Solar Energy 162 (2018) 196–204

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener

A diagnostic for advance detection of forecast busts of regional surface solar radiation using multi-center grand ensemble forecasts

Fumichika Uno[a,d,*], Hideaki Ohtake[a,d], Mio Matsueda[b,c], Yoshinori Yamada[d]

[a] Research Center for Photovoltaics, National Institute of Advanced Industrial Science and Technology, Japan
[b] Center for Computational Sciences, University of Tsukuba, Japan
[c] Department of Physics, University of Oxford, UK
[d] Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan

ARTICLE INFO

Keywords:
Ensemble forecast
Ensemble spread
Large forecast error detection

ABSTRACT

Large forecast errors (forecast busts) for surface solar radiation (SSR) and therefore photovoltaic power generation may lead to either a shortage of power supply or production of excessive surplus power. Ensemble forecasting with numerical weather prediction (NWP) models has been developed to reduce forecast errors by taking the average of individual forecasts and to consider forecast uncertainty and reliability by generating a probabilistic forecast of meteorological fields. A multi-center grand ensemble (MCGE) is recognized as a useful technique for further reducing the uncertainty of a weather forecast. An ensemble mean of MCGE (EM_g) has smaller forecast error than an ensemble mean of a single NWP center (EM). Moreover, the lognormal ensemble spread of MCGE ($LNES_g$) and single-NWP-center ensembles (LNES) relate to the forecast error, and can be used as a predictor of reliability for the weather forecast. 1- to 6-day ensemble forecasts at four leading NWP centers (European Centre for Medium-Range Weather Forecasts: ECMWF, Japan Meteorological Agency: JMA, National Centers for Environmental Prediction: NCEP, and the UK Met Office: UKMO) were used to detect the forecast busts of daily SSR over the Kanto Plain in central Japan in a day-ahead regional forecast operated by the Japan Meteorological Agency (JMA-MSM).

The magnitude of the forecast error of the EM_g was found to be comparable with that of the JMA-MSM. The correlations between the forecast error coefficient (Fc) and LNESg in winter season were higher than summer season. In the top 10%, 5% and 1% forecast busts in five winter months, the Receiver Operating Characteristic (ROC) scores of the MCGE in 1- to 6-day ahead forecast indicated statistical significance. The $LNES_g$ can therefore be a valuable predictor for detection of forecast busts in the regional forecast.

1. Introduction

Weather forecasting is an important technology for energy management. Prediction of surface solar radiation (SSR) using numerical weather prediction (NWP) models achieves better forecast accuracy than statistical models for forecast lead time in the range of several hours to several days (Diagne et al., 2013). One- to several-day forecasts are important in planning for an appropriate reserve capacity (e.g., unit commitment, pumped storage generation, and battery management).

Energy management using information from weather forecasts is vulnerable to the risks of both shortage of power supply and production of excessive surplus power owing to forecasts with large errors (forecast busts) of NWP models (e.g., Lorenz et al., 2011; Ohtake et al., 2016). Mean bias error (MBE) and root mean square error (RMSE) are commonly used to evaluate the forecast error. However, there are few examples in the literature of the assessment of forecast busts. Prediction of forecast busts is important for energy management systems.

Regional forecasts use global forecasts as initial and lateral boundary conditions. In previous studies, regional forecasts have been used to forecast SSR and grid-connected photovoltaic power generation, with machine learning and model output statistics (MOS) used to perform post-processing (Cornaro et al., 2015; Fonseca et al., 2012). It is well-recognized that post-processing reduces the forecast error from weather forecast models (e.g., Pelland et al., 2013; Sperati et al., 2016). Pierro et al. (2016) indicated the benefits of multi-center grand ensembles (MCGE) for a day-ahead forecasts. They used the Integrated Forecast System (IFS) model and the Weather Research and Forecast (WRF) model. The propagation of errors from the global (IFS model) to regional (WRF model) forecast occurs because the regional forecast model uses global forecast data as the initial and boundary conditions. Global forecast errors propagate to each subsequent step in, for instance, regional forecasts using a regional NWP model (dynamical

* Corresponding author at: Research Center for Photovoltaics, National Institute of Advanced Industrial Science and Technology, Japan.
E-mail address: uno.fumichika@aist.go.jp (F. Uno).

https://doi.org/10.1016/j.solener.2017.12.060
Received 17 February 2017; Received in revised form 25 December 2017; Accepted 28 December 2017
Available online 19 January 2018

Nomenclature

ECMWF	European Centre for Medium-Range Weather Forecasts
EM	ensemble mean
EMg	grand ensemble mean
ES	ensemble spread
ESg	grand ensemble spread
Fc	forecast error coefficient
IFS	Integrated Forecast System
JMA	Japan Meteorological Agency
JMA-GSM	Global Spectral Model by JMA
JMA-MSM	meso-scale model by JMA
LNES	lognormal ensemble spread
$LNES_g$	lognormal grand ensemble spread
MBE	mean bias error
MCGE	multi-center grand ensemble
MOS	model output statistics
NCEP	National Centers for Environmental Prediction
NES	normalized daily ES
NHM	non-hydrostatic model
NWP	numerical weather prediction
RMSE	root mean square error
ROC	Receiver Operating Characteristic
RRTM	rapid radiation transfer model
SS	skill score
SSR	surface solar radiation
SV	singular vector
UKMO	Met Office
WRF	Weather Research and Forecast

downscaling) and point forecasts using statistical models (statistical downscaling). The physical model, such as the regional NWP model, statistical models, the machine learning method or MOS use the global forecast as their input data. Then the forecast accuracy of the physical or statistical models used for regional and point forecasts is affected by the weather conditions of the global forecast.

The forecast error and uncertainty are caused by the incompleteness of the NWP models, errors in initial and lateral conditions, and a chaotic atmosphere (Lorenz, 1963). The forecast error is amplified with increasing forecast lead time (Lorenz et al., 2016), and varies with the physical variables and forecast lead time (Garcia-Moya et al., 2011; Stensrud et al., 2000). The error is influenced by the physical schemes of the model and initial and lateral boundary conditions (Clark et al., 2008; Liu et al., 2016). Moreover, the error is related to reproducibility of weather conditions in NWP model (e.g., Ohtake et al., 2015; Ohba et al., 2016).

Global ensemble forecasts are produced by several NWP centers to reduce weather forecast uncertainty and for evaluating risk assessment (Matsueda et al., 2007; Palmer, 2002). The benefits of ensemble forecasts are: (1) the ensemble mean (EM) reduces forecast error and uncertainty; (2) the ensemble spread (ES) provides information on the predictability and reliability of the forecast. Note that the EM indicates the mean value of all ensemble members in a single-NWP center ensemble forecast, ES indicates the standard deviation of the ensemble members, and ensemble member refers to an individual forecast in the overall ensemble of forecasts. Previous studies have shown that there is a high correlation between the ES and forecast skill (Grimit and Mass, 2007; Whitaker and Loughe, 1998). Thorey et al. (2015) demonstrated the benefits of an ensemble forecast using an MCGE for solar radiation. The RMSE of the EM of MCGE (EM_g) of the averaged SSR between 0600 and 1200 UTC was found to be lower than that of the EM from single-NWP centers for a six-hour forecast in the European region. These results indicate the usefulness of the MCGE for renewable energy fields. However, the availability of the ES of an MCGE (ES_g) for predicting SSR has not been discussed sufficiently in previous studies.

The purpose of this study is to evaluate the reliability and predictability of using ES and ES_g in Japan for short- to medium-range forecasts. In particular, we developed the prediction of forecast busts in regional deterministic forecasts using the ES and ES_g. The method for predicting the daily forecast error from one to several days is evaluated. The proposed detection method of forecast busts may be useful for the management of the auxiliary power supply and the battery when forecast error is expected.

2. Data and method

Short-range (1- to 3-day ahead) to medium-range (4- to 6-day ahead) forecasts are evaluated in this study. Fig. 1 shows the domain of the regional forecast model and the location of the Kanto Plain in Japan. The area-averaged daily (24-h average) SSR around the Kanto Plain (Fig. 1b) was used for the evaluation of forecast data. The 1- (6-) days ahead forecast of SSR is the value averaged from 00 (120) to 24 (144) hours of the forecast lead time. The evaluation period is for 2014–2016. The period covers from January 1 (6) to December 31 (26) for the 1- (6-) day ahead forecast. In this study, we mainly show as a specific example of the results in 2015. Missing values are defined when there are gaps in the observations or forecast data corresponding to 20% of the total evaluation period. Two forecast data sets were used, one regional and one global.

2.1. Regional deterministic forecast

The regional forecast is from the Japan Meteorological Agency (JMA) using a non-hydrostatic model (NHM) (Saito et al., 2006), and is referred to in this study as the meso-scale model (hereafter JMA-MSM). The forecast lead time of the JMA-MSM ranges from 1 h up to 39 h. The model domain is about 20–50°N and 100–160°E (see Fig. 1a). The horizontal resolution of the JMA-MSM is 5 km. A 24-h (day-ahead) forecast is used in this study. The SSR on a land surface grid is calculated as the area-averaged SSR of the Kanto Plain.

2.2. Global ensemble forecast

Global ensemble forecasts from four leading NWP centers were used to provide SSR forecasts. The data set has been archived by TIGGE (The International Grand Global Ensemble) project (Swinbank et al., 2016). TIGGE archives the ensemble forecasts for global forecast models from 10 NWP centers with a delay of 2 days, so the data set is mainly valuable for several-day forecasts. The four leading NWP centers are as follows: the ECMWF (European Centre for Medium-Range Weather Forecasts, Europe), the JMA, the NCEP (National Centers for Environmental Prediction, United States of America), and the UKMO (Met Office, United Kingdom). The JMA global ensemble forecast archived by TIGGE is the Global Spectral Model (Kanamitsu et al., 1983) (hereafter, JMA-GSM).

The leading four NWP centers provide the SSR forecast data for more than one week starting from the same initial forecast time. The SSR forecast data are archived as 6-h accumulated values. The area-averaged SSR of the ensemble forecast in Kanto Plain is calculated from six grid point values (see the plus symbols in Fig. 1b). We eliminate the other NWP centers because of their different initial forecast time or unarchived SSR forecast data. Moreover, several NWP centers have large RMSE on the SSR forecast in the European region (Thorey et al., 2015).

The global ensemble forecast models of the different NWP centers have different configurations. Table 1 shows an overview of the forecast models from the four forecast providers. Note that the ensemble size

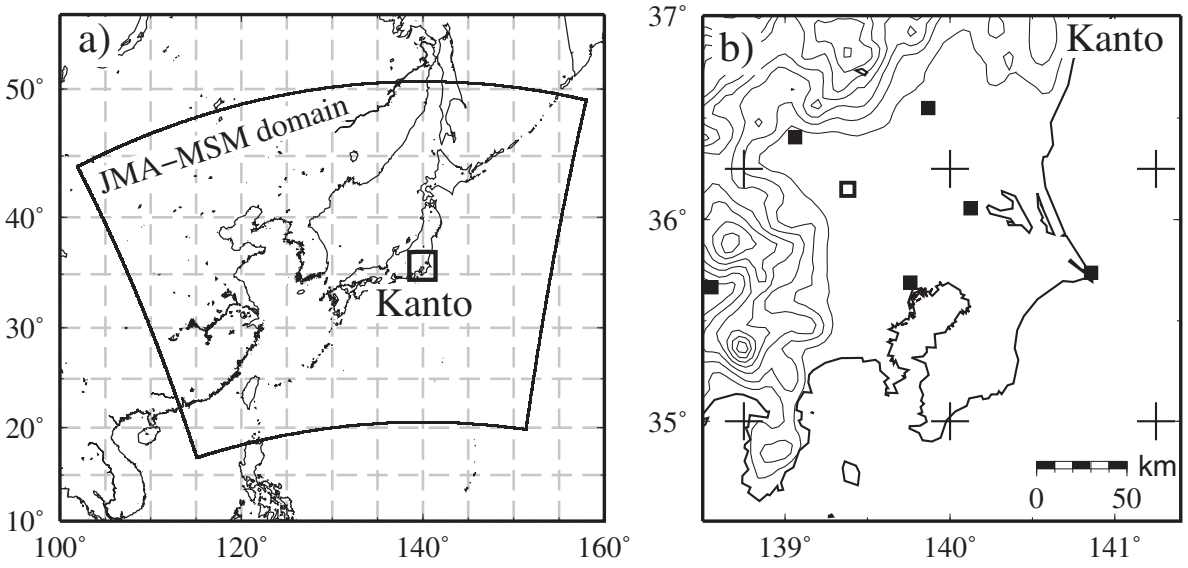

Fig. 1. Domain of the regional forecast model (a) and location of Kanto Plain (b). The solid squares indicate the locations of six JMA stations, and the open square is the temporary station (Kumagaya station); plus symbols are the central coordinates of interpolated global forecast data. The contour indicates the orography drawn at a 250 m interval.

Table 1
Overview of model configuration of NWP centers.

	ECMWF	JMA-GSM	NCEP	UKMO
Spatial resolution [km], (°)	40, (0.3 × 0.3)	140, (1.25 × 1.25)	110, (1.0 × 1.0)	40, (0.3 × 0.45)
Ensemble size	51	27	21	12 (24)

indicates the number of ensemble members, made up of one deterministic (unperturbed) and all the perturbed forecasts. The spatial resolution of ECMWF, NCEP and UKMO were interpolated to the same resolution as JMA-GSM (1.25°). The model physical processes and initial perturbation methods also vary between NWP centers. Note that the ensemble size of UKMO changed 24–12 in November. And JMA-GSM had been beginning to archive form February 25, 2014.

2.3. Validation data

The SSR data from surface observations are used as the reference data (see squares in Fig. 1b). The observational data are from six JMA operational observation sites (closed squares) and one temporary site (see Fig. 1b open square) at Kumagaya (Ohtake et al., 2013, 2015). Additionally, the SSR data from the ERA-Interim (Dee et al., 2011) and JRA-55 (Kobayashi et al., 2015) reanalysis were used to evaluate the bias at each of the NWP centers. The SSR of the reanalysis data has homogeneous spatial and temporal accuracy.

3. Method

3.1. Forecast error coefficient

The forecast error coefficient (Fc) of the SSR for the regional forecast is defined below. Fc is the normalized forecast error of the daily average SSR for the JMA-MSM, and is calculated as

$$Fc = \frac{I_f - I_o}{ext}, \tag{1}$$

where I_f is area-averaged SSR of the JMA-MSM, I_o is averaged SSR at the seven stations, and ext is extra-terrestrial radiation (Badescu, 2014). The ext is used as a normalizing parameter for removing seasonal variations of the forecast errors of daily SSR.

3.2. Multi-center grand ensemble means and spreads

The grand ensemble quantities EM_g and ES_g are often computed as a weighted average as a function of decreasing model bias and uncertainty. In previous studies the weights were determined by the forecast model (Jonson and Swinbank, 2009) or weather pattern (Greybush et al., 2008). However, the EM_g in this study was evaluated as an un-weighted average (the reason for this is explained in Section 4.1). The EM_g is calculated as follows:

$$EM_g = \frac{1}{N}\sum_{k=1}^{N} EM_k \quad N = 4. \tag{2}$$

Here, EM_k is EM at each of the NWP centers, and N is the number of NWP centers.

Next, for bias correction, ES_g is weighted by each month and for each NWP center:

$$ES_g = \frac{1}{N}\sum_{k=1}^{N} \frac{NES_k}{\overline{NES_{m,k}}} \quad N = 4, \tag{3}$$

where

$$NES_k = \frac{ES_k}{\overline{ES_{m,k}}}, \tag{4}$$

here, m is the month, NES_k is normalized daily ES, and $\overline{NES_{m,k}}$ is the normalized monthly NES_k. The ES and ES_g follow the well-known lognormal distribution. The lognormal ES and ES_g (hereafter, LNES and $LNES_g$) are predictors of the forecast skill (Whitaker and Loughe, 1998). The $LNES_g$ is calculated as follows:

$$LNES_g = \ln(ES_g). \tag{5}$$

3.3. Evaluation methods of forecast error and skill

There are several statistics that are often utilized for evaluating the forecast skill of a prediction system (e.g., Grimit and Mass, 2007). This study uses MBE and RMSE. The forecast error is the difference between the forecast value (y_k) and observed value (o_k) of SSR,

$$\mathrm{MBE} = \frac{1}{N}\sum_{k=1}^{N}(y_k - o_k) \tag{6}$$

and

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(y_k - o_k)^2} \tag{7}$$

where N is the total number of samples, y_k is the area-averaged daily SSR on Kanto Plain forecast by JMA-MSM and each NWP center. Here also, o_k indicates average daily SSR at the seven observation sites on Kanto Plain (see Fig. 1b).

Murphy (1988) suggested the forecast skill score (SS) using the MBE of the forecast to be evaluated and of the reference forecast. We used the persistence model as the reference forecast. The persistence model assumes that the conditions at the time of the forecast will not change. In this study, the smart persistence model assumes that the clear sky index (I_o/I_{clear}) on the next day is the same as the current day. The I_{clear} is clear sky radiation on the ground using Kumar model (Gueymard, 2012). Coimbra et al. (2013) also suggest a SS using the RMSE, $SS = 1 - RMSE_f/RMSE_p$, where $RMSE_f$ and $RMSE_p$ are the RMSE of the NWP and persistence forecasts, respectively.

3.4. Verification method of detectability of forecast busts

We evaluated the Receiver Operating Characteristic (ROC) curve and score (Wilks, 2005) to verify the detectability of forecast busts using LNES and $LNES_g$. The ROC curve is shown as a set of pairs of hit rate and false alarm rate. The hit rate (T/X) in this study is the number of detected forecast busts (T) relative to the total number of forecast busts that occur (X). The false alarm rate (F/N) in this study is also a relative frequency. F is the number of times the forecast bust event was forecast but did not occur, and N is the total number of times the bust did not occur. The ROC score is the area under the ROC curve. The predictor has no forecast skill when the ROC score is less than 0.5 (the ROC curve located the right side of a diagonal line). A perfect forecast has ROC score = 1.0. We used the T-test for evaluating the significance level (P-value) of the ROC scores.

4. Results

4.1. Bias of numerical weather prediction models

The SSR of each NWP model has an individual bias. We evaluated the biases of area-averaged SSR in the JMA-MSM domain (see Fig. 1a). Fig. 2 shows the seasonal variation of the monthly SSR for the four NWP models (colored circles). Additionally, we illustrate the monthly SSR from the ERA-Interim and JRA-55 reanalysis data (solid and dashed lines). The NCEP SSR forecast value was larger, and the UKMO smaller, than the others. The JMA-GSM and ECMWF were in agreement with ERA-Interim. We evaluated the difference in the monthly SSR forecast between ERA-interim and each NWP center. The percentage differences relative to the ERA-interim monthly average for ECMWF, JMA-GSM, NCEP, and UKMO were less than 5%, 5%, 15%, and 8%, respectively. The differences had no clear seasonal variation (figure not shown). The NCEP was closer to JRA-55 than to ERA-interim. It was not possible to determine whether the NCEP had a positive bias. These results indicate that it is difficult to evaluate the weight for calculating weighted average of multi-center grand ensemble. Therefore, EM_g was calculated as a simple averaged (un-weighted) SSR for this study.

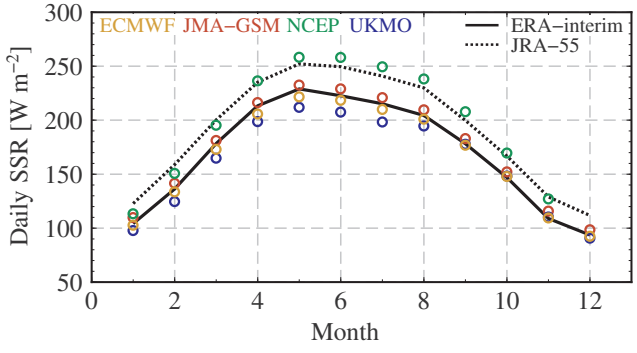

Fig. 2. Seasonal variation of daily SSR over the JMA-MSM domain in 2015. Colored circles (see key) indicate the EM of each NWP center. The black solid (dashed) line is the SSR from ERA-interim (JRA-55) reanalysis data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The errors in regional forecasts arise from three effects. First, the regional forecast error is affected by the forecast accuracy of the global forecast model used for the initial and boundary conditions. Second, the regional forecast model provides an additional error derived from the model uncertainty (e.g., iizumi et al., 2012). Third, the regional forecast model decreases the forecast error of the global forecast, because the regional forecast model can resolve smaller scale meteorological phenomena than the global forecast model. The disagreement in forecast busts between the regional and global forecasts is thought to be caused by the second and third effects.

4.2. Forecast error of ensemble mean

We evaluated the monthly forecast error for EM_g (Fig. 3). Fig. 3 shows the monthly RMSE for the day-ahead forecast using JMA-MSM and 1- to 6-day ahead forecasts from the four NWP centers. The JMA-MSM produced high forecast error in the summer season. The RMSE of EM_g increased consistently with the forecast lead time. In June and July, the RMSE of EM_g in 1- and 2-day forecasts was smaller than that of

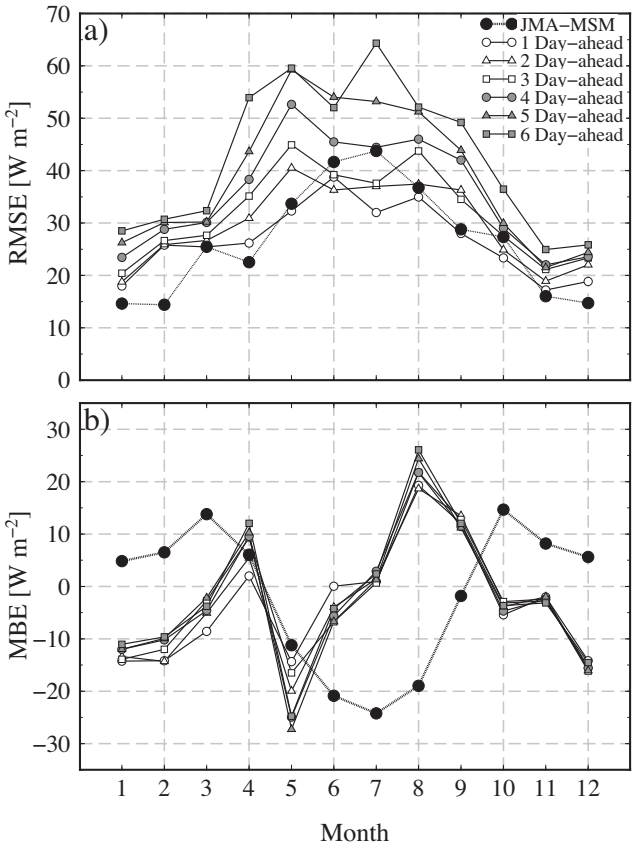

Fig. 3. Monthly RMSE (a) and MBE (b) values of EM_g from 1- to 6-day ahead forecasts and JMA-MSM day-ahead forecast in 2015. Solid circles indicate the JMA-MSM, other symbols (see key) show the EM_g for 1- to 6-day ahead forecasts.

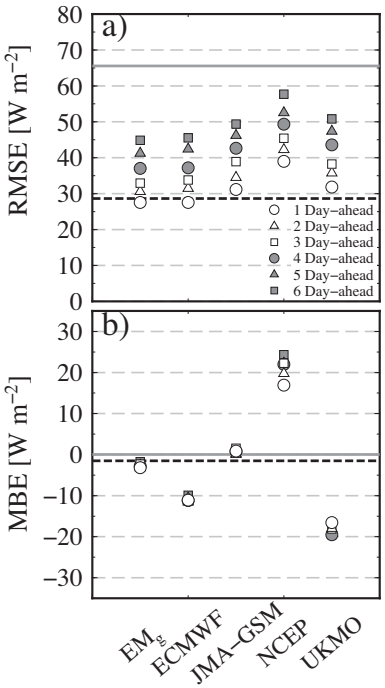

Fig. 4. Annual forecast errors from 1- to 6-day ahead forecasts. RMSE (a) and MBE (b) values of EM_g and EM for each NWP center in 2015. Thick dashed lines are the RMSE and MBE from the JMA-MSM, and the gray solid lines are the RMSE and MBE from the persistence forecast. Symbols as in Fig. 3 but without JMA-MSM.

JMA-MSM in a day-ahead forecast. This is caused by the large negative bias of the JMA-MSM in summer (Fig. 3b). The MBE of the JMA-MSM had negative (positive) values in winter (summer). The MBE of the EM_g had positive biases in April, July, August and September, and negative biases in other months. The MBE of JMA-MSM and EM_g had different seasonal variations. The seasonal variation of the JMA-MSM MBE is thought to be caused by the regional forecast model (JMA-NHM), because the MBE in the global forecast of the JMA-GSM has a similar seasonal variation to the EM_g (figure not shown).

Fig. 4 shows the annual forecast error for the 1- to 6-day ahead forecasts. The broken lines indicate the forecast error of the JMA-MSM for the day-ahead forecasts. The gray lines indicate the persistence forecast error for the day-ahead forecast. We mainly discuss the forecast errors for the day-ahead forecast. The RMSE (Fig. 4a) and MBE (Fig. 4b) in JMA-MSM are 28.6 W m^{-2} and −1.5 W m^{-2}, respectively. In Fig. 4a, the EM from the ECMWF and EM_g were close to the JMA-MSM forecast, with values of 27.5 W m^{-2} and 27.6 W m^{-2}, respectively. In Fig. 4b, the MBE of the JMA-GSM EM was 0.9 W m^{-2}, smaller than that of JMA-MSM. The MBE of EM_g and EM from the ECMWF, UKMO and NCEP were −3.18, −11.1, −16.6, and 16.9 W m^{-2}, respectively. The MBE had different biases (positive/negative) for each of the NWP centers. ECMWF has the lowest RMSE, but JMA-GSM has the lowest MBE. The forecast error in EM_g is little different from these centers. The SS of JMA-MSM, EM_g, ECMWF, JMA-GSM, NCEP, and UKMO were 56.7%, 58.3%, 58.3%, 52.8%, 41.0%, and 51.8%, respectively.

The impact of forecast lead time on the RMSE is small in the EM_g and EM. The RMSE of the EM from the NCEP had a higher value than the other NWP centers, and all NWP centers had a similar increase in RMSE with forecast lead time. Conversely, there were differences in the increase of MBE with forecast lead time between the NWP centers. The increases between the 1- and 6-day ahead forecasts from NCEP, UKMO, ECMWF, JMA-GSM, and for the EM_g were 7.2, 2.9, 1.1, 0.4, and 1.5 W m^{-2}, respectively.

4.3. Details of forecast busts

In the forecast busts in 2015, the regional forecast from the JMA-MSM was not comparable with the EM_g and EM from each NWP center. Table 2 shows the observed, estimated, and predicted results of area-averaged SSR for the worst five overestimated and underestimated values of JMA-MSM in the Kanto Plain. Note that the NCEP forecast was not available on December 17th, 2015. The observed SSR agreed with the EM for the 10 cases. The highest forecast bust occurred on October 10th, 2015, when the SSR from the JMA-MSM was 197 W m^{-2} and Fc was 0.28, while the observed SSR were 110 W m^{-2}. However, EM_g predicted 86 W m^{-2}, a higher forecast accuracy than JMA-MSM. In the 10 cases in Table 2, all EMs and EM_g were better than JMA-MSM. It is well known that the monthly and annual forecast accuracy of the EM_g is higher than that of individual EMs and the deterministic forecast (Thorey et al., 2015). However, the daily forecast accuracy of EM_g was not always higher than that of the EMs.

Fig. 5 shows the frequency of the absolute values of Fc (hereafter |Fc|) in 2015. In this study, the forecast busts are defined as the top 10% of |Fc| (see gray bars in Fig. 5), with a further classification into the top 10%, 5% and 1% of |Fc|. There is little difference in the frequency distribution of |Fc| in three years. We investigate the detectability of the forecast busts at the three threshold of |Fc| in the Section 5.2.

4.4. Ensemble spread of forecast busts

The ES is related to the forecast error and the reproducibility of the meteorological phenomena in the NWP model (Whitaker and Loughe, 1998). We investigate the relationship of ES and Fc in October, the month with the largest forecast bust in 2015, with Fc = 0.28 (see Overestimate in Table 2). Fig. 6 illustrates the time series of |Fc| for the

Table 2
Observed, estimated, and predicted SSR for the top five forecast busts (over- and under-estimates) in 2015.

				Observation [W m^{-2}]	Forecast [W m^{-2}]					
Month	Day	Fc	ext [W m^{-2}]	Surface	JMA-MSM	ECMWF	JMA-GSM	NCEP	UKMO	EM_g
Overestimate										
10	10	0.28	307	110	197	93	111	73	67	86
10	31	0.27	254	88	156	87	125	103	84	100
3	15	0.22	339	138	213	141	165	142	123	143
12	2	0.22	196	61	104	54	77	61	51	61
12	17	0.21	186	73	111	75	90	–	80	82
Underestimate										
7	24	−0.26	465	230	110	223	212	256	164	214
5	31	−0.22	476	298	194	244	236	282	231	248
1	14	−0.19	201	132	94	176	182	192	153	176
7	18	−0.19	471	173	84	105	105	113	90	103
8	18	−0.19	429	174	93	200	202	267	193	216

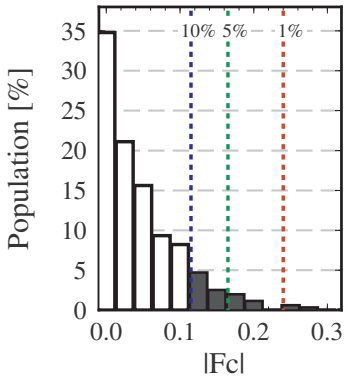

Fig. 5. Histogram of absolute values of Fc (|Fc|) in 2015. The gray bars indicate forecast busts. The blue, green and red dashed lines indicate the three thresholds of |Fc|. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

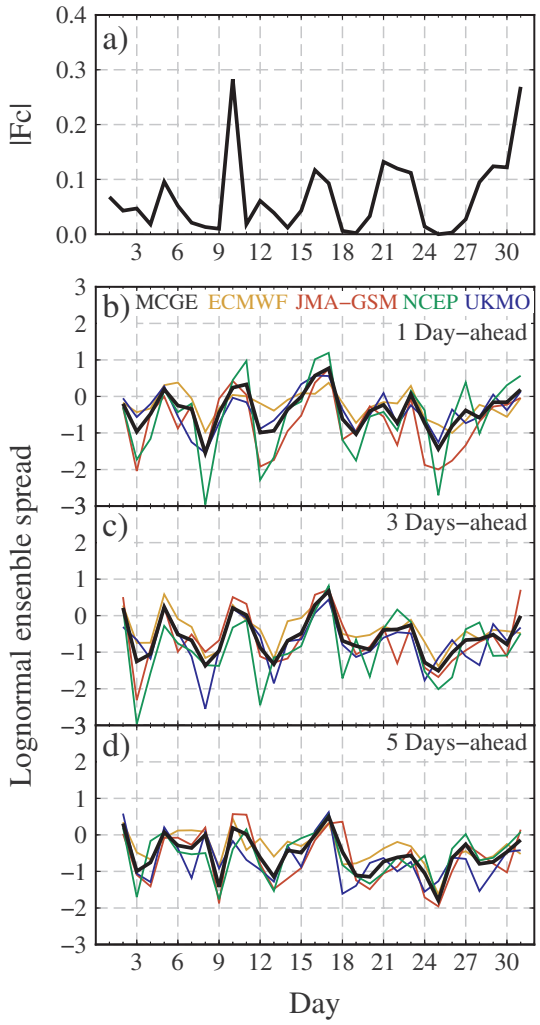

Fig. 6. Time series of |Fc| in JMA-MSM and ensemble spread in October 2015. (a) |Fc| in the JMA-MSM day-ahead forecast, and (b–d) LNES and $LNES_g$ in the 1-(b) 3-(c), and 5-(d) day ahead forecast. Thin colored lines indicate the LNES from each NWP center as shown in (b). Thick black lines in (b–d) are the MCGE $LNES_g$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

JMA-MSM (Fig. 6a), the LNES from each NWP center and the $LNES_g$ for 1-, 3-, and 5-day ahead forecasts (Fig. 6b–d) in October 2015. The LNES from each NWP center and the $LNES_g$ from the MCGE were consistent with |Fc|. The correlation coefficients (R) between $LNES_g$ and |Fc| for the 1-, 3-, and 5-day ahead forecasts were 0.68, 0.63, and 0.45, respectively. Note that the absolute value of Fc was used to evaluate the correlation, with no distinction between the positive and negative cases.

The |Fc| and $LNES_g$ were found to be consistent. For instance, on the 5th, 10th, 16th, 21st, and 23rd, the |Fc| was larger than 0.1 and $LNES_g$ was also relatively high, although the magnitude of LNES and $LNES_g$ did not always agree with that of |Fc|. The relationship was not consistent on the 1st, 22nd, 30th, and 31st. However, for the individual NWP centers, the LNES from the ECMWF (NCEP) had a large value on the 22nd (31st), and the relationship between LNES and |Fc| was consistent. Furthermore, the LNES for one or several NWP centers agrees with |Fc| over several days, if $LNES_g$ did not agree with |Fc|. The results imply the utility of multi-NWP-centers for the SSR forecast.

In the forecast busts on the 5th, 10th, and 16th, $LNES_g$ values remain high for the 1- to 5-day ahead forecasts. This shows that $LNES_g$ was a valuable predictor for the detection of forecast busts for several days ahead. Conversely, the $LNES_g$ value on the 23rd day is small in the 5-days ahead forecast, and increased with shortening forecast lead time. This forecast bust is difficult to detect in forecasts several days ahead.

The LNES and $LNES_g$ for short-range forecasts did not always have stronger relationships with |Fc| than those for the medium-range forecasts for individual cases. The $LNES_g$ for 3- and 5-day forecasts were higher than that for day-ahead forecasts for the forecast bust on October 10th, 2015. Naturally, the monthly and annual R is higher for short-range forecasts.

5. Discussion

5.1. Seasonal variation in sensitivity of ensemble spread for forecast busts

The LNES and $LNES_g$ have high positive correlation with |Fc| in October 2015. The seasonal variation of correlation between $LNES_g$ and |Fc| is shown in Fig. 7. The R was higher in winter than in summer, regardless of forecast lead time. However, the R in May was comparable with the winter months. Many months showed 95% or 90% statistical significance. Ten (seven) months achieved a 95% statistical significance for 1- (6-) day-ahead forecasts. The R from the UKMO was lower than the others for day-ahead forecasts (Fig. 7a). The small ensemble size in the UKMO is thought to be one of the causes of low R values. The seasonal variations of correlation coefficient show the same tendency in three years.

The annual R values for the $LNES_g$ for 1- to 6-day ahead forecasts are 0.40, 0.39, 0.38, 0.35, and 0.32, respectively. The seasonal variation of R values decreased as the forecast lead time increased from 1 to 6 days.

5.2. Detectability of forecast busts

The relationship between ensemble spread and |Fc| is statistically significant in winter seasons for the day-ahead forecast. Fig. 8 shows the ROC curve of LNES and $LNES_g$ for the day-ahead forecast for three years. The ROC curve in the top of 10%, 5%, and 1% forecast busts was evaluated using the all months in three years (Fig. 8a–c). The sample sizes were 108, 64, and 9 days, respectively. Inversely, the bottom figures (Fig. 8d–f) were used 5 months for evaluating ROC curve. These months were the top 5 months with high correlation coefficient between |Fc| and $LNES_g$. The sample sizes were 45, 23, and 5 days, respectively. The ROC scores in 5 months larger than that of all months in the forecast busts of 3 thresholds. For instance, The ROC score of MCGE of 5 months in the top of 10%, 5% and 1% were 0.69, 0.73, and 0.73, respectively. Otherwise, The ROC score of MCGE of all months were 0.60, 0.65, and 0.67, respectively.

We evaluated the ROC scores of MCGE and each NWP center for 1- to 6-day ahead forecast in the 5 months with high correlation

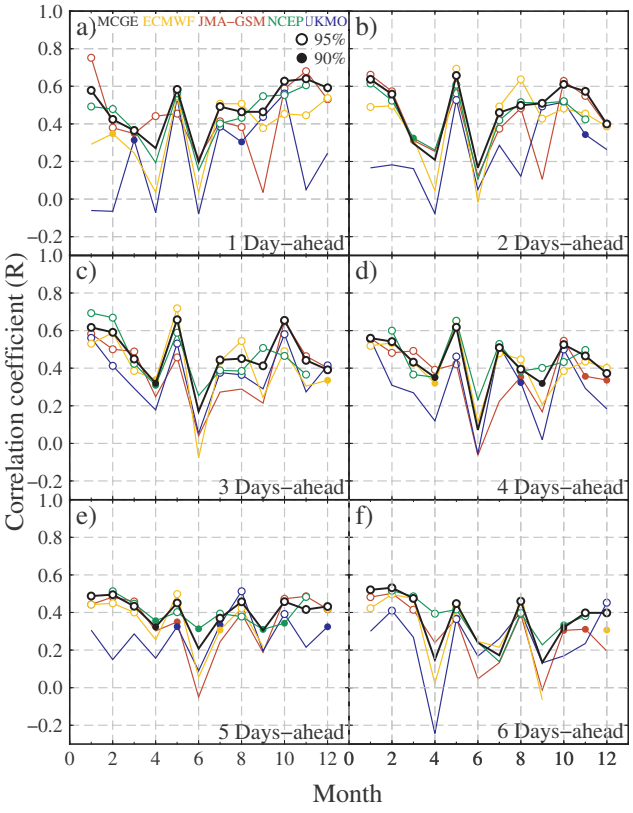

Fig. 7. Seasonal variation of the correlation between JMA-MSM forecast error (|Fc|) and LNES and $LNES_g$ for 1- to 6-day ahead forecasts. Colors are as in Fig. 6. The open and closed circles indicate months with 95% and 90% statistical significance, respectively. Note that some months are not shown because of missing data. For instance, ECMWF was missing in November for 6-day ahead forecasts, and NCEP was missing in December for 1- to 6-day ahead forecasts. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

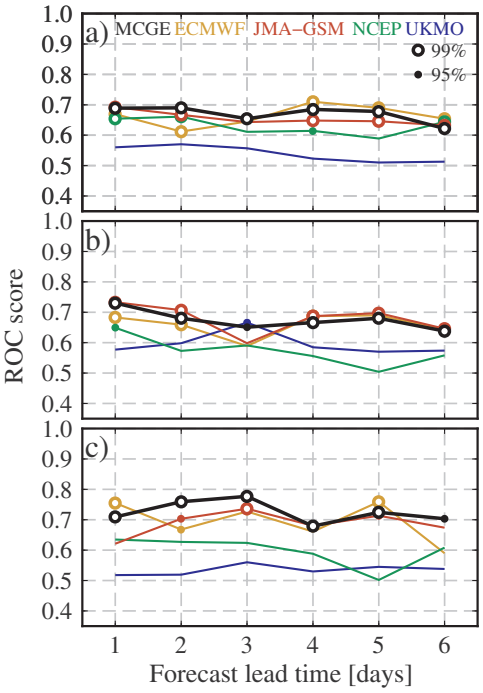

Fig. 9. Dependence of ROC scores on forecast lead time. It shows the top of 10(a), 5(b), 1%(c) forecast busts The open (closed) circles indicate 99 (95)% statistical significance of ROC scores. Color lines are as in Fig. 8. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

coefficient between |Fc| and $LNES_g$. (Fig. 9). The open (close) circle indicated the 99 (95)% statistical significance. In several cases, the ROC scores of single NWP center indicated the highest values (Fig. 9a–c). However, the ROC score of MCGE were higher than that of single NWP center in many cases. The difference of ROC scores each NWP centers in 1% forecast busts (Fig. 9c) were larger than the 10% and 5% forecast busts. The ROC scores of MCGE in the forecast busts of three thresholds were 99% or 95% statistical significance in 1- to 6-days ahead forecast. The magnitude of the 95% confidence interval of the ROC score in MCGE did not depend on the forecast lead time or the threshold of forecast busts (10%, 5%, and 1%). The average 95% confidence interval of MCGE was ± 0.04 in the 1- to 6-day ahead forecast and forecast busts of 3 threshold.

Matsueda and Nakazawa (2015) reported that the MCGE provides more reliable forecasts than single NWP center ensembles on the severe

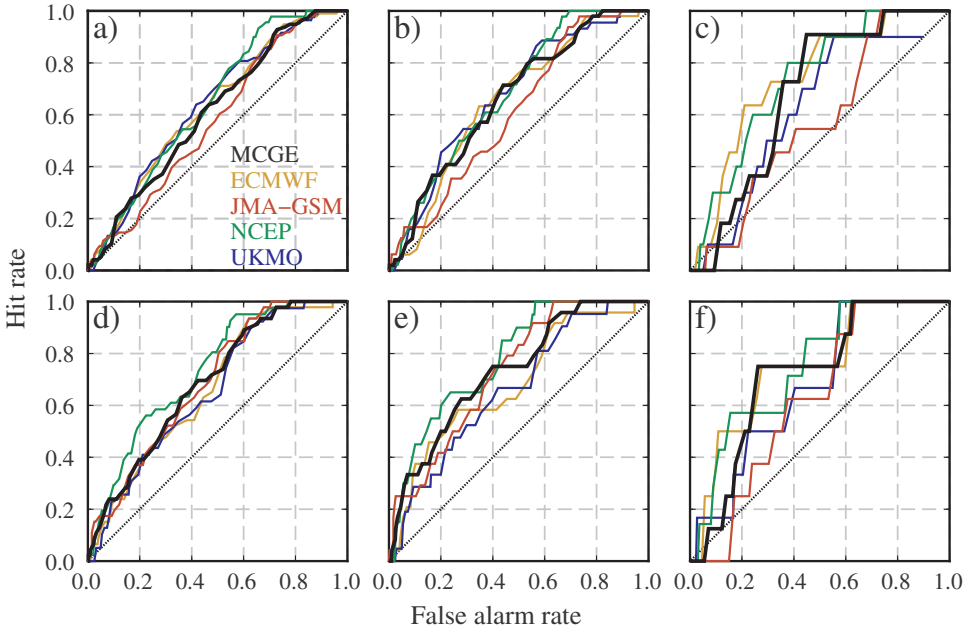

Fig. 8. The ROC curves in day-ahead forecast for three years. The evaluation period of the top figures are through a whole year, the bottom figures are winter seasons (5 months). These ROC curves are MCGE and NWP centers at the top of 10% (a, d), 5% (b, e), 1% (c, f) forecast busts in day-ahead forecast. Colored lines indicate the MCGE and each NWP center (see key). The diagonal dotted line corresponds to ROC score = 0.5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

weather event. In Section 4.2, there was no clear superiority of EM_g over single NWP center EMs. However, the ROC scores in MCGE were higher than single NWP center ensemble. The results indicate the value of MCGE. Also, the $LNES_g$ is a valuable predictor of forecast busts on an area-averaged daily SSR over a broad area (e.g., Kanto Plain). Therefore, $LNES_g$ in global ensemble forecasts can detect the forecast uncertainty for regional forecasts (e.g., JMA-MSM).

This proposed detection method is highly likely to detect forecast busts several times a year. The LNES and $LNES_g$ for global ensemble forecasts are valuable predictors of the forecast uncertainty and forecast error for regional forecasts. However, the method cannot be adopted directly in an energy management system because the false alert rate is not small enough. Moreover, the number of samples for evaluating the detectability is small. We need to increase the number of samples, for instance by extending the study period by several years.

6. Conclusions

This study shows the applicability of single-NWP center ensemble forecasts and the MCGE to a SSR forecast. We focus on the detectability of forecast busts for a regional forecast using the LNES and $LNES_g$. The results showed the following:

1. The EM_g values for area-averaged daily SSR evaluated by global ensemble forecasts have a comparable forecast accuracy to a regional forecast in the Kanto Plain area (about 400 km^2) for 1- and 2-day ahead forecasts.
2. LNES and $LNES_g$ were highly correlated with $|Fc|$ in 1- to 6-day ahead forecasts.
3. LNES and $LNES_g$ were statistically significant in several months, particularly in winter.
4. ROC score of MCGE showed 99% or 95% statistical significance for all forecast lead time in the top of 10%, 5%, and 1% forecast busts.

The area-averaged SSR over a broad area is more important than the SSR over a local area for energy management systems. Monthly and annual forecast errors (RMSE and MBE) in SSR showed little difference between global and regional forecasts for 1- and 2-day ahead forecasts.

The ensemble spreads in global forecasts was highly correlated with the forecast error of regional forecasts. Moreover, the $LNES_g$ has high ROC score in forecast busts. The uncertainty and forecast error of global forecasts propagate into the regional forecast. LNES and $LNES_g$ are valuable as predictors for evaluating the reliability and forecast skill of the regional forecast. The predictor has worldwide availability since it makes use of the global ensemble forecast.

We need to assess whether the predictor can apply to several-hour ahead forecasts, because this study only evaluated the index using daily SSR. It is likely that it will be difficult to detect precipitous changes of SSR (called “ramp” events) on diurnal variations (Zhang et al., 2015a, 2015b).

In the renewable energy field, ensemble forecasts, particularly MCGE, have not yet been widely used for constructing probabilistic forecasts. Introducing MCGE approaches into the field would be useful for more stable electricity provision.

Acknowledgments

This study is funded by the Japan Science and Technology Agency (JST) Core Research for Evolutional Science and Technology (CREST) project Grant Number JPMJCR15K1. We thank the staff of Kumagaya station in JMA who carefully observed SSR. The development of the manuscript benefited from discussions and comments by T. Oozeki. Reviews by two anonymous reviewers further strengthened the paper. Finally, the authors would like to thank two editors and three anonymous reviewers for their useful comments which have improved the manuscript.Edited by: Drs. Jan Kleissl and David Renne.

Edited by: Drs. Jan Kleissl and David Renne.

References

Badescu, V., 2014. Modeling solar radiation at the earth surface. In: Badescu, V. (Ed.). Springer.

Clark, A.J., Gallus Jr., W.A., Chen, T.-C., 2008. Contributions of mixed physics versus perturbed initial/lateral boundary conditions to ensemble-based precipitation forecast skill. Mon. Weather Rev. 136, 2140–2156. http://dx.doi.org/10.1175/2007MWR2029.1.

Coimbra, C., Kleissl, J., Marquez, R., 2013. Overview of solar forecasting methods and a metric for accuracy evaluation. In: Kleissl, J. (Ed.), Solar Resource Assessment and Forecasting. Elsevier, Waltham, Massachusetts, pp. 171–194.

Cornaro, C., Pierro, M., Bucci, F., 2015. Master optimization process based on neural networks ensemble for 24-h solar irradiance forecast. Sol. Energy 111, 297–312.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bildlot, J., Bormaan, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolado, C., Thépaut, J.-N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597. http://dx.doi.org/10.1002/qj.828.

Diagne, M., Mathieu, D., Lauret, P., Boland, J., Schmutz, N., 2013. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. Renew. Sustain. Energy Rev. 27, 65–76. http://dx.doi.org/10.1016/j.rser.2013.06.042.

Fonseca Jr., J.G.S., Oozeki, T., Takashima, T., Koshimizu, G., Uchida, Y., Ogimoto, L., 2012. Use of support vector regression and numerically predicted cloudiness to forecast power output of a photovoltaic power plant in Kitakyushu, Japan. Prog. Photovolt. 20, 874–882.

Garcia-Moya, J.-A., Callado, A., Escriba, P., Santos, C., Santos-Munoz, D., Simarro, J., 2011. Predictability of short-range forecasting: a multimodel approach. Tellus A 63, 550–563. http://dx.doi.org/10.1111/j.1600-0870.2010.00506.x.

Greybush, S., Haupt, S.E., Young, G., 2008. The regime dependence of optimally weighted ensemble model consensus forecasts of surface temperature. Weather Forecast. 23, 1146–1161.

Grimit, E.P., Mass, C.F., 2007. Measuring the ensemble spread–error relationship with a probabilistic approach: stochastic ensemble results. Mon. Weather Rev. 135, 203–221. http://dx.doi.org/10.1175/MWR3262.1.

Gueymard, C.A., 2012. Clear-sky irradiance predictions for solar resource mapping and large-scale applications: improved validation methodology and detailed performance analysis of 18 broadband radiative models. Sol. Energy 86, 2145–2169. http://dx.doi.org/10.1016/j.solener.2011.11.011.

Iizumi, T., Uno, F., Nishimori, M., 2012. Climate downscaling as a source of uncertainty in projecting local climate change impacts. J. Meteorol. Soc. Jpn. 90B, 83–90.

Jonson, C., Swinbank, R., 2009. Medium-range multimodel ensemble combination and calibration. Q. J. R. Meteorol. Soc. 135, 777–794. http://dx.doi.org/10.1002/qj.383.

Kanamitsu, M., Tada, K., Kudo, T., Sato, N., Isa, S., 1983. Description of the JMA operational spectral model. J. Meteorol. Soc. Jpn. 61, 812–828.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., 2015. The JRA-55 reanalysis: general specifications and basic characteristics. J Meteorol. Soc. Jpn. 93, 5–48.

Liu, Y., Shimada, S., Yoshida, J., Kobayashi, T., Miwa, Y., Furuta, K., 2016. Ensemble forecasting of solar irradiance by applying a mesoscale meteorological model. Sol. Energy 136, 597–605. http://dx.doi.org/10.1016/j.solener.2016.07.043.

Lorenz, E., Kühnert, J., Heinemann, D., Nielsen, K.P., Remund, J., Müller, S.C., 2016. Comparison of global horizontal irradiance forecasts based on numerical weather prediction models with different spatio-temporal resolutions. Prog. Photovolt. 24, 1626–1640.

Lorenz, E., Scheidsteger, T., Hurka, J., Heinemann, D., Kurz, C., 2011. Regional PV power prediction for improved grid integration. Prog. Photovolt. 19, 757–771. http://dx.doi.org/10.1002/pip.1033.

Lorenz, E.N., 1963. Deterministic nonperiodic flow. J. Atmos. Sci. 20, 130–141.

Matsueda, M., Kyoda, M., Tanake, H.L., Tsuyuki, T., 2007. Daily forecast skill of multicenter grand ensemble. SOLA 3, 29–32. http://dx.doi.org/10.2151/sola.2007-008.

Matsueda, M., Nakazawa, T., 2015. Early warning products for severe weather events derived from operational medium-range ensemble forecasts. Meteorol. Appl. 22, 213–222. http://dx.doi.org/10.1002/met.1444.

Murphy, A.H., 1988. Skill scores based on the mean square error and their relationships to the correlation coefficient. Mon. Weather Rev. 116, 2417–2424.

Ohba, M., Kadokura, S., Nohara, D., 2016. Impacts of synoptic circulation patterns on wind power ramp events in East Japan. Renew. Energy 96, 591–602.

Ohtake, H., Fonseca Jr., J.G.S., Takashima, T., Oozeki, T., Shimose, K.-I., Yamada, Y., 2015. Regional and seasonal characteristics of global horizontal irradiance forecasts obtained from the Japan Meteorological Agency mesoscale model. Sol. Energy 116, 83–99. http://dx.doi.org/10.1016/j.solener.2015.03.020.

Ohtake, H., Shimose, K.-I., Fonseca Jr., J.G.S., Takashima, T., Oozeki, T., Yamada, Y., 2013. Accuracy of the solar irradiance forecasts of the Japan Meteorological Agency

mesoscale model for the Kanto region, Japan. Sol. Energy 98, 138–152. http://dx.doi.org/10.1016/j.solener.2012.10.007.

Ohtake, H., Takashima, T., Oozeki, T., Fonseca Jr., J.G.S., Yamada, Y., 2016. A case study of outlier event on solar irradiance forecasts from the two NWPs with different horizontal resolutions. Renew. Energy Environ. Sustain. 37, 1–4. http://dx.doi.org/10.1051/rees/2016049.

Palmer, T.N., 2002. The economic value of ensemble forecasts as a tool for risk assessment: from days to decades. Q. J. R. Meteorol. Soc. 128, 747–774.

Pelland, S., Galanis, G., Kallos, G., 2013. Solar and photovoltaic forecasting through post-processing of the Global Environmental Multiscale numerical weather prediction model. Prog. Photovolt. 21, 284–296.

Pierro, M., Bucci, F., Felice, M.D., Maggioni, E., Moser, D., Perotto, A., Spada, F., Cornaro, C., 2016. Multi-model ensemble for day ahead prediction of photovoltaic power generation. Sol. Energy 134, 132–146. http://dx.doi.org/10.1016/j.solener.2016.04.040.

Saito, K., Fujita, T., Yamada, Y., Ishida, J., Kumagai, Y., Aranami, K., Ohmori, S., Nagasawa, R., Kumagai, S., Muroi, C., Kato, T., Erro, H., 2006. The operational JMA nonhydrostatic mesoscale model. Mon. Weather Rev. 134, 1266–1298. http://dx.doi.org/10.1175/MWR3120.1.

Sperati, S., Alessandrini, S., Monache, L.D., 2016. An application of the ECMWF Ensemble Prediction System for short-term solar power forecasting. Sol. Energy 133, 437–450.

Stensrud, D.J., Bao, J.W., Warner, T.T., 2000. Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. Mon. Weather Rev. 128, 2077–2107.

Swinbank, R., Kyouda, M., Buchanan, P., Frouda, L., Hamill, T.M., Hewson, T.D., Keller, J.H., Matsueda, M., Methven, J., Pappenbarger, F., Scheuerer, M., Titley, H.A., Wilson, L., Yamaguchi, M., 2016. The TIGGE Project and its achievements. Bullet. Am. Meteorol. Soc. 97, 49–67. http://dx.doi.org/10.1175/BAMS-D-13-00191.1.

Thorey, J., Mallet, V., Chaussin, C., Descamps, L., Blanc, P., 2015. Ensemble forecast of solar radiation using TIGGE weather forecasts and HelioClim database. Sol. Energy 120, 232–243. http://dx.doi.org/10.1016/j.solener.2015.06.049.

Whitaker, J.S., Loughe, A.F., 1998. The relationship between ensemble spread and ensemble mean skill. Mon. Weather Rev. 126, 3292–3302.

Wilks, S.D., 2005. Statistical Methods in the Atmospheric Sciences, second ed. Academic press, Ny.

Zhang, J., Florita, A., Hodge, B.-M., Lu, S., Hamann, H.F., Banunarayanan, V., Brockway, A.M., 2015a. A suite of metrics for assessing the performance of solar power forecasting. Sol. Energy 111, 157–175.

Zhang, J., Hodge, B.-M., Lu, S., Hamann, H.F., Lehman, B., Simmons, J., Campos, E., Banunarayanan, V., Black, J., Tedesco, J., 2015b. Baseline and target values for regional and point PV power forecasts: toward improved solar forecasting. Sol. Energy 122, 804–819.