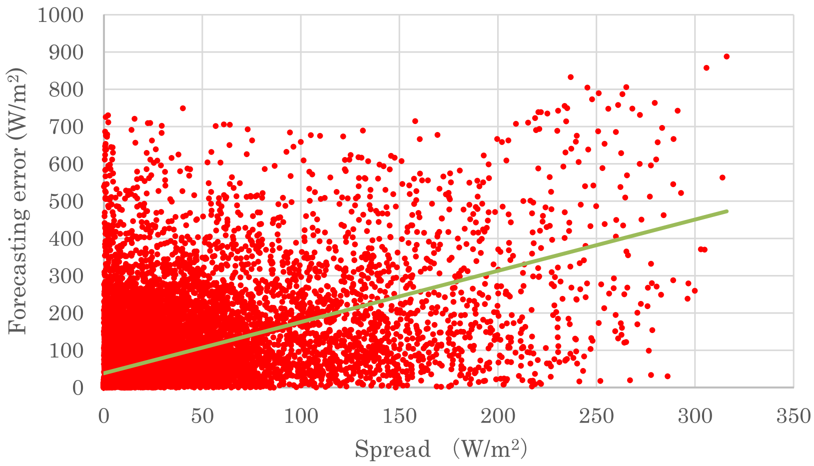
Confidence intervals of the solar irradiance forecasting are evaluated as follows. First, the data plotted in Fig. 6 are grouped separately for each 25 W m^{-2} of the spread. The groups are called bins. The frequency charts in Fig. 7 correspond to the forecasting error distributions in each bin. Next, the absolute forecasting errors are determined to contain a specified percentage of the plotted data in each bin, starting with the lowest error. The determined values correspond to the confidence intervals of the forecasting for the specified probability of forecasting error. Figure 8 shows the lines of confidence intervals plotted on the plots about the relationship between the spreads and the absolute forecasting error. The dot-plotted data in this figure are the same as the ones in Fig. 6. The confidence intervals for each bin are drawn with staircase lines in this figure. The yellow, orange,



(a) Winter (from November to mid-March)



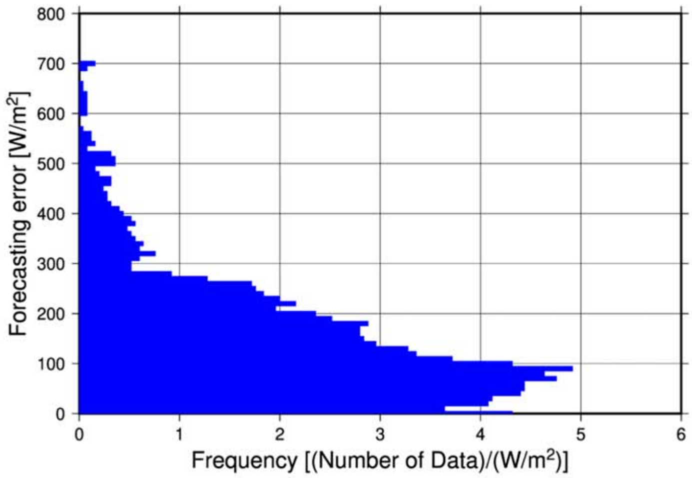
(b) Summer (from mid-March to mid-May)



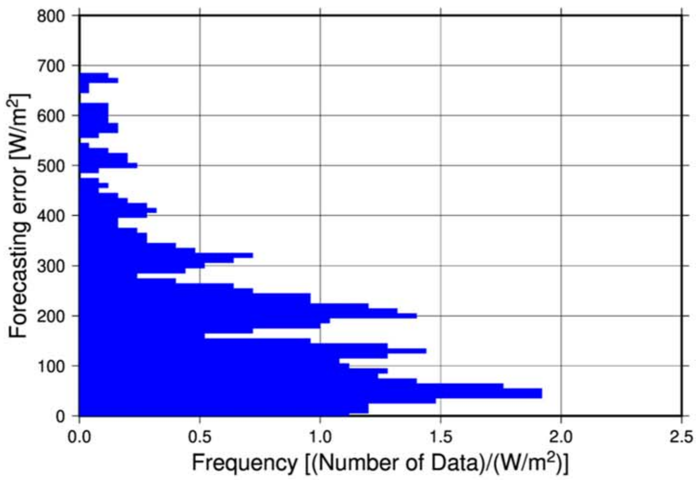
(c) Rainy season (from mid-May to October)

Fig. 6. Relationship between spreads and absolute forecasting error. The lines indicate linear regression lines. Correlation coefficients during the (a) winter, (b) summer, and (c) rainy seasons are 0.47, 0.50, and 0.54, respectively.

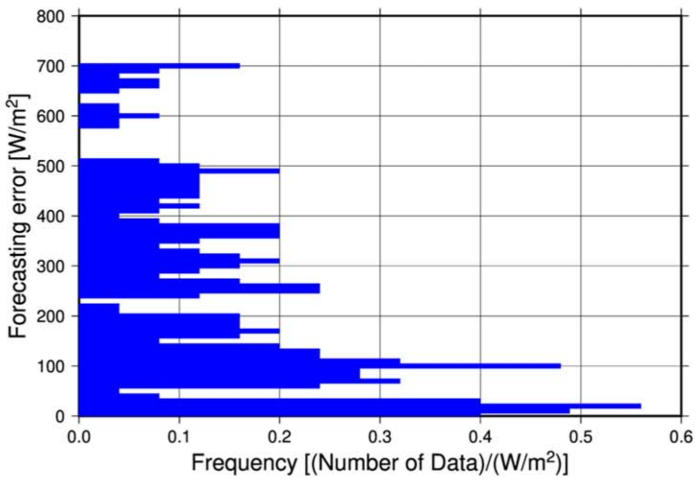
and brown lines denote the intervals for confidence levels 50%, 80%, and 90%, respectively. The confidence intervals increase as the spreads become large in each season as shown in Fig. 8. In some cases, the intervals reduce when the spreads are large because the bins with large spreads contain a small number of plotted data and it makes the accuracy of the evaluated interval worse. As the probability of forecasting errors becomes larger, the intervals also increase. The intervals in winter, Fig. 8(a) are smaller than the ones in the summer and rainy seasons, Figs. 8(b) and 8(c). This also indicates the difficulty of solar irradiance forecasting during these seasons.



(a) Spread from 25 to 50 W/m^2



(b) Spread from 100 to 125 W/m^2

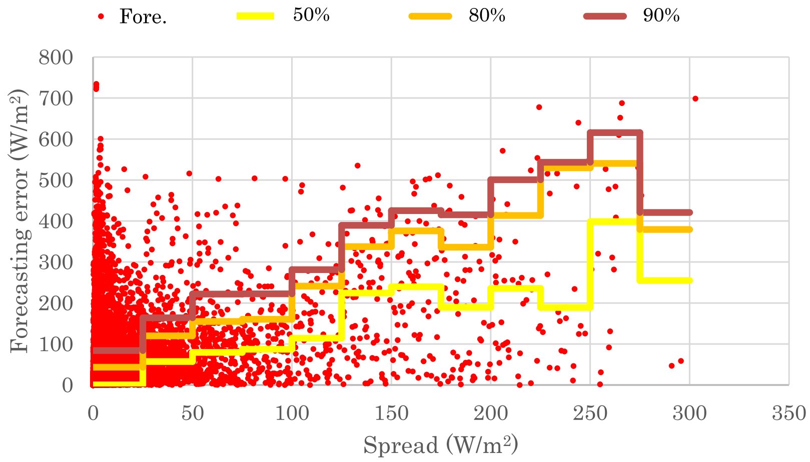


(c) Spread from 200 to 225 W/m^2

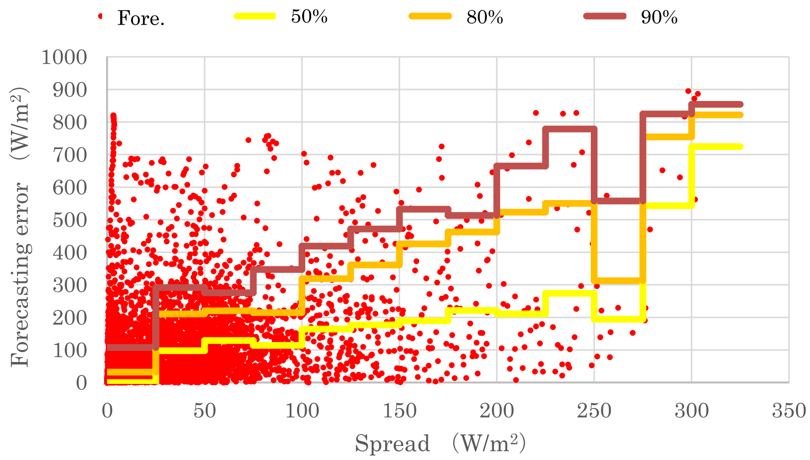
Fig. 7. Frequency of forecasting errors for the spreads between (a) 25 and 50 W m^{-2}, (b) 100 and 125 W m^{-2}, and (c) 200 and 225 W m^{-2} during the rainy season.

3.3. Probability forecasting with ensemble forecasting

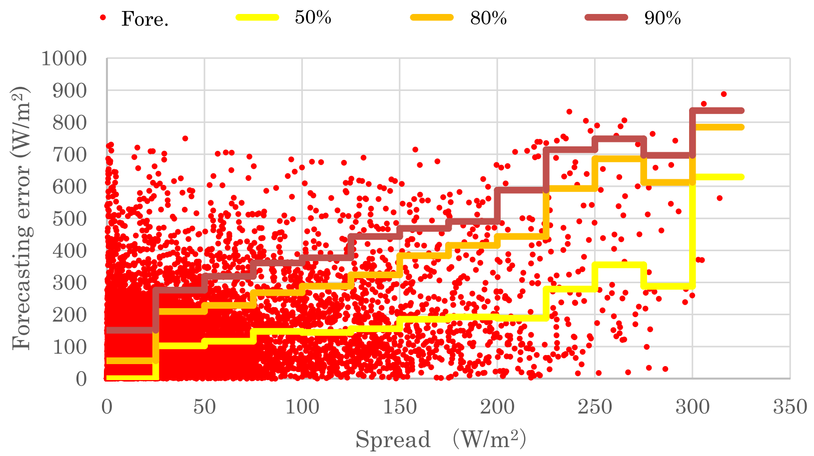
The most probable values of solar irradiance forecasting and their confidence intervals have been evaluated in the previous sections. The probability prediction of solar irradiance is performed with ensemble forecasting in this section. The most probable values of forecasted daily solar irradiance are plotted in light blue in Fig. 9. The width of the lines corresponds to the confidence interval. The areas in yellow, orange, and brown represent the intervals with confidence



(a) Winter (from November to mid-March)



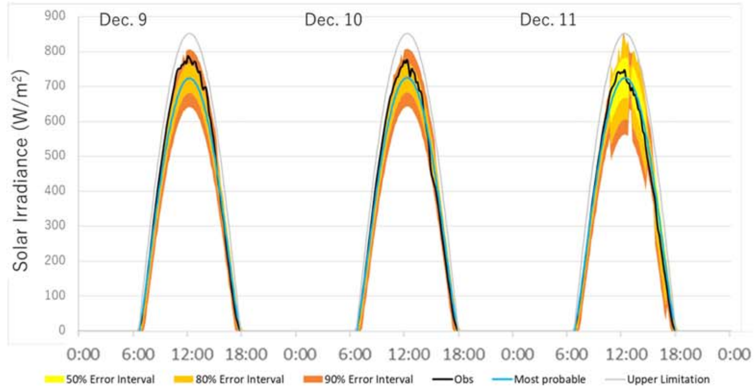
(b) Summer (from mid-March to mid-May)



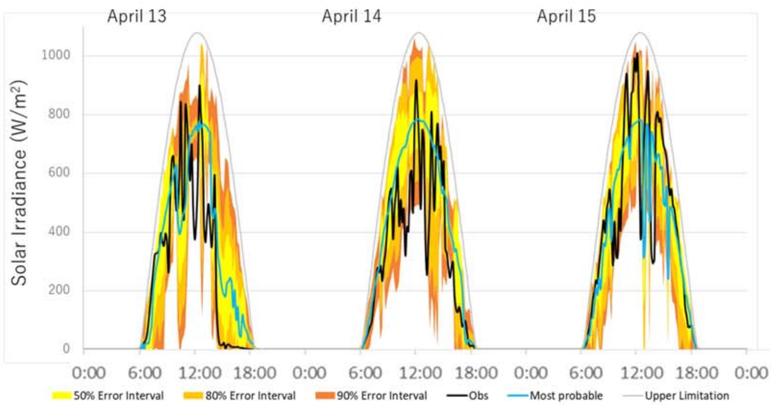
(c) Rainy season (from mid-May to October)

Fig. 8. Relationship between spreads and absolute forecasting error. The lines indicate the confidence intervals with confidence levels of 50%, 80%, and 90%.

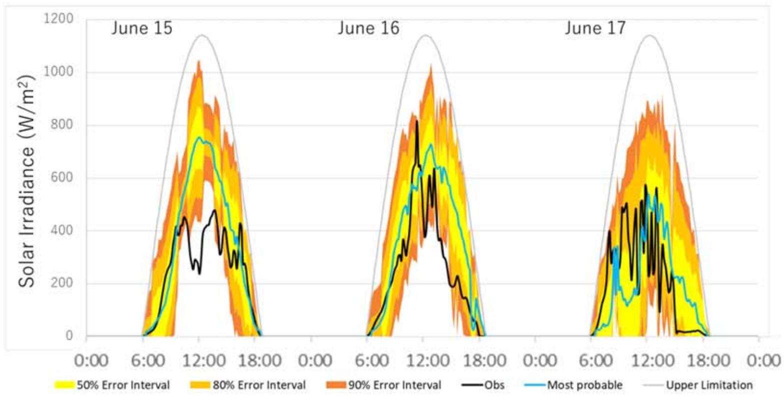
levels of 50%, 80%, and 90% of the forecasting, respectively. We put the limits on the forecasted irradiances and their intervals; the irradiance under a clear sky is for the upper limit, and irradiance zero is for the lower limit. The observed irradiances are also plotted in black in the figure. The confidence interval with a confidence level of 90% is wider than the ones with levels 50% and 80%. The confidence interval in winter, Fig. 9(a), is smaller than the ones in the summer and rainy seasons, Figs. 9(b) and 9(c). In winter, the most probable value of the forecasting traces well with the



(a) Winter season (from November to mid-March)



(b) Summer season (from mid-March to mid-May)



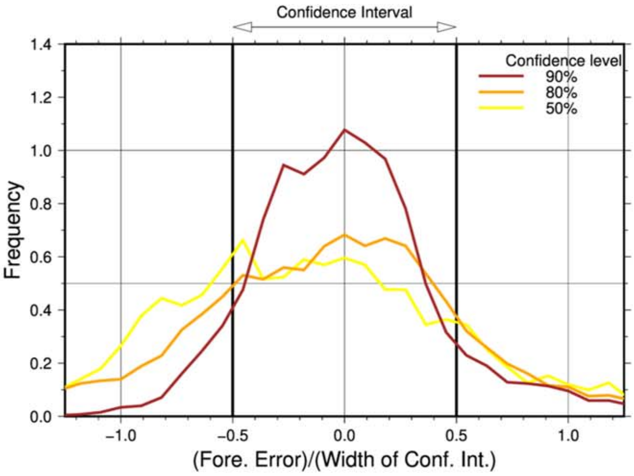
(c) Rainy season (from mid-May to October)

Fig. 9. Probability forecasting of daily solar irradiance. The most probable values of the forecasted irradiance and confidence intervals are plotted with thick light blue and black lines, respectively. The confidence intervals correspond to the widths of the thick black lines. The upper limits of the confidence intervals are 851, 1080, and 1141 W m^{-2} as shown by the graphs in panels (a)–(c), respectively.

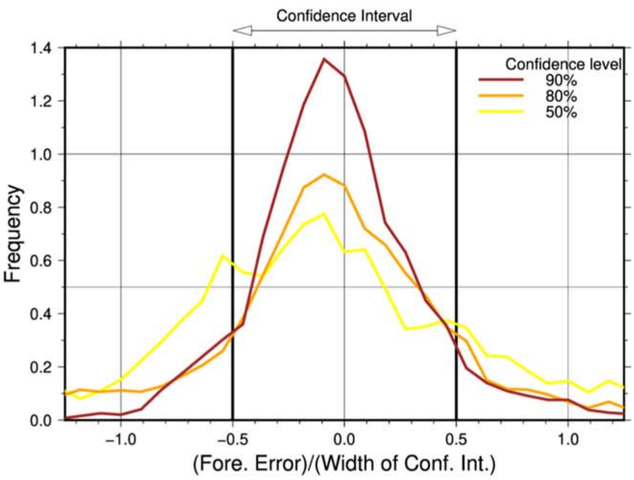
observation and the confidence intervals cover the observation as shown in Fig. 9(a). On the other hand, in the summer and rainy seasons, the most probable values do not represent the observations, but the observation is in the range of the confidence intervals as shown in Figs. 9(b) and (c). Sometimes, like on 15 June as shown in Fig. 4(c), all ensemble members were incorrect in their forecasting, and the confidence interval cannot capture the observation as shown in Fig. 9(c). The risk of the observation being outside the confidence interval in the forecasting depends on the specified confidence level for evaluating the interval. If the level is low, the risk of an incorrect prediction is high. When the level is increased, the risk is reduced, but the confidence interval becomes wider.

3.4. Verification of the accuracy of probability forecasting

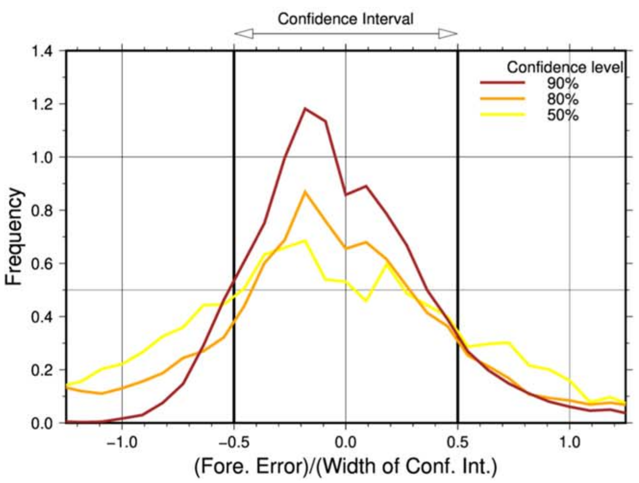
The accuracy of the probability forecasting performed in the previous section is verified. Figure 10 shows histograms of forecasting errors normalized with the width of the confidence interval in each season. The forecasting errors in the horizontal axes are normalized, and the range from −0.5 to +0.5 is in the interval. This means that the forecasting is correct if the errors are in the range. As shown in Fig. 9, the confidence intervals become wider as the confidence levels



(a) Winter season (from November to mid-March)



(b) Summer season (from mid-March to mid-May)



(c) Rainy season (from mid-May to October)

Fig. 10. Histogram of forecasting error normalized with the width of the confidence interval.

become higher, from 50% to 80% and 90%. The forecasting errors are normalized as shown in Fig. 10, and their distributions become sharper instead as the levels become higher. In winter, the distributions of normalized errors are almost symmetrical as shown in Fig. 10(a). In other seasons, summer and rainy seasons, the normalized errors inside the confidence intervals are slightly shifted to the negative direction as shown in Figs. 10(b) and 10(c). The shifting means that the forecasted solar irradiances are slightly smaller than the observation when the observations are in the confidence interval. The medium values of the seasons, whose quantiles are 50%, are negative as shown in Fig. 5. The shifting shown in Figs. 10(b) and 10(c) corresponds to the negative values.

Table III. Observed relative frequency of forecasting errors normalized with the width of the confidence interval.

(Forecasting error)/(width of CI)		~ -0.5	$-0.5 \sim +0.5$	$+0.5 \sim$
Confidence interval		Outside (under-forecasting)	Inside	Outside (over-forecasting)
Season	Specified confidence level	Observed relative frequency of forecasting error (%)		
Winter	50%	34.9	44.0	21.1
	80%	21.7	58.2	20.1
	90%	8.4	79.2	12.4
Summer	50%	27.8	48.8	23.4
	80%	20.7	64.3	15.0
	90%	8.5	82.7	8.8
Rainy season	50%	30.6	47.2	22.2
	80%	22.7	60.0	17.3
	90%	9.7	79.7	10.6

Observed frequencies of the forecasting errors normalized with the width of the confidence interval are listed and plotted in Table III and Fig. 11. The green lines in the figure indicate the ideal case in which observed frequencies of the forecasting errors are the same as the specified confidences. The quantiles inside the confidence interval correspond to the actual confidence levels of the probability forecasts. The actual levels are lower than the specified ones in any case as indicated in the table and the figure. In this study, the confidence interval of the forecasting is evaluated as a combination of the most probable forecasted irradiance and the width of the interval. Each of them is evaluated separately and also contains an error. Therefore, the total error is expected to be larger when combining them to evaluate the interval. To reduce the decrease in the actual confidence level, higher accuracies of the most probable irradiance and the width of the interval are required. A series of studies by the authors, optimization of parameterization options,[17] and the introduction of the Kalman filter as a post-processing,[22] have improved the forecasting accuracy of WRF. On the other hand, there is still room for improvement in the evaluation of confidence intervals. In this study, there are only three ensemble members, and the accuracy of their spreads is not considered to be high.

4. Conclusions

The probability prediction of solar irradiance in Thailand is performed with ensemble forecasting.

In solar irradiance forecasting with NWP, there is a limit to the improvement of forecasting accuracy through model improvement, etc. due to the uncertainty in the representation of the models and weather forecasting products. It would be beneficial to the users if the forecasting could provide its reliability as well. The probability prediction with ensemble forecasting provides both the forecasting value and its reliability.

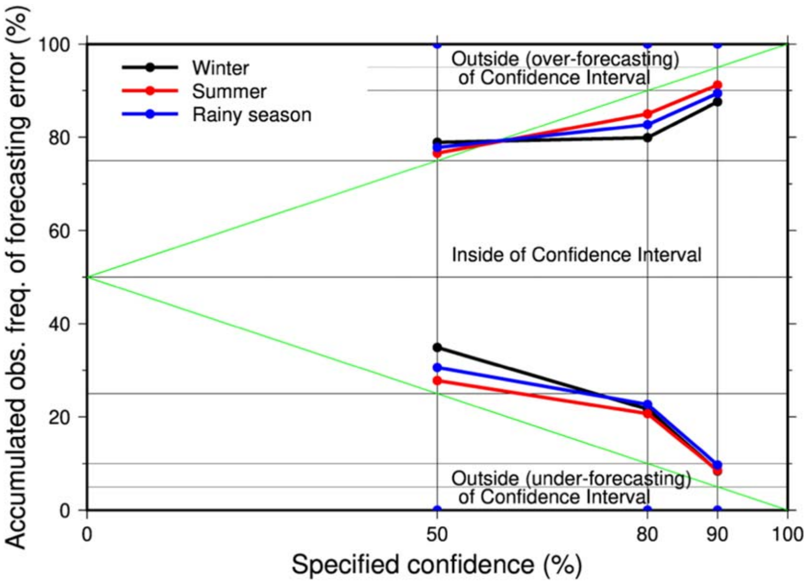


Fig. 11. Accumulated observed relative frequencies of forecasting errors in each season for the specified confidences. The data corresponds to that in Table III. Green lines denote the case in which the observed relative frequencies of forecasting errors are the same as the specified confidences.

In this study, the numerical weather model WRF, is employed for forecasting solar irradiance. The optimized physical parameterization options in the WRF are used to improve the accuracy of the forecasting in Thailand. The nonlinear Kalman filter is also introduced as post-processing of WRF and adjusts the irradiance computed with WRF. For ensemble forecasting, the LAF method is employed. It is easy to construct ensemble members with this method, and this method is efficient, especially for short-period forecasting. The validation of the accuracy of the forecasting was performed with different scopes of forecasting, and the intraday forecasting is selected as the most probable value of the forecasted irradiance. The relationship between spreads, which relate to the variance of ensemble members, and forecasting errors is investigated, and the confidence interval for each spread is evaluated. Finally, the probability prediction of the solar irradiance was performed with the most probable values and the confidence intervals. The observation was captured within the confidence interval of the prediction. However, the actual confidence level evaluated from the probability prediction is lower than the specified one. To explain this, the small number of ensemble members is pointed out. The results of this study indicate that ensemble forecasting of solar irradiance in Thailand works precisely.[44]

1) T. Stetz, J. Von Appen, F. Niedermeyer, G. Scheibner, R. Sikora, and M. Braun, "Twilight of the grids: The impact of distributed solar on Germany's energy transition," IEEE Power Energy Mag. **13**, 50 (2015).
2) A. Detollenaere, J. Van Wetter, and G. Masson, Snapshot of Global PV Markets 2020 IEA-PVPS T1-37:2000 IEA, 2020 20.
3) H. M. Diagne, M. David, P. Lauret, and J. Boland, "Solar irradiation forecasting: State-of-the-art and proposition for future developments for small-scale insular grids," *World Renewable Energy Forum, WREF 2012, Including World Renewable Energy Congress XII and Colorado Renewable Energy Society (CRES) Ann. Conf.* 2012) Vol. 6, p. 4802.
4) D. Heinemann, E. Lorenz, and M. Girodo, "Solar irradiance forecasting for the management of solar energy systems," *American Solar Energy Society - Solar 2006: 35th ASES Ann. Conf., 31st ASES National Passive Solar Conf., 1st ASES Policy and Marketing Conf., ASME Solar Energy Division Int. Solar Energy Conf.* 2006) Vol. 1, p. 1.
5) E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **2**, 2 (2009).
6) V. Lara-Fanego, J. A. Ruiz-Arias, D. Pozo-Vázquez, F. J. Santos-Alamillos, and J. Tovar-Pescador, "Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain)," Sol. Energy **86**, 2200 (2012).
7) S. Shimada, Y. Liu, H. Xia, J. Yoshino, T. Kobayashi, A. Itagaki, T. Utsunomiya, and J. Hashimoto, "Accuracy of solar irradiance simulation using the WRF-ARW model," J. Jpn. Sol. Energy Soc. **38**, 41 (2012), [in Japanese].
8) A. W. Aryaputera, D. Yang, and W. M. Walsh, "Day-ahead solar irradiance forecasting in a tropical environment," J. Sol. Energy Eng. ASME **137**, 051009 (2015).
9) P. Chinnavornrungsee, S. Kittisontirak, N. Chollacoop, S. Songtrai, K. Sriprapha, P. Uthong, J. Yoshino, and T. Kobayashi, "Solar irradiance prediction in tropics using weather forecasting model," Jpn. J. Appl. Phys. **62**, SK1050 (2023).
10) Department of Alternative Energy Development and Efficiency, *Thailand PV Status Report 2018* (Department of Alternative Energy Development and Efficiency, Ministry of Energy, Bangkok, 2019).
11) A. H. Sobel, "Tropical weather," Nat. Educ. Knowl. **3**, 2 (2012).
12) New Straits Times, Weather in tropics hard to predict, 17 August 2019, https://nst.com.my/news/nation/2019/08/513461/weather-tropics-hard-predict (accessed December 13, 2022).
13) M. Shenoy, P. V. S. Raju, and J. Prasad, "Optimization of physical schemes in WRF model on cyclone simulations over Bay of Bengal using one-way ANOVA and Tukey's test," Sci. Rep. **11**, 24412 (2021).
14) J. W. Yoon, S. Lim, and S. K. Park, "Combinational optimization of the WRF physical parameterization schemes to improve numerical sea breeze prediction using micro-genetic algorithm," Appl. Sci. **11**, 11221 (2021).
15) R. Arasa, I. Porras, A. Domingo-Dalmau, M. Picanyol, B. Codina, M.ª Ángeles González, and J. Piñón, "Defining a standard methodology to obtain optimum WRF configuration for operational forecast: Application over the Port of Huelva," Atmos. Clim. Sci. **6**, 329 (2016).
16) H. Verbois, R. Huva, A. Rusydi, and W. Walsh, "Solar irradiance forecasting in the tropics using numerical weather prediction and statistical learning," Sol. Energy **162**, 265 (2018).
17) D. Harada, P. Chinnavornrungsee, S. Kittisontirak, N. Chollacoop, S. Songtrai, K. Sriprapha, J. Yoshino, and T. Kobayashi, "Optimization of numerical weather model parameterizations for solar irradiance prediction in tropics," Jpn. J. Appl. Phys. [Submitted].
18) Japan Meteorological Agency, *Training Textbook 2018 for Numerical Forecasting—Tenth Generation Numerical Analysis and Forecasting Systems and Fundamentals of Numerical Forecasting, Numerical Forecasting Explanatory Materials* (Japan Meteorological Agency, Japan, 2018) Numerical Forecasting Manual, Vol. 51, p. 168 [in Japanese].
19) J. Yang, M. Astitha, E. N. Anagnostou, and B. M. Hartman, "Using a Bayesian regression approach on dual-model windstorm simulations to improve wind speed prediction," J. Appl. Meteor. Climatol. **56**, 1155 (2017).

20) M. Diagne, M. David, J. Boland, N. Schmutz, and P. Lauret, "Post-processing of solar irradiance forecasts from WRF model at Reunion Island," Sol. Energy **105**, 99 (2014).
21) A. Rincón, O. Jorba, J. M. Baldasano, and L. D. Monache, "Short-term irradiance forecasting based on post-processing tools applied on WRF meteorological simulations," Geophys. Res. Abstr. **13**, EGU2011 (2011).
22) N. Moriai, P. Chinnavornrungsee, S. Kittisontirak, N. Chollacoop, S. Songtrai, K. Sripapha, J. Yoshino, and T. Kobayashi, "Post-processing of solar irradiance predicted with a numerical weather model," Jpn. J. Appl. Phys. [Submitted].
23) M. S. Hossain and H. Mahmood, "Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast," IEEE Access **8**, 10 (2020).
24) K. Zhang and G. Zou, "Photovoltaic output prediction method based on weather forecast and machine learning," J. Phys.: Conf. Ser. **2320**, 012032 (2022).
25) A. Callado, P. Escribà, J. A. García-Moya, J. Montero, C. Santos, D. Santos-Muñoz, and J. Simarro, *Ensemble Forecasting* (IntechOpen, London, 2013), https://doi.org/10.5772/55699.
26) European Centre for Medium-Range Weather Forecasts, Forecast charts and data, https://ecmwf.int/ (accessed December 24, 2022).
27) National Centers for Environmental Prediction, National Weather Service, https://weather.gov/ncep/ (accessed December 24, 2022).
28) Japan Meteorological Agency, Weather, Climate & Earthquake Information, https://jma.go.jp/jma/en/menu.html) (accessed December 24, 2022).
29) Y. Liu, S. Shimada, J. Yoshino, T. Kobayashi, Y. Miwa, and K. Furuta, "Ensemble forecasting of solar irradiance by applying a mesoscale meteorological model," Sol. Energy **136**, 597 (2016).
30) P. Singla, M. Duhan, and S. Saroha, "An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and BiLSTM deep learning network," Earth Sci. Inform. **15**, 291 (2022).
31) W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, D. Barker, M. G. Duda, X. Huang, W. Wang, and J. G. Powers, *A Description of the Advanced Research WRF Version 3* (University Corporation for Atmospheric Research., Boulder, CO, USA, 2008), No. NCAR/TN-475+STR, https://doi.org/10.5065/D68S4MVH.
32) N. Salvação and C. Guedes Soares, "An operational forecast system for wind conditions in the Portuguese pilot area of Aguçadoura," *Renewable Energies Offshore* (Taylor & Francis, London, 2015) p. 229, https://doi.org/10.1201/b18973.
33) S. C. Pryor, T. J. Shepherd, R. J. Barthelmie, A. N. Hahmann, and P. Volker, "Wind farm wakes simulated using WRF," J. Phys.: Conf. Ser. **1256**, 012025 (2019).
34) Y. Lin and B. A. Colle, "A new bulk microphysical scheme that includes riming intensity and temperature-dependent ice characteristics," Mon. Wea. Rev. **139**, 1013 (2011).
35) M. D. Chou and M. J. Suarez, "A solar radiation parameterization for atmospheric studies," NASA Tech. Memo **15**, 40 (1999).
36) M. D. Chou, M. J. Suarez, X. Z. Liang, and M. M. H. Yan, "A thermal infrared radiation parameterization for atmospheric studies," NASA Tech. Memo **19**, 68 (2001).
37) J. Dudhia, "Numerical study of convection observed during the winter monsoon experiment using a mesoscale two- dimensional model," J. Atmos. Sci. **46**, 3077 (1989).
38) J. S. Kain, "The Kain-Fritsch convective parameterization: an update," J. Appl. Meteor. Climatol. **43**, 170 (2004).
39) S.–Y. Hong, Y. Noh, and J. Dudhia, "A new vertical diffusion package with an explicit treatment of entrainment processes," Mon. Wea. Rev. **134**, 2318 (2006).
40) NCEP GDAS/FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids, https://rda.ucar.edu/datasets/ds083.3/ (accessed December 13, 2022).
41) R. N. Hoffman and E. Kalnay, "Lagged average forecasting, an alternative to Monte Carlo forecasting," Tellus A **35**, 100 (1983).
42) W. Ebisuzaki and E. Kalnay, "Ensemble experiments with a new lagged average forecasting scheme," WMO Res. Activities Atmos. Ocean. Model. Rep. **15**, 308 (1991).
43) Y. Miyamoto, Y. Kajikawa, R. Yoshida, T. Yamaura, H. Yashiro, and H. Tomita, "Deep moist atmospheric convection in a subkilometer global simulation," Geophys. Res. Lett. **40**, 4922 (2013).
44) K. Zhang and G. Zou, J. Phys.: Conf. Ser. **2320**, 012032 (2022).

