Integrated Regional Solar PV Forecasting by Hybrid Method: Case Study of an Indian State

Aravindakshan Ramanan
GIZ GmbH

Naveen Kumaar
NIWE

Chris Alice Abraham
NIWE

Arindam Roy
NIWE

Barun Kumar
NIWE

Elena Barykina
Overspeed GmbH

Hans Peter Waldl
Overspeed GmbH

Felix Dierich
Overspeed GmbH

Indradip Mitra
GIZ GmbH

Karthik Ramanathan
NIWE

Prasun Kumar Das
NIWE

Balaraman Kannan
NIWE

***Abstract*—In India, cumulative power forecasts for the State Load Dispatch Centers are mandated under regulatory framework and accordingly, eleven Renewable Energy Management Centers are being established to implement RE forecasting activities. For the system operator, cumulative regional forecast of renewable energy generation is very important for maintaining the stability of grid. It will help them in scheduling the conventional generators accordingly and ensuring effective evacuation of renewable energy.**

Till now in India, the approach practiced was to forecast each plant and add them up for estimation at regional level. The concept of upscaling using few selected reference plants and arriving at the regional level is new in India. This study is the first attempt, following international practices, to approach regional level forecasting using upscaling technique. The initial results are promising and it is expected that such approach may soon become mainstream for future forecasting activities in India.

In this work, 12 solar power plants across an Indian state taken as Reference Plants, ensuring geographical distribution of solar power plants across the state, for providing Day Ahead Forecast services. Based on these plant specific forecasts, a cumulative regional forecast model was developed. The forecasts were done based on combination of both physical and statistical methods, using multiple sources of Numerical Weather Prediction models in the forecasts. Numerical Weather Prediction models from various sources from the US, Europe and India were interpolated to 15 minutes from their original time resolution. Diffuse fraction, tilt conversion models, DC, AC conversion models were employed with associated post processing steps to arrive at the solar power forecasts. Measured radiation data from the network of Solar Radiation Resource Assessment (SRRA) stations in the state were utilized in the model development. SRRA data was also used for generating bias correction and combination coefficients. Various model optimizations were done in the model chain to improve the accuracy of forecasts. The forecast for the region was arrived based on adopting suitable upscaling technique for deriving the forecast for the entire state. For the regional forecasts, Normalized Root Mean Square Errors of regional power production have been improved significantly compared to the single farm values, as expected. Detailed Accuracy measures for the individual power plants compared with the entire region were evaluated. Outcomes on the appropriateness of selection of reference power plants for regional forecasting were documented.

Keywords- solar forecasting; agggregation; regional forecasting

I. INTRODUCTION

With the increase in penetration of renewable energy, forecasting has become extremely crucial and important for evacuation of renewable energy into the grid. Grid operators are more concerned with forecasting of renewable energy for the entire control area. In the initial step of upscaling, suitable reference farms are selected. Forecasting is performed for the chosen reference farms. Based on the above, aggregated forecasts are generated for the entire control area.

In India, forecasting for aggregate area is performed by forecasting the power production of every power plant and summing them up individually. There is not much available literature related to the forecasting procedure for the regional forecasting using selected reference power plants. Deriving the forecasts of all the individual power plants in order to derive the aggregate forecasts is computation and time intensive exercise.

This is the first kind of attempt where day ahead forecast of solar power production is performed for an entire Indian state in 15 minutes time resolution by selection of appropriate reference power plants. All the steps in the different models where developed indigenously with maximum extent of Indian data sources. Validation of results are done with measured aggregated data of the solar power production of the entire state. Forecasting of single reference plants is done as a preliminary step towards arriving at the forecast of the entire control area. The results of the forecasts of single plants are also evaluated. The accuracy of forecasts of single plants and regional forecasts are cross-compared. There are multiple sources of Numerical Weather Prediction Models (NWP) namely European Centre for Medium Range Weather Forecasting (ECMWF), National Centre for Medium Range Weather Forecasting (NCMRWF) and Global Forecasting System (GFS) which was used for generation of AC power production forecasts. Accuracy of the above models for generating forecasts are also evaluated.

II. LITERATURE REVIEW

In the initial step of upscaling, suitable reference farms are selected. Forecasting is performed for the chosen reference farms. After performing the forecasts for the reference farms, aggregated forecasts are generated for the entire control area. There are number of methods for arriving at the aggregate forecasts. The following sections details out the criteria to be followed for reference farm selection and the techniques available in the literature for aggregate forecasting.

A. Selection of reference farms

The selection of reference farms is very important step that determines the accuracy of the forecasting model chain. Number of criteria needs to be taken into account while choosing the reference power plants as detailed in the following [9]:

- Regional Spread of the power plants must be ensured.
  - Farms which are quite distant have lower correlation of measurements and forecasts. Lower correlation improves regional forecasts as the errors tend to even out.
  - Including several nearby plants with high correlation does not improve the accuracy significantly.
  - Regional selection still needs to be representative of installed capacities.
- Representation for core areas of installed capacity.
- Representation of technology/ size mix is very important. The different technologies of the modules like poly crystalline, mon crystalline, etc needs to represented well in the reference power plant selection.
- Representation of power plants in different climatic zones as per installation needs to be considered.
- The relevance of number of reference farms to be selected in relation to the accuracy is of lesser importance compared to the above mentioned criteria.

B. Methods for regional forecasting

There are different methods mentioned in the literature for arriving at the regional day-ahead forecasts from the individual power plant forecasts. Some of them are enumerated below:

- Proportional Sum Method

In this method, all the non-reference power plants are considered as on big power plant. The forecast for the aggregate area is developed according to the following formula:

$$Upscaled\ solar\ power\ forecast\ (MW) = \frac{\sum Solar\ \ power forecast\ of\ reference\ plants\ (MW)}{\sum Installed\ Solar\ capacity\ of\ reference\ plants\ (MW)} * \sum Installed\ Solar\ capacity\ of\ entire\ State\ (MW)$$
(1)

- Nearest Neighbor Method

In this method, the forecasts of all the non reference power plants are calculated in relation to the forecasts of the nearest reference farms. Therefore, all the power plants are mapped according to the most proximate neighboring reference farm. The formula used in the calculations is mentioned in the following:

$$Forecast\ of\ non-reference\ power\ plant\ (MW) = \frac{Solar\ power\ forecast\ of\ neighbouring\ farm\ (MW)}{Installed\ capacity\ of\ neigbouring\ reference\ plant\ (MW)} * Installed\ capacity\ of\ non\ reference\ \ plant (MW)$$

$$Upscaled\ solar\ power\ forecast\ (MW) = \sum Forecast\ of\ all\ non\ reference\ farms\ (MW) + \sum Forecast\ of\ all\ reference\ farms (MW)$$
(2)

- Inverse Distance Weighted Method

The first step in this method is to normalize the production of each of the reference power plants by its nominal capacity so that all yield values are comparable to each other irrespective of the sizes. For each of the non-reference power plants, individual weighting factors with respect to reference power plants are estimated. This is in relation to the distance between power plants. The more proximate reference power plants will have greater influence on the predictions[5].

The yield of the unknown PV plants is then estimated by spatially interpolating the yield of the reference plants to the location of the unknown plants. Interpolated yield values are then normalized by the nominal capacity of the unknown PV plants and finally aggregated over the considered region to assess the regional PV power generation. The formula mentioned below is utilized for this purpose.

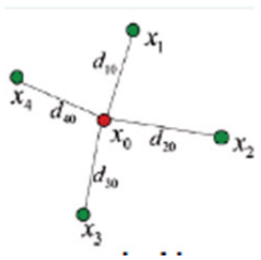

Figure 1. Representation of Inverse Distance Weighted Method

$$y(x_0) = \sum_{i=1}^{n} w(x_0, x_i) * y(x_i) \qquad (3)$$

$$w(x_0, x_i) = \frac{d(x_0,x_i)^{-p}}{\sum_{k=1}^{n} d(x_0,x_i)^{-p}} \qquad (4)$$

III. DATA SETS USED

A. NWP Data

Three Numerical Weather Prediction models provided by ECMWF, NCMRWF and GFS were used in the current work. NCMRWF was running the UK met office based model . The time and spatial resolutions of NWP models used in this exercise are given in Table 1.

TABLE I. DESCRIPTION OF SPATIAL AND TEMPORAL RESOLUTION OF NWP MODEL [1,2,3]

NWP Model Name	Spatial Resolution	Original Temporal resolution	Interpolated Temporal resolution (not given by the provider)
NCMRWF	0.25 degree x 0.25 degree (Approx. 25 km * 25 km)	1 hour	15 minutes
GFS	0.25 degree x 0.25 degree (Approx. 25 km * 25 km)	1 hour	15 minutes
ECMWF	0.25 degree x 0.25 degree (Approx. 25 km * 25 km)	3 hours	15 minutes

B. Solar Radiation Resource Assessment (SRRA) Data

SRRA network is a series of high quality, well maintained radiation measurement stations maintained all over India. Many meteorological parameters namely Global Horizontal Irradiance (GHI), Diffuse Horizontal Irradiance (DHI), Direct Normal Irradiance (DNI), etc are measured. State of the art, secondary standard pyranometers are installed for the measurements of the radiation. This network is owned and maintained by National Institute of Wind Energy (NIWE) [4]. All the parameters are run through state of the art quality control algorithms.

C. Power Plant Data

Static and dynamic data received from totally 12 reference power plants were utilized for the forecasting model development. The only dynamic data received from the reference solar power plants was AC power generation. Aggregate AC power measurement data of all the solar power plants in the state was also received. The AC power data was quality controlled (stuck value check, missing values, etc) and used in the model development.

IV. METHODOLOGY

Forecasting model is developed and operational system is implemented for Day ahead forecasting of solar power generation for entire state. The flowchart of the methodology followed for the developed day ahead system is shown in Figure 2. The original NWP's are interpolated to 15 minutes time resolution which is the scheduling resolution. The interpolation is performed by clear sky method.

Bias correction coefficients are calculated based on historical measured ground measured data. Fourth order polynomial as function of cosine of zenith angle and clear-sky index is calculated. Ineichen clear sky model was employed in this exercise. The coefficients in the equation 5 are calculated by fitting with ground measured data. The radiation measurements from one of the SRRA stations is used as ground measurements for calculation of bias correction coefficients. The calculated bias corrected coefficients were used in the forecasting model chain for all the reference power plants in the state.

The combination coefficients were obtained by fitting the coefficients according to linear regression for the data according to equation 5.

$$Combi_{NWP} = a1 * NWP1 + a2 * NWP2 + a3 * NWP3 + a4 \qquad (5)$$

The ground measured SRRA data of one station was taken as reference for calculation of combination coefficients. The best performing individual NWP will have the highest weightage in the step of combination.

Two empirical models: Diffuse fraction model and tilt conversion model are used for the conversion of irradiance from horizontal to tilt plane. The diffuse fraction model used in this work was Chandrasekar and Kumar which is mentioned in the following set of equations [6].

$$k_d = 1.0086 - 0.178k \; for \; k \leq 0.24 \quad (6)$$
$$k_d = 0.9686 + 0.1325k + 1.4183k^2 - 10.1860k^3 + 8.3733k^4 \; for \; 0.24 < k \leq 0.8 \quad (7)$$
$$k_d = 0.197 \; for \; k > 0.8 \quad (8)$$

- Where k_d= diffuse fraction, k= clearness index

Klucher model was used for the tilt conversion as mentioned in the equations 9 and 10 [7].

$$I_T = \frac{(I_H - I_D) * \cos(\psi)}{\sin(\alpha)} + I_D\left(\frac{1+\cos\varepsilon}{2}\right)\left(1 + Fsin^3\left(\frac{\varepsilon}{2}\right)\right)\left(1 + Fcos^2(\psi)sin^3(90-\alpha)\right) \quad (9)$$
$$F = 1 - \left(\frac{I_D}{I_H}\right)^2 \quad (10)$$

- I_T=insolation on surface tilted toward the equator at angle ε
- I_H=total insolation received on horizontal surface
- I_D=diffuse insolation received on horizontal surface
- (α) = solar elevation angle
- (ψ)= angle between sun direction and normal direction of tilted surface

Forecasted DC power is calculated as a function of estimated tilted irradiance and Module temperature as per equation 11. The parameters in the GTI to DC conversion model are chosen from the static data received from the reference power plants.

$$DCPower = GTI * Number\ of\ Panels * Panel\ Area * Efficiency * (1 + \text{alpha} * (\text{temp\_mod} - 25)) \quad (11)$$

- alpha= temperature coefficient of PV module.

Schmidt and Sauer model was used in the inverter modelling as per equations 12, 13 and 14. The parameters of the Schmidt and Sauer model (Pself, Vloss, rloss) was fitted from the power plant where the DC and AC power measurements were available. The same set of coefficients are applied to all the reference power plants which are to be forecasted.

$$Ploss = Pself + Vloss.Pout + rloss.P^2out \quad (12)$$

- Pself, Vloss, rloss are inherent, magnetic, resistive loss coefficients respectively for inverter model
- Equation (13) is valid when rloss is positive. Equation 14 is applicable when rloss is negative.

$$\eta * = -\frac{1+v*_{loss}}{2.r_{loss}.p_{in}} + \sqrt{\frac{(1+v*_{loss})^2}{(2.r*_{loss}.p_{in})^2} + \left(\frac{p_{in}-p*_{self}}{r*_{loss}.p_{in}^2}\right)} \quad (13)$$

$$\eta * = -\frac{1+v*_{loss}}{2.r_{loss}.p_{in}} - \sqrt{\frac{(1+v*_{loss})^2}{(2.r*_{loss}.p_{in})^2} + \left(\frac{p_{in}-p*_{self}}{r*_{loss}.p_{in}^2}\right)} \quad (14)$$

Twelve reference plants where chosen as basis for performing the forecast for the control area. The reference power plants were selected to represent the proportion, technology, etc of all the installations in the state. All the parameters detailed in the previous section was considered in their selection.

Based on the forecasts of the individual reference plants, forecast is derived for the entire control area by the proportional sum and nearest neighbor methods (formula explained in previous section).

The aggregated forecasts are modified according to the bias correction factor. Bias correction factor is derived based on linear fitting technique between historical aggregated forecasts and measurements.

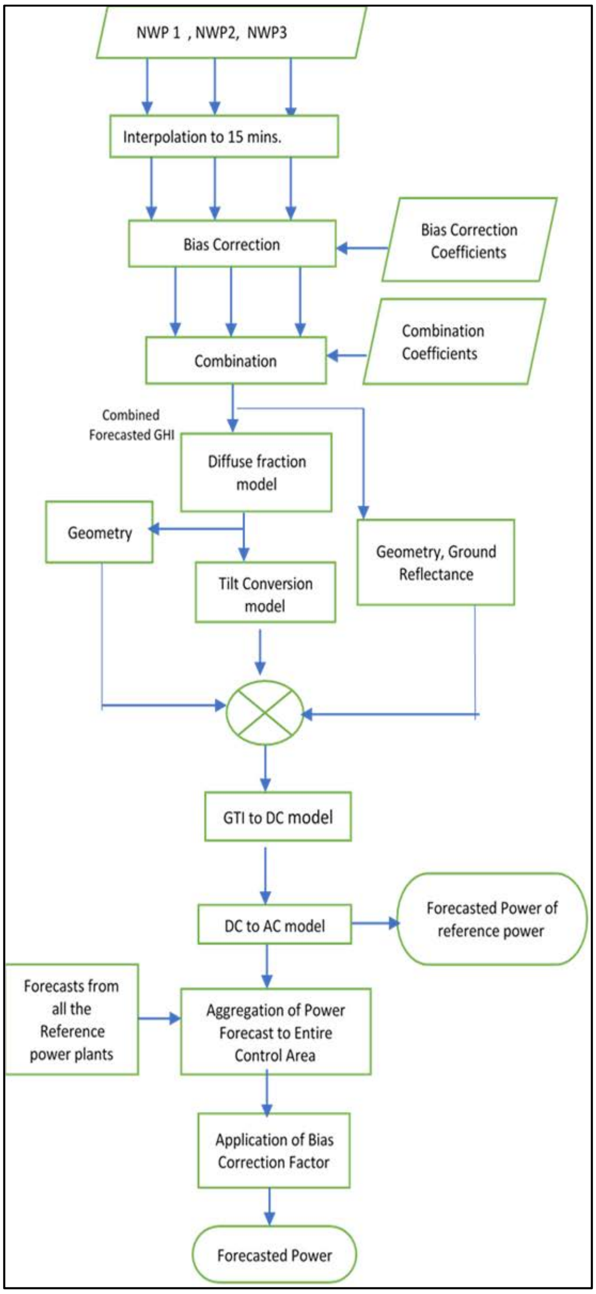

Figure 2. Flow chart depicting methodology followed for the exercise of forecasting

V. RESULTS

A. Bias Correction and Combination of Irradiance

Bias correction coefficients were calculated by fitting the deviation between forecasted and measured GHI for historical dataset (1st May 2017 - 31st January 2018), as a fourth order polynomial function of clear sky index and the cosine of apparent zenith angle.

The combination coefficients were obtained by fitting the coefficients according to linear regression.

While estimation of Key Performance Indicator (KPI's) , the night time values were excluded from the analysis. Normalizing factor applied for KPI calculation is 1000 W/m2.

The comparison of KPI values is shown in Table 2.

TABLE II. REPRESENTATION OF KPI'S OF NWP (RAW, BIAS CORRECTED) AND COMBINATION PROCESS FOR IRRADIANCE FORECASTS IN RELATIVE TERMS.

NWP name	Raw/ Bias Corrected	NRMSE(%)	NBias (%)
NWP1	Raw	X1	Y1
	Bias Corrected	0.93X1	0.133Y1
NWP2	Raw	1.37X1	2.238Y1
	Bias Corrected	1.11X1	0.3813Y1
NWP3	Raw	1.504X1	2.094Y1
	Bias Corrected	1.236X1	0.287Y1
Combination	Bias Corrected (ECMWF, NCMRWF and GFS)	0.888X1	-0.0056Y1

B. DC to AC Model

Schmidt and Sauer model as explained in the previous section was used in this study [8]. The inverter coefficients were derived from 10 MW power plant were DC power and AC power data were available. The value of coefficients derived based on fitting of equation are mentioned in Table 4. The Schmidt Sauer model was able to model the performance of the inverter to varying irradiation profiles quite accurately. The modelled and measured efficiency plots of validation dataset is shown in Figure 3. The same set of coefficients were used in the forecasting model chain of all the reference power plants.

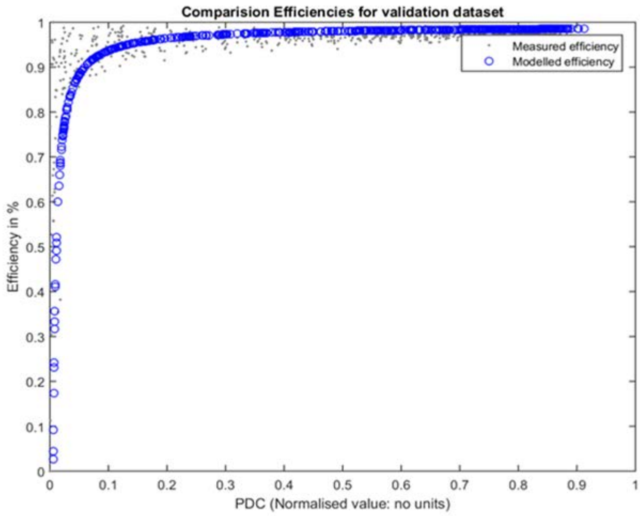

Figure 3. Modelled and measured efficiency (Validation data set) for 10 MW power plant

C. Performance of forecast for Reference Power Plants

In this section, analysis of **AC power forecasts** for the reference power plants for different NWP's and combination is presented. The combination of NWP's is done for the irradiance prediction in the forecasting chain. The time period of validation of forecasts is **8th December 2018 to 26th February 2019**. The normalization factor for KPI are based on the installed capacity of the solar power plant. The results of the forecasts of reference power plants is indicated in Figure 4 and Figure 5. The KPI's are represented in relative terms for all the power plants.

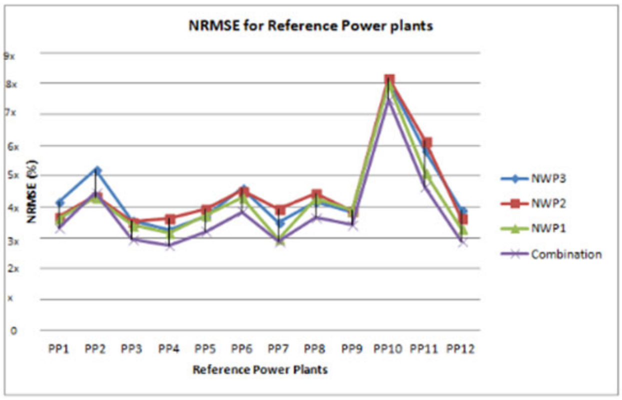

Figure 4. Representation of NRMSE for 12 reference plants in relative terms.

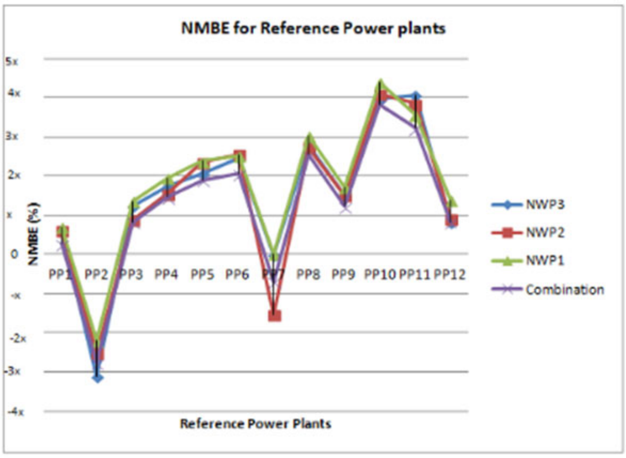

Figure 5. Representation of NMBE for 12 reference plants in relative terms.

As it can be noticed from Figure 4 and Figure 5, NWP 1 is performing relatively the best compared to all the NWP's. Combination of forecasts is showing best results compared to all the individual NWP's. Power plant 10 is showing higher errors due to the curtailment of generation in certain days. This curtailment information is not captured by the forecasting system.

D. Performance of Aggregate forecast based on historical data sets

Since NWP1 forecasts are the best performing compared to other NWP's, the upscaling of forecasts for the entire state is attempted with NWP1 model chain. Two types of upscaling techniques (proportional sum and nearest neighbor) are attempted.

Forecast accuracy evaluation at the AC Power end is performed for a period from ***8th December 2018 to 26th February 2019***, which is the same period considered for individual power plants. The methodology as prescribed in previous section is followed. Bias correction factor is not applied at AC power end. The KPI's for the process is shown in Table 3 (Shown in relative terms). The KPI's are calculated for distinct set of dates because, installed capacity was varying for the state in the validation time period. The normalization factor for KPI's are based on the installed solar power plant capacity of the entire state.

As it can be observed, proportional sum is performing much more accurately compared to nearest neighbor method.

TABLE III. REPRESENTATION OF KPI'S FOR THE AGGREGATE FORECASTS IN RELATIVE BASIS

DATE RANGE	Proportional Sum		Nearest Neighbor	
	NRMSE (%)	**NMBE (%)**	**NRMSE (%)**	**NMBE (%)**
08-12-2018 to 10-12-2018	X1	Y1	1.92X1	2,358Y1
11-12-2018 to 04-01-2019	0.989X1	0.989Y1	1.89X1	2.33Y1
05-01-2019 to 21-01-2019	0.966X1	0.966Y1	1.85X1	2.28Y1
22-01-2019 to 21-02-2019	0.944X1	0.944Y1	1.81X1	2.23Y1
22-02-2019 to 26-02-2019	0.912X1	0.912Y1	1.75X1	2.15Y1

E. Performance of operational forecasting system for aggregate forecasts

In this section, the results for operational forecasting system is represented. Operational forecasting system for solar power generation was setup based on NWP1 forecasts. In the previous section, two upscaling methods were tested and "Proportional Sum" method was giving better results. Therefore, operational forecasting system is implemented with "Proportional Sum" Method.

Forecast accuracy evaluation at the aggregate AC Power is for the period between 7th June 2019 to 7th July 2019 for the operational forecasting system. The normalization factor for KPI's and plots are based on the installed solar power plant capacity of the entire state.

The sample time series of forecasts and measurement is represented in Figure 6. The aggregate forecasts with and without the bias correction factor are represented.

DATE RANGE	Bias Correction Factor at AC power	NRMSE (%)	NMBE (%)	NMAE (%)
7th June till 7th July, 2019	No	X1	Y1	Z1
7th June till 7th July, 2019	Yes	0.922X1	-0.786Y1	0.847Z1

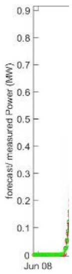

Figure 6. Representation of sample time series of aggregate forecasts and measurement.

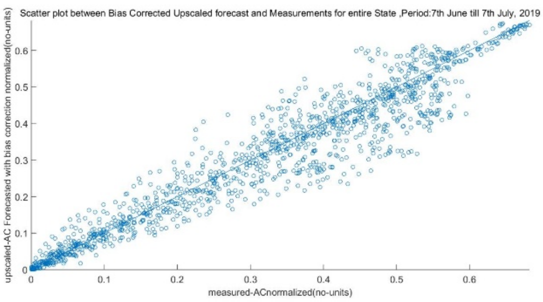

Figure 7. Scatter Plot between forecasted power with bias correction and measured AC Power

The KPI's calculated for operational forecasting system are mentioned in the Table 4. It could be seen that NRMSE is slightly improving after the bias correction factor is applied. Normalizing factor for the error metrics is the total installed capacity of solar power plants in the state. Bias is getting shifted from positive to negative side after application of correction factor.

TABLE IV. KPI'S FOR THE OPERATIONAL FORECAST SYSTEM IN RELATIVE TERMS

VI. DISCUSSION AND CONCLUSION

Grid operators primarily require the forecast for the entire Control area. As this will help them in scheduling and grid management. In this work, development of indigenous day ahead forecasting for an Indian state was attempted. Combination of physical and statistical methodologies was adopted. REMC's being established in India will support the state SLDC's in this very crucial functionality of aggregate forecasting [10]. The general practice in India is to perform the forecast of all individual plants and arriving at the forecast of the control area. This study is the first attempt where the forecasts are generated for the control area, by choosing certain selected set of power plants.

Considering all the important criteria, twelve reference plants was chosen as basis for performing the forecast for the control area. The forecasting system developed is showing reasonably good accuracy. It was seen that combination of NWP models is improving the forecasting accuracy of individual plants. It is also expected, implementation of combination process will result in accuracy enhancement of the aggregate forecasting system.

In order to fine tune the forecasting system, combination methods (linear and non-linear), advanced diffuse models, etc are in the process of being implemented. Algorithms for including uncertainty information in the forecasts are also being implemented which will be very important for the grid operators to manage and plan for reserves.

Static and dynamic data of the power plants are crucial information for the development of solar power forecasting system. The measured data of solar power generation (of reference plants and entire state solar generation) was only received for the modelling exercise. The accuracy of the model could be significantly improved if other dynamic parameters like GHI, Global Tilted Irradiance, Module Temperature, etc are also obtained. The radiation data from SRRA station in one of the locations in the state was used for generating bias correction and combination coefficients for all the 12 reference solar power plants. Access to power plant specific radiation measurements will help to calculate distinct set of coefficient for each of the reference locations.

AC power measurement data was received only from 12 reference plants. Therefore, this forecasting methodology assumed rest of the power plants in the state are in good working condition. Access to the measurement data from all the solar power plants can help to finetune the final forecast generated by analysis of historic data.

It is important to receive the data in the least possible time resolution from the provider. In the analysis of historic aggregate forecast and measurements, there were time-shift between the two signals. This could be because, the measurement data was received only in 15 minutes time resolution. However, the time shift was eliminated in operational forecasting system when the data was started to receive in one minute time resolution.

This work is expected to be further researched with further advanced techniques as mentioned. Efforts are underway to obtain further possible datasets that will assist in improvement of results. It is expected that promising indigenous forecasting system that has been developed will continue to evolve into a mature and robust system.

ONGOING AND FURTHER WORK

Satellite-based forecasting for intra-day time horizon.
Improvement modelling the influence of aerosols.
Improvement of soiling modelling [11].

REFERENCES

[1] https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs

[2] https://www.ncmrwf.gov.in/

[3] https://www.ecmwf.int/

[4] https://niwe.res.in/department_srra.php

[5] Yves-Marie Saint-Drenan. (2015), *A Probabilistic Approach to the Estimation of Regional Photovoltaic Power Generation using Meteorological Data*. PhD Thesis. Faculty of Electrical Engineering and Computer Science Of the University of Kassel.

[6] J. Chandrasekaran and S. Kumar, “Hourly diffuse fraction correlation at a tropical location”, Solar Energy, vol. 53, no. 6, pp. 505-510, 1994.

[7] T. Klucher, “Evaluation of models to predict insolation on tilted surfaces”, Solar Energy, vol. 23, no. 2, pp. 111-114, 1979.

[8] H.Schmidt, D.U.Sauer, (1996). Wechselrichter-Wirkungsgrade. Sonnenenergie 4, 43-47, 1996

[9] Felix Dierich , Seminar on Up scaling procedure/ Aggregation by Overspeed, Germany at NIWE, Chennai, India, Feb 2-4, 2019

[10] Tender for Renewable Energy Management Centre , Power Grid corporation of India Limited.

[11] Sahana L, N. Kumar GM, H. P. Waldl, P. K. Das, K. Ramanathan, B. Kannan, I. Mitra, “Impact of Soiling on Energy Yield of Solar PV Power Plant and Developing Soiling Correction Factor for Solar PV Power Forecasting”, 2nd Large Scale Grid Integration Conference, September 2019.