# Comparison and Sensitivity Analysis of Methods for Solar PV Power Prediction

Mashud Rana[1(✉)], Ashfaqur Rahman[2], Liwan Liyanage[3], and Mohammed Nazim Uddin[4]

[1] Data61, CSIRO, Sydney, NSW, Australia
`mdmashud.rana@data61.csiro.au`
[2] Data61, CSIRO, Sandy Bay, TAS, Australia
`ashfaqur.rahman@data61.csiro.au`
[3] School of Computing, Engineering and Mathematics,
Western Sydney University, Sydney, Australia
`l.liyanage@westernsydney.edu.au`
[4] Department of Computer Science and Engineering, East Delta University,
Chittagong, Bangladesh
`nazim@eastdelta.edu.bd`

**Abstract.** The variable nature of solar power output from PhotoVoltaic (PV) systems is the main obstacle for penetration of such power into the electricity grid. Thus, numerous methods have been proposed in the literature to construct forecasting models. In this paper, we present a comprehensive comparison of a set of prominent methods that utilize weather prediction for future. Firstly, we evaluate the prediction accuracy of widely used Neural Network (NN), Support Vector Regression (SVR), k-Nearest Neighbours (kNN), Multiple Linear Regression (MLR), and two persistent methods using four data sets for 2 years. We then analyze the sensitivities of their prediction accuracy to 10–25% possible error in the future weather prediction obtained from the Bureau of Meteorology (BoM). Results demonstrate that ensemble of NNs is the most promising method and achieves substantial improvement in accuracy over other prediction methods.

**Keywords:** Solar power prediction · Sensitivity analysis · Neural networks · Support Vector Regression · Nearest neighbours · Regression

## 1 Introduction

Solar power is one of the prominent sources of renewable energy and expected to contribute to a major share of electricity generation in future. Many countries have been installing large-scale solar PV plants and connecting them to grid to meet their electricity demand. For example, in 2015–2016 solar power generation in Australia increased by 24% and accounted for 3% of its total energy generation [1]. By the year 2050, Australia also aims to generate 29% of the electricity from solar PV systems [2].

However, unlike the traditional sources of energy, power output from solar PV plants fluctuates because of its dependency on meteorological conditions. This fluctuation imposes substantial challenges on achieving high level penetration of solar

power into the electricity grid [3]. A rapid unexpected change in the PV output power creates grid operational issues and negatively affects the security of supply. Reliable prediction of solar PV power output at different horizons is therefore critical to compensate the negative consequences related to variability in generation.

The prominent methods for solar power prediction are based on machine learning techniques such as NN (e.g. [4, 5]), SVR (e.g. [6]), and kNN (e.g. [7]); and statistical methods such as MLR and Autoregressive methods [8, 9]. Pedro and Coimbra [10] studied the performance of several methods for 1 and 2 h ahead prediction using a data set consisting of only previous solar power data. Although various weather information (such as solar exposure, temperature, and rainfall) have significant influence on the power output, they excluded the weather information and used only previous power data as inputs as their primary goal was to develop the baselines for further evaluation.

Long et al. [11] evaluated the performance of four methods using a data set that included weather parameters in addition to previous power as the inputs. However, they considered forecasting the cumulative power output for a day. Predicting daily total power output has very limited use in practical applications for real-time grid operation as the power output largely varies at different times of a day (see Fig. 1 showing the variability of solar power outputs at different times during the days) depending on the weather condition.
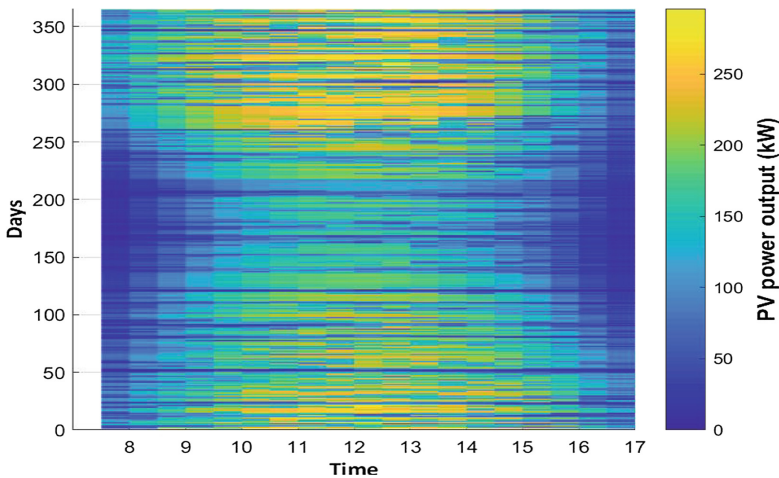


**Fig. 1.** Variability of solar PV power output at different times from 7:00 am to 5:00 pm for all days in 2015 (sample data set name UQC)

A review of the relevant literature suggests the predictive performance of the existing methods were evaluated on different data sets and for different forecasting scenarios (such as different prediction horizons). Hence, their computational accuracy reported in the literature is not readily comparable and does not convince to demonstrate the superiority of any single method over others. Therefore, it can be concluded that despite numerous approaches proposed and many notable achievements cited

systematic comparison of different methods utilizing weather data accompanied by analysis of their sensitivities to error in future weather prediction is very limited. In this paper, we address this deficit in the literature and provide a comprehensive evaluation of a group of solar power forecasting methods that utilize weather prediction for future. In contrast to the previous work our contributions can be summarized as follows:

We compare a set of prominent methods for the task of predicting the output power profile for a day - i.e. predicting all the power outputs at half-hourly intervals for a day. Forecasting daily power profile is crucial for real time unit dispatching, gird reliability, and supporting energy trading in the deregulated markets. Firstly, we evaluate the 4 state-of-the-art and 2 baseline methods using 4 different data sets for two years 2015–2016 collected from the largest flat-panel grid connected PV plant in Australian. We then analyze the sensitivities of those methods to error in weather prediction for future by considering 10–25% noise in weather data obtained from BoM. This has been done to evaluate the robustness of the prediction methods in dealing the uncertainty associated with weather prediction. To the best of our knowledge this has not been investigated before in the literature.

## 2 Data Sets and Problem Statement

### 2.1 Solar Power Data

We use the data from the largest flat-panel PV system in Australia which has been located at the St Lucia campus of University of Queensland, Brisbane. It has a capacity of 1.22 MW and consists of about 5000 solar panels installed across four different sites: University of Queensland Centre (UQC), Sir Llew Edwards Building (LEB), Car Park 1 (CP1) and Car Park 2 (CP2). We use data from all four sites and consider the data from each site as a separate case study.

For each site, we collect the solar power output data in 1 min interval for 2 years – 2015 and 2016. For each day we use the data from 7:00 am to 5:00 pm since the solar power outputs have been recorded as either zero or not available outside this 10-h window, due to the absence (or very small amount) of solar irradiation. This leads to $2 \times 365 \times 600 = 438,000$ 1-min measurements for each site. The 1-min data is publicly available at [12]. The 1-min data was then aggregated into 30-min intervals by averaging every 30 consecutive measurements, resulting in 20 measurements per day and $2 \times 365 \times 20 = 14,600$ measurements in total for each site.

### 2.2 Weather Data

We also collect the meteorological data for the four variables – solar exposure, rainfall, sunshine hours, temperature and wind speed, for the same time period and from the nearest weather station of the PV sites, from BoM Australia [13]. These are among the widely cited weather variables affecting the solar power output from PV systems.

All data (PV power, solar exposure, rainfall, sunshine hours, temperature, and wind speed) has been normalized to the range of [0, 1].

### 2.3 Problem Statement

Given

1. a time series of previous solar power outputs up to the day $d$: $P = [P_1, P_2, P_3, \ldots, P_d]$, where $P_i = [p_i^1, p_i^2, p_i^3, \ldots, p_i^n]$ represents the power profile for day $i$, i.e. $n$ observations of the power outputs measured at half-hourly intervals ($n = 20$ for our case);
2. a time series of previous weather data from BoM up to day $d$: $W = [W_1, W_2, W_3, \ldots, W_d, ]$, where $W_i$ is the weather data for day $i$. $W_i$ is a 6-dimensional vector of the daily global solar exposure (SE), rainfall(R), sunshine hours (SH), maximum wind speed (WS), and maximum and minimum temperature (T), $W_i = [SE^i, R^i, SH^i, WS_{max}^i, T_{max}^i, T_{min}^i]$.
3. predicted weather data $W_{d+1}$ for day $d + 1$.

Goal: forecast $P_{d+1}$ - half-hourly solar PV power profile for the next day $d + 1$ using predicted weather data $W_{d+1}$ for day $d + 1$ as input. It is important to note that unlike other methods, only kNN requires: (1) previous weather data up to day $d$ to select the $k$ neighbours (days) nearest to predicted weather data $W_{d+1}$ for target day $d + 1$, and (2) previous power outputs of the $k$ selected neighbours (days) to compute the prediction for $P_{d+1}$.

## 3 Methods

For comparison, we consider four most prominent methods in the literature: NN, SVR, kNN, and MLR, and two persistent methods as baselines. All the methods use forecasted weather profile as input to predict the daily solar power curve for next day.

### 3.1 NN

To build NN based prediction model, we use multilayer perceptron NN. Approaches based on such NN are the most popular for solar power prediction. NN can learn and estimate complex functional relationships between the input and output variables from examples. However, the performance of NN depends on the network architecture and the random initialization of weights. To reduce this sensitivity, we apply an ensemble of NNs. Ensembles are shown to be more accurate than a single NN in previous works (e.g. [14]) using the same data sets.

To develop ensemble of NNs we follow [14, 15]. We first construct $V$ different NN structures by varying the number of neurons in hidden layer from *1 to V*. For each structure $S_{v \varepsilon V}$, we then build an ensemble $E_v$ that consists of $n$ NNs (we used $n = 10$). Each member of the ensemble $E_v$ has the same structure $S_v$, i.e. the same number of hidden neurons, but is initialized to different random weights. Each of the $n$ members of ensemble $E_v$ has been trained separately on the training data using the Levenberg Marquardt (LM) algorithm. The NN training has been stopped when there is no improvement in the error for 20 consecutive epochs or a maximum number of 1000 epochs is reached.

To predict the half-hourly power outputs for a given day, the individual predictions of the ensemble members are combined by taking their median value, i.e. the prediction for time (half an hour) $h$ for day $d + 1$ is: $\hat{P}_h^{d+1} = median\left(\hat{P}_{h,NN_1}^{d+1}, \ldots, \hat{P}_{h,NN_n}^{d+1}\right)$, where $\hat{P}_{h,NN_j}^{d+1}$ is the prediction for $h$ generated by an ensemble member $NN_j$, $h = 1,\ldots, 20$ and $j = 1, \ldots, n$.

The performance of each ensemble $E_v$ was evaluated on the validation set. The best performing ensemble, i.e. the one with the lowest prediction error, is then selected and used to predict testing data.

## 3.2   SVR

SVR is an advanced machine algorithm that has shown excellent performance for solar power forecasting [6, 16]. The key idea of SVR is to map the input data into a higher dimensional feature space using a non-linear transformation and then apply linear regression in the new space. The task is formulated as an optimization problem. The main goal is to minimize the error on the training data, but the flatness of the line and the trade-off between training error and model complexity are also considered to prevent overfitting.

Solving the optimization problem requires computing dot products of input vector in the new space which is computationally expensive in high dimensional spaces. To help with this, kernel functions satisfying the Mercer's theorem are used - they allow the dot products to be computed in the original lower dimensional space and then mapped to the new space.

Since SVR can have only one output, we divide the daily load curve prediction task into 20 subtasks (i.e. predicting power output for each half-hourly time separately) and build a separate SVR prediction model for each subtask.

The selection of kernel function is important for SVR, and is done by empirical evaluation. After experimenting with different kernel functions, we selected the Radial Basis Function (RBF) kernel as it achieved the best performance on the validation data.

## 3.3   kNN

kNN is an instance based method for forecasting. The main concept of kNN is selecting a subset of training examples whose inputs are similar to the inputs of test example, and use the outputs of that training subset to predict the outputs for the test example.

To forecast the power outputs for the next day $d + 1$, kNN firstly obtains the weather profile $W_{d+1}$ for $d + 1$ from the weather forecast report. It then finds the $k$ nearest neighbors of $d + 1$. This is done by matching the weather profile of all previous days ending with $d$ and finding the $k$ most similar days. This leads to a neighbour set $NS = \{q_1, \ldots, q_k,\}$ where $q_1, \ldots, q_k$ are the $k$ closest days to day $d + 1$, in order of closeness computed using a distance measure between weather profiles of the neighbors to that for $d + 1$. The prediction for the new day is the weighted linear combination   of   the   power   outputs   for   the   $k$   nearest   neighbors: $\hat{P}_{d+1} = \frac{1}{\sum_{s \in NS} \alpha_s} \cdot \sum_{s \in NS} \alpha_s \cdot P_s^h.$

The weights $\alpha_s$ are computed by following [17]: $\alpha_s = \frac{dist\left(W_{q_k}, W_{d+1}\right) - dist(W_s, W_{d+1})}{dist\left(W_{q_k}, W_{d+1}\right) - dist\left(W_{q_1}, W_{d+1}\right)}$, where $dist$ is the Euclidian distance measure, and $k$ is the number of neighbors. The optimal value for $k$ can be set by applying kNN method on the training data set, i.e. it is the one that minimizes the prediction error for training data set.

### 3.4 MLR

MLR is a classical statistical method for forecasting. It assumes linear relationship between the predictor variables and the variable that is predicted, and uses the least square method to find the regression coefficients. In this work we apply Weka's implementation of linear regression. It has an inbuilt mechanism for input variable selection based on the M5 method. This method firstly builds a regression model using all inputs and then removes the variables, one by one, in decreasing order of their standardized coefficient until no improvement is observed in the prediction error given by the Akaike Information Criterion (AIC). Similar to SVR, for MLR we train one model for each half-hourly time since MLR can have only one output.

### 3.5 Persistent Methods

We also implement two persistent methods as baselines for comparison.

The first persistent method ($B_{msday}$) firstly selects the most similar previous day ($s$) in the historical data based on the weather profile where the similarity is measured by Euclidean distance between the weather profiles of $d + 1$ and $s$. It then uses the power output for the day $s$ as the predictions for the day $d + 1$. This means the prediction for $\hat{P}_{d+1} = \left(\hat{P}_{d+1}^1, \hat{P}_{d+1}^2, \ldots, \hat{P}_{d+1}^{20}\right)$ is given by $P_s = \left(p_s^1, p_s^2, \ldots, p_s^{20}\right)$. Obviously $B_{msday}$ is same as kNN if only a single neighbour is considered in kNN.

The second persistent method ($B_{pday}$) considers the power outputs from the previous day $d$ as the predictions for the next day $d + 1$, i.e. the predictions for $\hat{P}_{d+1} = \left(\hat{P}_{d+1}^1, \hat{P}_{d+1}^2, \ldots, \hat{P}_{d+1}^{20}\right)$ are given by $P_d = \left(p_d^1, p_d^2, \ldots, p_d^{20}\right)$.

## 4   Simulation Settings

### 4.1 Training and Testing Data

We divide the data for each case study into two non-overlapping subsets – *training* and *testing*. The training set consists of all the samples from the year 2015 and has been used to build the prediction models. This applies to all the models expect NN and SVR. For NN and SVR 90% of the samples from the training set has been used to train the models and remaining 10% (validation set) has been used for parameter selection (such as selecting the number of hidden neurons for NN and kernel for SVR). During the training phase we used actual weather data $W_{d+1}$ for target day $d + 1$ as input since training was performed offline and we do not have access to the historical weather prediction.

On the other hand, testing set consists of all the samples from the year 2016 and has been used to evaluate the accuracy of the prediction models.

### 4.2  Evaluation Metrics

To evaluate the accuracy of forecasting models, we use Mean Relative Error (MRE). MRE is one of the widely cited measure for the accuracy of solar power prediction and defined as: $\text{MRE} = \frac{1}{D}\frac{1}{H}\sum_{d=1}^{D}\sum_{h=1}^{H}\left|\frac{p_d^h - \hat{P}_d^h}{R}\right| \times 100\%$.

where $p_d^h$ and $\hat{P}_d^h$ are the actual and predicted power outputs for day $d$ at time $h$, respectively; $D$ is the number of instances (days) in the testing data; $H$ is the total number of predicted power outputs for a day ($H = 20$ for our task), and $R$ is the range of the power output.

For comparison of prediction models, we also compute the improvement in accuracy between two prediction models $A$ and $B$ as: $\text{improvement}(A, B) = \frac{|MRE_A - MRE_B|}{MRE_B} \times 100\%$.

## 5  Results and Discussion

Table 1 presents the accuracy (MRE) of the prediction models using actual weather data for target day as input as we do not have access to the historical weather prediction. The statistical significance for differences in accuracy for each pair of prediction models are also shown Table 2. Results show that ensemble of NNs is the most accurate model and outperforms all other prediction models including two baselines in all data sets, except SVR for EBD data set. The overall improvements of accuracy for NNs ensemble are 0.18–2.94% over SVR, 23.77–56.06% over kNN, 1.04–3.14% over MLR, 5.34–30.35% over $B_{msday}$, and 14.82–39.25% over $B_{pday}$. All the improvements of NNs ensemble are also statistically significant at p $\leq$ 0.05 except the difference between NNs ensemble and SVR for CP2 data (see Table 2).

**Table 1.** Accuracy results (MRE) of all prediction models evaluated on four data sets.

|      | NN    | SVR   | kNN   | MLR   | $B_{msday}$ | $B_{pday}$ |
|------|-------|-------|-------|-------|-------------|------------|
| CP1  | 9.43  | 9.50  | 19.19 | 9.53  | 11.72       | 11.95      |
| CP2  | 7.75  | 7.77  | 17.65 | 7.98  | 11.13       | 13.46      |
| EBD  | 12.01 | 11.89 | 15.76 | 12.24 | 12.69       | 14.10      |
| UQC  | 8.45  | 8.70  | 16.38 | 8.72  | 11.26       | 13.90      |

SVR and MLR come next in the ranking with SVR being slightly better. Even though SVR and MLR shows similar performance, the difference in accuracy between them is statistically significant for all data sets except UQC. Besides, although NN shows the highest accuracy, the performance of SVR and MLR is not too far behind: MRE = 7.75–12.01% for NNs ensemble vs 7.77–11.89% for SVR and 7.98–12.24%

for MLR. In addition, both SVR and MLR outperform kNN and two baselines and their improvements over kNN and baselines are statistically significant too (see Table 2).

On the other hand, kNN provides the lowest accuracy among all the prediction models – they even unexpectedly outperformed by the baselines $B_{msday}$ and $B_{pday}$. From the poor performance of kNN compared to $B_{msday}$, it can be said the using more than 1 similar day to compute the prediction next day is not quite beneficial.

**Table 2.** Statistical significance test (two-sample t-test) of pair-wise differences in accuracy for the prediction models: T = statistically significant at $p \leq 0.05$ and F = not statistically significant. Four letters in a cell indicates results on the four data sets – CP1, CP2, EBD and UQC respectively; for example T, T, F, T in row #1, column #2 indicate that difference in accuracy between NN and SVR are statistically significant for CP1, EBD, and UQC data sets, but not significant for CP2 data set.

|  | NN | SVR | kNN | MLR | $B_{msday}$ | $B_{pday}$ |
|---|---|---|---|---|---|---|
| NN |  | T, F, T, T | T, T, T, T | T, T, T, T | T, T, T, T | T, T, T, T |
| SVR |  |  | T, T, T, T | T, T, T, F | T, T, T, T | T, T, T, T |
| kNN |  |  |  | T, T, T, T | T, T, T, T | T, T, T, T |
| MLR |  |  |  |  | T, T, T, T | T, T, T, T |
| $B_{msday}$ |  |  |  |  |  | T, T, T, T |

**Table 3.** Accuracy of all prediction models after adding noise in weather data.

|  | NN | SVR | kNN | MLR | $B_{msday}$ | $B_{pday}$ |
|---|---|---|---|---|---|---|
| *With 10% added noise in weather data* |  |  |  |  |  |  |
| CP1 | 9.32 | 9.72 | 18.98 | 9.85 | 12.60 | 11.95 |
| CP2 | 8.17 | 8.39 | 18.31 | 8.72 | 12.14 | 13.46 |
| EBD | 12.50 | 12.50 | 17.05 | 12.70 | 14.37 | 14.10 |
| UQC | 8.90 | 9.49 | 17.55 | 9.57 | 12.35 | 13.90 |
| *With 15% added noise in weather data* |  |  |  |  |  |  |
| CP1 | 9.91 | 10.14 | 18.04 | 10.38 | 12.62 | 11.95 |
| CP2 | 8.83 | 9.09 | 17.80 | 9.55 | 12.34 | 13.46 |
| EBD | 12.48 | 13.08 | 17.40 | 13.14 | 14.78 | 14.10 |
| UQC | 9.43 | 10.26 | 17.24 | 10.41 | 12.83 | 13.90 |
| *With 20% added noise in weather data* |  |  |  |  |  |  |
| CP1 | 10.14 | 10.72 | 17.52 | 11.12 | 12.62 | 11.95 |
| CP2 | 9.42 | 9.90 | 17.55 | 10.54 | 12.20 | 13.46 |
| EBD | 13.16 | 13.82 | 17.52 | 13.74 | 14.37 | 14.10 |
| UQC | 10.16 | 11.12 | 17.04 | 11.42 | 12.69 | 13.90 |
| *With 25% added noise in weather data* |  |  |  |  |  |  |
| CP1 | 10.47 | 11.40 | 17.37 | 12.02 | 12.84 | 11.95 |
| CP2 | 9.88 | 10.76 | 17.48 | 11.65 | 12.67 | 13.46 |
| EBD | 13.73 | 14.63 | 17.62 | 14.47 | 14.73 | 14.10 |
| UQC | 10.49 | 12.03 | 17.01 | 12.53 | 13.13 | 13.90 |

It is important to note that results in Table 1 have been computed using measured weather data for future as we do not have access to historical weather prediction. However, in practical applications, the models require predicted weather data $W_{d+1}$ for day $d + 1$ from BoM. The performance of the solar power forecasting models substantially depends on the accuracy of such weather prediction. Therefore, to check the robustness of the prediction models and analyse their sensitivity to the error in weather prediction, we evaluate their accuracy by adding 10–25% random noise to measured weather data.

Table 3 presents accuracy of all prediction models after adding different level of noise in weather information in the test data sets. Figures 2, 3, 4 and 5 shows the comparison of how much the accuracy (MRE) for different models are affected by error in future weather prediction for CP1, CP2, EBD and UQC data sets respectively (kNN are excluded for better visualization since its MRE range is much higher). We can see that only the accuracy of $B_{pday}$ is unaffected by the error in weather data since it considers the power outputs from previous day as the outputs for next day irrespective of the changes in weather.
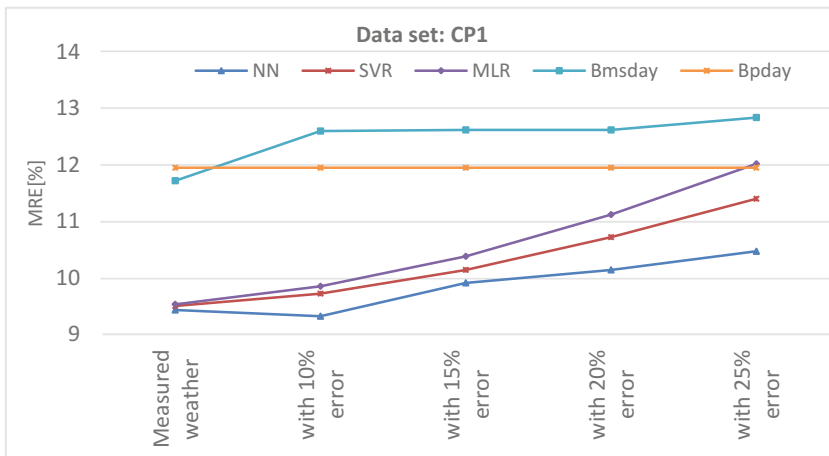


**Fig. 2.** MRE of prediction models for different level of error in weather data: CP1 data set

The prediction error (MRE) for all other models (except kNN) rises and shows an increasing trend as the error in the future weather information increases from 10–25%. The MRE after adding noise in the weather data reaches to the range of 8.17–13.73% for NNs ensemble, 8.39–14.63% for SVR, 8.72–14.47% for MLR, and 11.13–14.78% for $B_{msday}$.

On the other hand the MRE of kNN shows slight improvement or remains similar as error in weather goes up. Although it is unexpected and quite opposite to the case for remaining models, it does not make any difference in the ranking of models as the MRE of kNN is still far behind than two baselines.
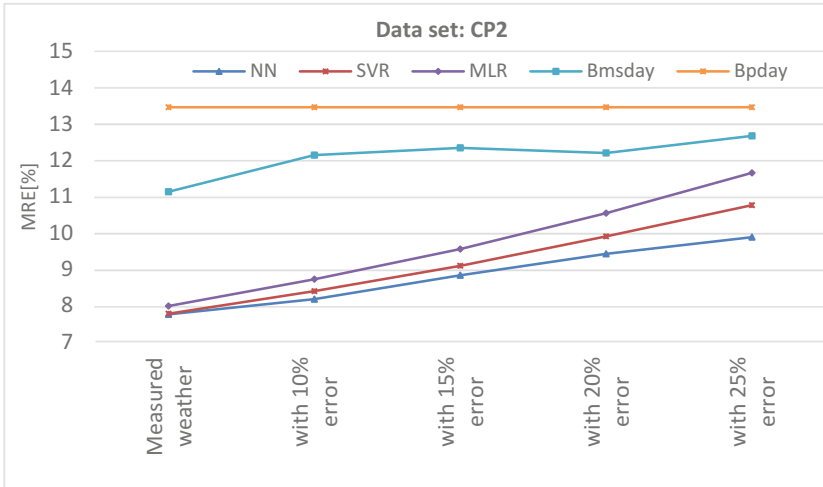
**Fig. 3.** MRE of prediction models for different level of error in weather data: CP2 data set
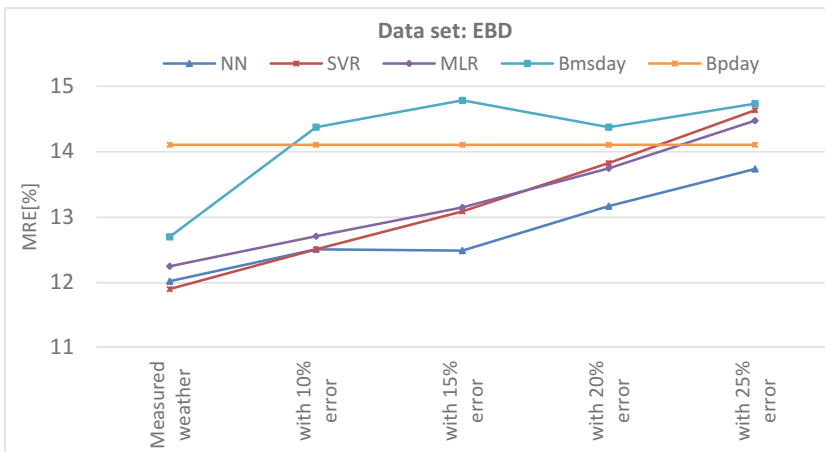


**Fig. 4.** MRE of prediction models for different level of error in weather data: EBD data set

Moreover, comparison of MRE results from Tables 1 and 3 shows that the ranking of all prediction models after adding noise to actual weather data follows the same order as it was before adding the noise: NNs ensemble being the most accurate followed by SVR, MLR, baselines, and kNN. Despite the performance of the all prediction models (except kNN) deteriorates after adding noise to weather data, the difference in accuracy between the NNs ensemble and other models becomes more prominent as the noise goes from 10–25% (see Figs. 2, 3, 4 and 5). This indicates that NN model is more robust in handling the error in weather prediction for future and able to forecast the solar power outputs for next day even the weather profile for next day does not exactly matches with the weather prediction obtained from BoM.
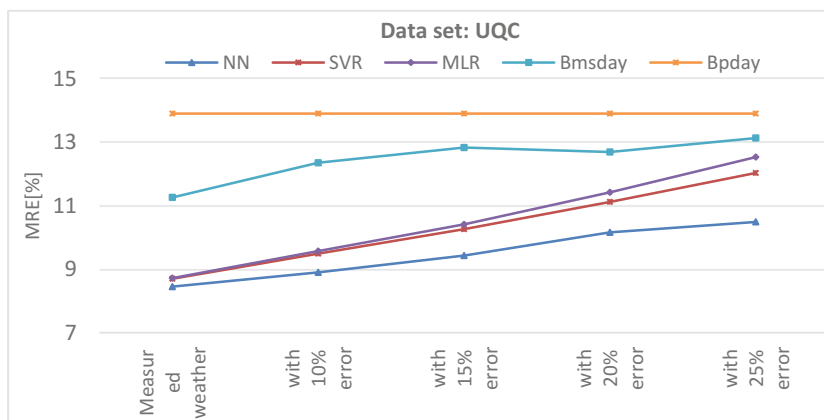
**Fig. 5.** MRE of prediction models for different level of error in weather data: UQC data set

In summary, considering the overall accuracy and ability to handle the uncertainty associated with the weather data for future, ensemble of NNs is the most effective among all the models used for comparison.

## 6    Conclusion

In this paper, we present a comprehensive assessment of a set of prominent methods for forecasting day ahead solar power output profile. We evaluate the performance of ensemble of NNs, kNN, SVR, MLR and two baselines using 4 different sets of data collected for 2 years. The presented results show that ensemble of NNs is the most accurate prediction method and achieves considerable improvement of accuracy over all other methods. Ensemble of NNs also has been found to be very successful in dealing the error in weather prediction - its performance is less sensitive to inaccuracies in weather prediction for future. Although the performance of SVR and MLR also found comparable to NNs, the difference in accuracy between any of these two models and NNs increases significantly as the error in weather forecast increases. Therefore, we conclude that ensemble of NNs is more viable for practical application for forecasting solar power outputs from PV systems.

## References

1. Department of the Environment and Energy, Australian Government: Australian Energy Update (2017). https://www.energy.gov.au/sites/g/files/net3411/f/energy-update-report-2017.pdf. Accessed 05 May 2018
2. European Photovoltaic Industry Association: Connecting the Sun-Solar Photovoltaics on the Road to Large Scale Grid Integration. http://www.pvtrin.eu/assets/media/PDF/Publications/other_publications/263.pdf. Accessed 01 Aug 2017

3. Rana, M., Koprinska, I., Agelidis, V.G.: Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. Energy Convers. Manag. **121**, 380–390 (2016)
4. Chu, Y., Urquhart, B., Gohari, S.M.I., Pedro, H.T.C., Kleissl, J., Coimbra, C.F.M.: Short-term reforecasting of power output from a 48 MWe solar PV plant. Sol. Energy **112**, 68–77 (2015)
5. Rana, M., Koprinska, I.: Neural network ensemble based approach for 2D-interval prediction of solar photovoltaic power. Energies **9**, 829–845 (2016)
6. Shi, J., Lee, W.-J., Liu, Y., Yang, Y., Wang, P.: Forecasting power output of photovoltaic systems based on weather classification and support vector machines. IEEE Trans. Ind. Appl. **48**, 1064–1069 (2012)
7. Wang, Z., Koprinska, I., Rana, M.: Clustering based methods for solar power forecasting. In: International Joint Conference on Neural Networks (IJCNN) (2016)
8. Yang, C., Thatte, A., Xie, L.: Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation. IEEE Trans. Sustain. Energy **6**, 104–112 (2015)
9. Yang, D., Ye, Z., Lim, L.H.I., Dong, Z.: Very short term irradiance forecasting using the lasso. Sol. Energy **114**, 314–326 (2015)
10. Pedro, H.T., Coimbra, C.F.: Assessment of forecasting techniques for solar power production with no exogenous inputs. Sol. Energy **86**, 2017–2028 (2012)
11. Long, H., Zhang, Z., Su, Y.: Analysis of daily solar power prediction with data-driven approaches. Appl. Energy **126**, 29–37 (2014)
12. UQ Solar Photovoltaic Data. http://solar.uq.edu.au/user/reportPower.php. Accessed 01 Aug 2017
13. Climate Data Online. http://www.bom.gov.au/climate/data/. Accessed 01 Aug 2017
14. Rana, M., Koprinska, I., Agelidis, V.G.: Forecasting solar power generated by grid connected PV systems using ensembles of neural networks. In: International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland (2015)
15. Rana, M., Koprinska, I., Agelidis, V.G.: Solar power forecasting using weather type clustering and ensembles of neural networks. In: International Joint Conference on Neural Networks (IJCNN), Canada (2016)
16. Rana, M., Koprinska, I., Agelidis, V.G.: 2D-interval forecasts for solar power production. Sol. Energy **122**, 191–203 (2015)
17. Lora, A.T., Santos, J.M.R., Exposito, A.G., Ramos, J.L.M., Santos, J.C.R.: Electricity market price forecasting based on weighted nearest neighbors techniques. IEEE Trans. Power Syst. **22**, 1294–1301 (2007)