Energy 202 (2020) 117743

Contents lists available at ScienceDirect

Energy

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

Combining forecasts of day-ahead solar power

Chaman Lal Dewangan*, S.N. Singh, S. Chakrabarti

Department of Electrical Engineering Indian Institute of Technology Kanpur, Kanpur, 208016, India

ARTICLE INFO

Article history:
Received 12 September 2019
Received in revised form
20 March 2020
Accepted 27 April 2020
Available online 3 May 2020

Keywords:
Combined-forecast
Day-ahead
Machine learning
Solar power
Numerical weather prediction

ABSTRACT

Solar power forecasting is important for the reliable and economic operation of power systems with high penetration of solar energy. The solar power forecasts for the day-ahead time horizon are more erroneous than the hour-ahead time horizon. Numerical weather prediction (NWP) variables such as irradiance, cloud cover, precipitation etc. are used as input to day-ahead forecasting models. The uncertainty in NWP varies with weather conditions. Different forecasting algorithms based on a single method are available in the literature. Combination of individual forecasting algorithms increases the accuracy of the forecasts. However, the combined-forecast has yet not been analysed much in the area of day-ahead solar power forecasting. This paper thus explores different combined-forecast methods such as mean, median, linear regression and non-linear regressions using supervised machine learning algorithms. The number of models required for day-ahead solar power forecasts is studied. One for all hour (same) or separate models for each hour of the day are possible. The effects of retraining frequency on the performance of the forecasting models, which is important for the computational burden of the system, are also studied. Forecasting algorithms are applied to three solar plants in Australia.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

The intermittent nature of solar and wind power due to their dependency on weather increases the vulnerabilities of power grids. The accurate forecasting of wind and solar conditions [1] helps to manage variability and ramp events in power output of these renewable energy sources. The forecasting of renewable energy, such as solar and wind energy, and load are taken as the input for the unit commitment of hydro and thermal generators, reserve management, balancing, and electricity-market settling [2–6]. Solar power forecasting is also used for battery storage scheduling in distribution systems with high photovoltaic penetration to control energy arbitrage and peak shaving [7–9].

When historical measurements are available, machine learning and statistical techniques are used to determine the relationship between weather forecasts and the observed power. For new renewable energy plant installations, where historical measurements are unavailable, physically based models are employed, which convert weather forecasts into power output by using plant parameters [10].

The inputs for a solar power forecasting model depend on the time horizon of forecasting. The time series of recorded power is sufficient for intra-hour or hour-ahead predictions. Satellite imaging is very useful in forecasts for 1 min to 5 h. Numerical weather prediction (NWP) involves different variables such as the surface irradiance, cloud cover, wind speed, cloud water content, and ice content. NWP variables are used as inputs for 3-h-ahead to two-weeks-ahead solar power forecasting [11–13].

A point forecast is usually the mean of the distribution of a future observation in the time series and is conditional on the previous recorded time series. A probabilistic forecast represents an estimation of the respective probabilities for all the possible future outcomes of a random variable. The probabilistic forecast represents the uncertainty in the forecasted variable in form of a prediction interval or probabilistic distribution [14,15].

Many deterministic and probabilistic forecasting algorithms exist in the literature. The importance of some of these algorithms is reviewed. The solar irradiance and correction factor is predicted to forecast day-ahead solar power generation. The correction factor is adaptive to the degradation of solar modules and change in factors such as the weather, rain, and dust [16]. Solar power forecasting algorithms based on the variational mode decomposition of the solar power time series, the information-theoretic feature selection method, and neural networks trained with hybrid training algorithms, including the dual-population chaotic shark smell optimisation (DPC-SSO) and modified Levenberg-Marquardt

* Corresponding author.
E-mail address: chamanlal.iitk@gmail.com (C.L. Dewangan).

https://doi.org/10.1016/j.energy.2020.117743
0360-5442/© 2020 Elsevier Ltd. All rights reserved.

algorithms, were proposed [17,18]. A day-ahead solar irradiance forecasting model were proposed using deep learning with long-short term memory (LSTM) layers. The prediction accuracy of proposed model was found better than the model using back propagation neural network, and SVR [19–21]. In Refs. [22], a 5-min-ahead solar power forecasting model based on the weighted Gaussian process, which is robust to outlier data, was modelled. A combination of the weighted local outlier factor and nonlinear correlation measurements were used for preparing the forecasting model. Transmissivity-based normalisation and 2D representation of solar irradiation based on support vector regression (SVR) were used to capture the seasonal and daily variations in the forecasting model. The forecasting accuracy was found better in the SVR models than the radial basis function neural network models [23]. Low-pass filters using the Fourier transform were built to capture the annual cyclic nature of solar radiance and power. Gradient-boosting-based regression was used to convert NWP data to solar power data [15]. Linear predictive coding and the 2D-predictive-image-filtering-based forecasting algorithm for solar radiation and wind speed are presented in Ref. [24]. The forecasting methods in Ref. [15,24] are based on filter concepts.

A probabilistic forecast gives uncertainty in the forecasting point. The probabilistic forecast can be obtained directly from the input variables. The probabilistic forecast can also be obtained first through a point forecast and then extended to the probabilistic forecast. Huang et al. [15] estimated the probability distribution of solar power generation in the form of quantiles from deterministic forecasting by using the k-nearest neighbour (kNN) regression method. A hybrid method [14] involving the combination of wavelet decomposition, feature selection based on mutual information, bootstrap approaches in an ensemble structure, and an extreme learning machine, was proposed for the point forecast. The variance was estimated using the maximum likelihood estimation and particle swarm optimisation algorithm to obtain the prediction interval from the point forecast. The radial basis function extracted the nonlinear dependencies of the input variables. Then, multiple quantile regression methods were used to estimate the probability distribution of future outcomes [25]. This proposed method is based on a direct prediction of the probabilistic forecast from the input variables.

When different forecasting models have similar accuracy, deciding the best model is difficult [26]. combined-forecasts generally increase the accuracy of forecasting by averaging out the bias and reducing the error variance so that their error correlation is small. Load forecasting using many combination methods, such as simple mean, median, linear regression, and rolling window are analysed [27]. The regression combination method outperformed the other competitive combination methods. A combination of day-ahead forecast and the ramp-rate forecast was done to forecast hour-ahead solar power and was found better than compared individual models [28].

To the best of the authors' knowledge, combined-forecast of day-ahead solar power has not been extensively studied yet. Improving accuracy and reducing the computational burden of forecasting models are the primary motivation behind the research work proposed in this paper. The main contributions of the research work are as follows:

- Different combined-forecast methods, such as simple mean, median, linear regression, and nonlinear regression such as SVR and GPR, were analysed for the day-ahead solar power forecasting models.
- Number of models required for day-ahead solar power forecasts was studied. One-for-all-hours or separate models for each sunny hour of the day can be utilized, the performance of the forecasting models was analysed.
- When the forecast models are prepared and implemented. The forecast models have to be retrained so that the forecast models can learn from the recent data and maintain the accuracy of the forecasting. Different retraining frequencies such as daily, weekly, and monthly were examined.

The individual and combined-forecasts are implemented to three solar plants test-beds. The remainder of the paper is arranged as follows: Section 2 describes the data used in the analysis. Section 3 explains the procedure of day-ahead solar power forecast. Section 4 includes simulation results and discussions. Section 5 concludes the study and informs about future works.

2. Data description

Twelve NWP variables and recorded power time series from April 1, 2012 to June 30, 2014 for three solar farms in Australia were used in this study. The data was used in the Global Energy Forecasting Competition 2014 (GEFCom2014) and is available online [29]. The NWP data were available for midnight, and the aim was forecasting the solar power for next 24 h of the day. The list of variables used is provided in Table 1. Some variables were directly taken as NWP output, whereas others were derived or converted into a useable format [25]. Solar radiation was obtained in the form of cumulative energy (J/m^2) and had to be converted into the power format (W/m^2). Differential cloud cover refers to the difference in the cloud cover at time t with respect to that in the previous or coming hours, $t \pm h$. The differential surface pressure is similarly defined and represents the change in weather. The wind chill index is an indicator of the cooling of the solar photovoltaic module due to the blowing wind. The power output of the solar photovoltaic panel depends on the module temperature. The module is cooled by wind and heated by solar radiation. Clear sky solar radiation depends on the annual and diurnal periodicity (i.e., the day of the year and hour of the day).

Normalisation of every variable was performed using the min-max normalisation formula. Normalisation makes the training less sensitive to the scale of input variables and well-conditioned for optimisation in machine learning algorithms. The normalisation formula is given as follows:

$$\mathbf{u} = \frac{\mathbf{z} - \min(\mathbf{z})}{\max(\mathbf{z}) - \min(\mathbf{z})} \tag{1}$$

where $\mathbf{z}$ and $\mathbf{u}$ represent the actual and normalised value of the variables, respectively. Because the minimum value of solar power output is zero, the formula becomes,

$$\mathbf{u} = \frac{\mathbf{z}}{\max(\mathbf{z})} \tag{2}$$

3. The procedure of day-ahead solar power forecasting

3.1. Input variable selection

In day-ahead solar power forecasting, NWP output variables and previously recorded power are used. Deciding the effective variables and their lags to be used is important for preparing an accurate forecasting model (i.e., with good generalisation capability). To obtain the optimal feature subset, the redundant variables should be eliminated [30,31].

Table 1
Input variables for solar power forecasting.

Input Type	Input parameters
NWP variables	Total column liquid water (kg/m^2)
	Total column ice water (kg/m^2)
	Surface Pressure (sp) in (Pa)
	Relative humidity at 1000 mbar
	Total cloud cover (cc) value between 0 and 1
	Total precipitation (m)
	2-m temperature T_{amb} in degree Celcius
Derived variables	Surface solar radiation down (I_s) in (W/m^2)
	Top net solar radiation (W/m^2)
	Surface thermal radiation (W/m^2)
	10 m equivalent wind speed (v) (m/s)
	Total cloud cover $\times$ relative humidity
	Total cloud cover $\times$ surface solar radiation
	Differential cloud cover: ($cc_t - cc_{t+h}$)
	Differential surface pressure: ($sp_t - sp_{t+h}$)
	Wind chill index:
	$(10\sqrt{v} - v + 10.5) \times (33 - T_{amb})$
	Photovoltaic module temperature: $T_{amb} + I_s.e^{-3.473-0.0594v}$
	Periodicity due to day of the year:$\cos\left(\frac{day}{365}2\pi\right), \sin\left(\frac{day}{365}2\pi\right)$
	Periodicity due to hour of the day:
	$\cos\left(\frac{hour}{24}2\pi\right), \sin\left(\frac{hour}{24}2\pi\right)$

The sequential feature selection algorithm was used for selecting the input variables. The direction of the sequential selection can be either forward or backward. In the forward direction, the algorithm starts from an empty feature set and the next best feature is added in every iteration. In the backward direction, the algorithm starts from a full feature set and the worst feature is removed in every iteration. The algorithm stops when the mean square error stops decreasing. Forward selection may skip highly informative combinations of input variables. Backward selection may be unable to re-evaluate the feature after it has been removed. Both directions should be examined, and the subset that reduces the mean square error to a minimum should be found [32,33].

3.2. Forecasting models

A forecasting model can be viewed as a regression model. Regression refers to the function approximation between the input and output sets [34]. Day-ahead solar photovoltaic (PV) power output forecasting is a nonlinear function of weather forecast variables such as the irradiance, temperature, cloud cover, and liquid water content in the cloud [25]. The regression model provides the average prediction for the given input by learning the training patterns and decreases the effect of uncertainty present in the input variables given by NWP.

Mathematically, regression can be expressed as $f : \mathbf{X} \rightarrow \mathbf{Y}$ mapping $\mathbf{x} \in \mathbf{X}$ to $y \in \mathbf{Y}$, where $\mathbf{x}$, $\mathbf{X}$, y, $\mathbf{Y}$ are input variables vector, input space, solar power output, and output space, respectively. Training data set is $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x_m}, y_m)\}$, where m is the number of training samples and $\mathbf{x} \in \mathbf{R}^d$, d is the size of the input vector.

The following regression models were used in this study:

3.2.1. Persistence

The persistence method is the simplest forecasting technique. The persistence method does not consider the effects of change in the inputs. It gives forecast value $\widehat{y}(t)$ equal to previous recorded value $y(t-1)$. This method performs well when previous and forecast conditions are similar. This method is generally taken as reference for comparing the performance of the other forecasting models.

3.2.2. Linear regression interacts (LRI)

Linear regression is a simple supervised learning method. It gives a linear function between inputs and output. For two input variables and one output variable, it can be expressed as,

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \varepsilon \tag{3}$$

where $\alpha_0, \alpha_1, \alpha_2$ are the parameters, x_1, x_2 are the inputs, y is the output, and ε is the error term. When an input variable has a different effect on the output depending on the values of other input variables, this is called interaction between input variables. The interactions can be considered in the modelling of the linear regression as,

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \beta_1 x_1 x_2 + \varepsilon \tag{4}$$

where β_1 is the interaction parameter.

2.2.3. Feed forward neural network (FFNN)

An FFNN is a nonlinear parametric regression model. In this network, the data flows in only one direction, from the input neurons, through the hidden neurons, and to the output neurons. The number of hidden neurons and activation functions for the hidden and output layers are decided to create the network. The activation functions for the hidden layer, such as tan-sigmoid, log-sigmoid, and wavelet functions, are generally nonlinear and differentiable functions so that the network can grasp the nonlinear relationships between the input and output vectors in the training patterns. The output layer of the FFNN is generally linear when the FFNN is used for regression applications. Training, validation, and testing are three important steps of modelling. Training starts with the initialisation of weights and biases of the network and parameters of the activation functions. Backpropagation training algorithms, such as Levenberg-Marquardt, scaled conjugate gradient, and gradient descent, are used to obtain the optimal weights, biases, and parameters. Validation avoids the overfitting of the FFNN during training. Training stops when the performance criterion (mean square error) is satisfied or the iterations reach the maximum set value. Testing compares the performance of different models [35,36]. The architecture of the FFNN with four input

variables, three hidden neurons and one output variable is displayed in Fig. 1.

3.2.4. Deep neural networks with long short-term memory (LSTM)

Deep neural networks (DNN) have multiple layers with non-linear activation functions, which makes it an expert in learning or approximating non-linear function between inputs and outputs [37]. DNN with long-short term memory (LSTM) layers is found very successful in sequence forecasting [19–21]. The architecture of the LSTM unit is similar to a recurrent neural network (RNN). The general architecture of an LSTM unit has a memory and three gates: an input gate, an output gate and a forget gate. These three gates are responsible for updating the information of dependencies between the elements in the input sequence, in the memory cell of the LSTM unit [38,39]. The training of DNN with LSTM is similar to FFNN such as error backpropagation and optimize the parameters of the network using gradient descent. LSTM units also eliminate gradient vanishing problem in conventional RNN during back-propagation training. The DNN with LSTM layers, which is used in the analysis, is shown in Fig. 2.

3.2.5. Support vector regression (SVR)

SVR is a method for function estimation using support vector machines. Function predictions depend on support vectors, which are a subset of the training data [40]. Linear function approximation using SVR can be expressed as follows:

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2 + E\sum_{i=1}^{m}(\beta_i + \beta_i^*) \\ &\text{subject to} \quad y_i - \langle \mathbf{w}, \mathbf{x}_i\rangle - c \le \varepsilon + \beta_i \\ &\langle \mathbf{w}, \mathbf{x}_i\rangle + c - y_i \le \varepsilon + \beta_i^* \\ &\beta_i, \beta_i^* \ge 0 \end{aligned} \tag{5}$$

where ε is the allowed deviation from true output value (errors lower than ε are not considered),**w** is the weight vector, and c is intercept of the approximated regression line. A small value of **w** is required for a flat function. β_i, β_i^* are slack variables that allow soft margin errors so that a feasible solution can be obtained. E determines the importance of the flatness of a function and the number of errors larger than ε that are tolerated.$\langle \mathbf{w}, \mathbf{x}_i\rangle$ represents dot product of two vectors.

Problem formulation includes a quadratic cost function with linear constraints. A linear epsilon-insensitive loss function is used in the problem formulation. Lagrange dual formulation can be used to easily solve the optimisation problem, which allows support vector machines to be extended to nonlinear functions. The sequential minimal technique is generally applied to determine the optimal number of support vectors from the training patterns [41–43]. Different kernel functions, such as linear, Gaussian, and polynomial kernel functions, are available. The generalising capability of SVR depends on the kernel function used to prepare the forecasting model.

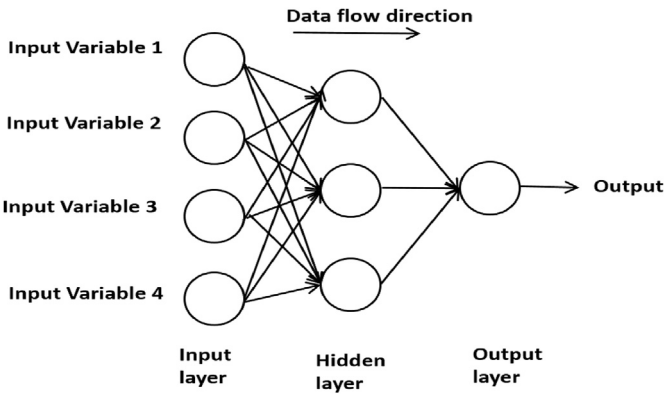


Fig. 1. FFNN architecture.

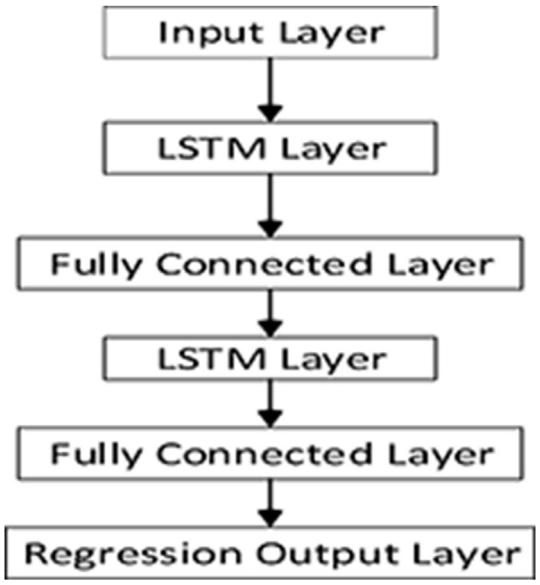


Fig. 2. DNN with LSTM layers.

3.2.6. Gaussian process regression (GPR)

GPR is a nonparametric kernel-based probabilistic model. A GPR model defines the response by using latent variables $g(\mathbf{x}_i, i = 1, 2, \ldots, m)$ from a Gaussian process (GP) and explicit basis functions e. The covariance function of the latent variables learn the smoothness of the response, and the basis functions project the inputs **x** into a q-dimensional space.

A GP is a collection of random variables such that any finite collection of the random variables have a joint Gaussian distribution. If $g(\mathbf{x})$ is a GP, the training samples $\mathbf{x}_1$,$\mathbf{x}_2$,…, $\mathbf{x}_m$ and joint distribution of the random variables $g(\mathbf{x}_1)$, $g(\mathbf{x}_2)$,…, $g(\mathbf{x}_m)$ are Gaussian. A GP has a mean function $b(\mathbf{x})$ and covariance function $h(\mathbf{x}, \mathbf{x}')$ that can be expressed as follows:

$$\begin{aligned} &E(g(\mathbf{x})) = b(\mathbf{x}) \\ &Cov[g(\mathbf{x}), g(\mathbf{x}')] = E[\{g(\mathbf{x}) - b(\mathbf{x})\}\{g(\mathbf{x}') - b(\mathbf{x}')\}] = h(\mathbf{x}, \mathbf{x}') \end{aligned} \tag{6}$$

In GPR, a response y can be modelled as follows:

$$P\left(y_i | g(\mathbf{x}_i), \mathbf{x}_i\right) \sim N\left(y_i | \mathbf{r}(\mathbf{x}_i)^T \boldsymbol{\eta} + g(\mathbf{x}_i), \sigma^2\right) \tag{7}$$

where y_i is the response of input vector $\mathbf{x}_i$, $g(\mathbf{x}_i)$ is a latent variable, $\mathbf{r}(\mathbf{x}_i)$ transforms the d dimensions of $\mathbf{x}_i$ into q-dimensional space, $\boldsymbol{\eta}$ is a q-dimensional vector of the basis function coefficients, and σ^2 is the estimated error variance for $\mathbf{x}_i$.

The covariance function of the latent variables, which can be defined by many kernel functions, specifies the similarity between two latent variables. The prediction accuracy of the model depends on the covariance function selected. Different available kernel (covariance) functions include the squared exponential, rational quadratic, and Matern functions. The hyperparameters of the

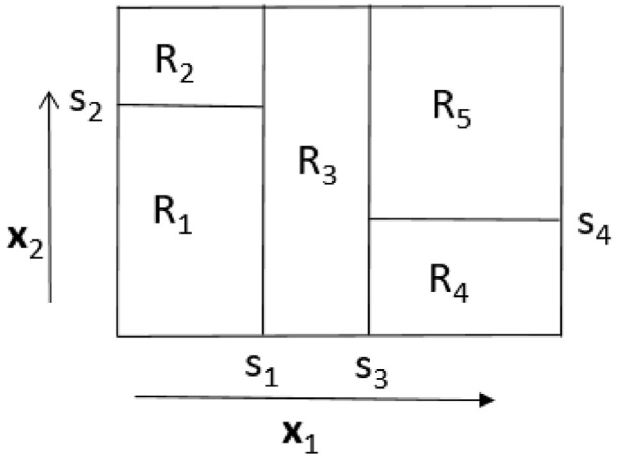


Fig. 3. Splitting of the predictor space [33].

covariance functions are obtained by maximising the log-likelihood function [41,44,45].

3.2.7. Ensemble trees regression (ETR)

Tree-based regression methods are also nonparametric. Tree-based regression methods involve segmenting the predictor space into numerous simple regions. The predictor space is divided by the values of d input variables $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d)$ into V nonoverlapping regions $R_1, R_2, \ldots, R_V$. The set of splitting rules is applied to grow the tree. Therefore, these methods are also called decision tree methods. For each region, R_v, a prediction is performed by calculating the mean or mode of the observations present in that region [46]. Recursive binary splitting is applied to grow the tree from top to down. Trees begin from the root (beginning node) and end at the leaf nodes. A leaf node provides the prediction for the given input. We wish to determine the values of variables that split the nodes and minimise the residual sum of squares $\sum_{v=1}^{V} \sum_{i \in R_v} (y_i - \hat{y}_{R_v})^2$, where $\hat{y}_{R_v}$ is the prediction value of the region R_v, and y_i is the actual value of the response. The intermediate nodes are split according to the importance rankings of the prediction variables such that prediction error reduces until each terminal node has less than a minimum number of training observations. The splitting of the two-dimensional predictor space into five regions and the corresponding tree are displayed in Figs. 3 and 4, respectively. Where S_1, S_2, S_3, and S_4 are values at which the predictor space are split.

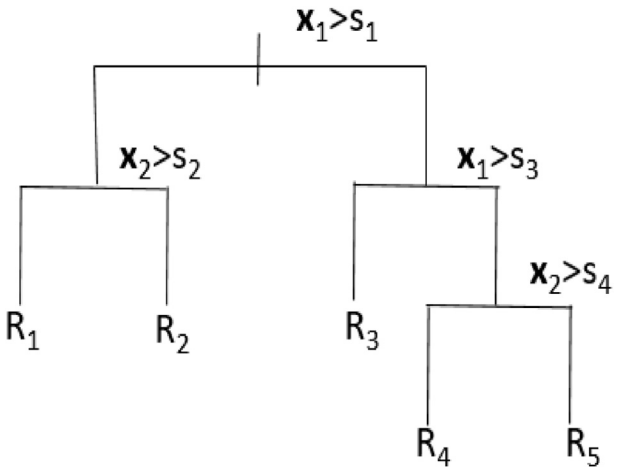


Fig. 4. Interpreting the regions in form of a tree [33].

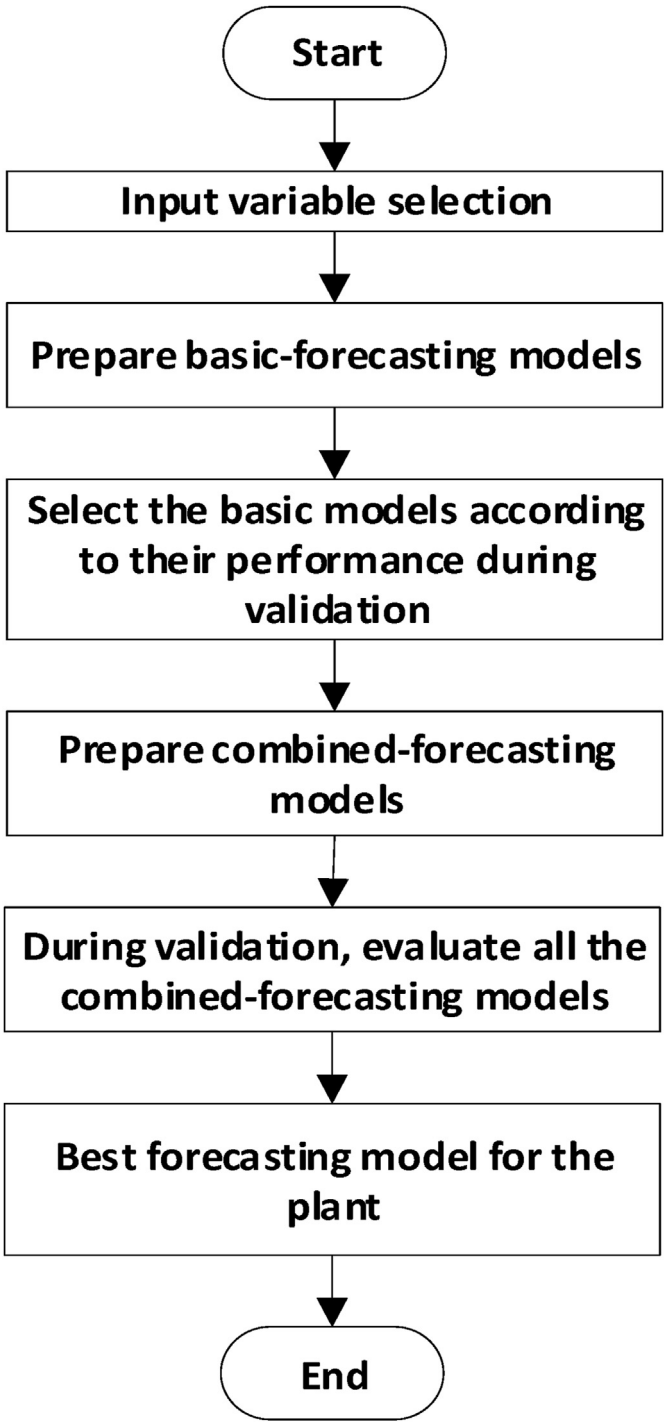


Fig. 5. Flowchart of the combining forecasts algorithm.

The trees are pruned to avoid overfitting and poor test performance. A simple tree and complex tree suffer from high bias and high variance, respectively. Ensemble algorithms such as bagging and boosting, which combine many trees, can improve the predictive performance [46,47]. In bagging, bootstrapping is performed by taking repeated samples from the training data sets. The model is trained using each bootstrapped dataset, and the

Table 2
Comparison of hourly and one-for-all-hours forecasting models based on LRI method.

Days	Root mean square error in percentage					
	Plant 1		Plant 2		Plant 3	
	Hourly	one-for-all-hours	Hourly	one-for-all-hours	Hourly	one-for-all-hours
Month 1	11.32	9.53	21.07	21.17	13.53	10.75
Month 2	15.75	11.9	20.42	20.29	14.24	11.18
Month 3	17.47	12.85	13.03	12.88	16.2	11.5
Month 4	14.5	11.91	10.72	10.79	16.11	10.19

Table 3
Comparison of hourly and one-for-all-hours forecasting models based on SVR method.

Days	Root mean square error in percentage					
	Plant 1		Plant 2		Plant 3	
	Hourly	one-for-all-hours	Hourly	one-for-all-hours	Hourly	one-for-all-hours
Month 1	10.93	9.21	19.31	18.83	12.17	9.63
Month 2	14.8	11.33	14.12	14.29	14.73	11.8
Month 3	18.94	12.67	12.8	12.35	14.79	11.58
Month 4	14.18	11.09	9.58	10.01	12.1	9.95

prediction is obtained. Finally, the average of all the bagged predictions gives the prediction for the given input. In boosting, trees are grown successively by using information from previously grown trees. A tree is fit to the residuals of the previously grown trees. Then, the decision tree is added to the fitted function for updating the residuals.

3.2.8. Combined-forecasts

Combined-forecasts are done when the task of determining the best or dominant forecasting model among several similar performing forecasting models is difficult [35]. Different forecasting models have different methods of learning. Thus, the forecast outputs have different information, which can be blended to obtain an accurate model. Candidate forecasting models should have satisfactory accuracy to participate in combined-forecasts, and poorly performing forecasting models should not be considered when forming the combined-forecasting model. Combined-forecasting methods include simple average with equal weights, median, linear regression, and nonlinear regression.

2.3. Performance evaluation

The prediction accuracies of the forecasting models are evaluated in terms of the mean absolute percentage error (MAPE) and root mean square error (RMSE). The percentage of the MAPE and RMSE is calculated with respect to the maximum power of the power plant.

$$MAPE = \frac{1}{R}\sum_{p=1}^{R}\left|y_p - \hat{y}_p\right| * 100\% \tag{8}$$

$$RMSE = \sqrt{\frac{1}{R}\sum_{p=1}^{R}\left|y_p - \hat{y}_p\right|^2} * 100\% \tag{9}$$

where $\hat{y}_p$ and y_p are the normalised forecasted and measured power of a time point p, respectively, and R is the total number of forecasting points.

The algorithm of combined-forecast is illustrated in Fig. 5 with the help of a flow chart.

4. Simulation results and discussions

The proposed forecasting algorithm was applied to three solar power plants in Australia (as described in Section 2) and simulated using MATLAB software [48].

Using the sequential feature selection algorithm, twenty one effective variables were obtained for the day-ahead solar power forecasting models, which also included the lag of some variables, such as the previous and next hours. The variables included NWP variables and recorded solar power time series.

Five-fold cross-validation was done, while preparing the forecasting models based on machine-learning methods. Data were divided into two sets, namely; training and testing data. The training data were further split into five parts. Four parts were used for training the model, and the fifth part was used for validating the model.

Table 4
Effects of different retraining frequency on the performance of LRI method.

Root mean square error in percentage						
Days	Plant 1			Plant 2		
	Daily	Weekly	Monthly	Daily	Weekly	Monthly
Month 1	9.47	9.49	9.52	21.22	21.2	21.17
Month 2	11.87	11.85	11.9	20.27	20.29	20.27
Month 3	12.77	12.77	12.85	12.25	12.49	12.88
Month 4	11.81	11.87	11.91	10.64	10.82	10.79

Table 5
Effects of different retraining frequency on the performance of SVR method.

Root mean square error in percentage						
Days	Plant 1			Plant 2		
	Daily	Weekly	Monthly	Daily	Weekly	Monthly
Month 1	9.17	8.93	9.01	19.09	18.75	18.96
Month 2	10.84	11.4	11.17	13.65	13.84	13.91
Month 3	12.58	12.67	12.94	13.08	12.91	12.65
Month 4	11.29	11.2	11.16	9.97	10.36	10.12

Table 6
Abbreviations used in the following tables.

ETR	Ensemble tree regression
FFNN	Feed forward neural network
GPR	Gaussian process regression
GPRMa	Gaussian process regression with Matern kernel
LR	Linear regression
LRI	Linear regression interacts
Persis	Persistence
SVRG	Support vector regression with Gaussian kernel
SVRL	Support vector regression with linear kernel
SVRQ	Support vector regression with quadratic kernel

Note: suffix 'C' is used with the name of regression to represent the combined-forecast models.

An FFNN with ten hidden neurons, Levenberg-Marquardt training algorithm, tan-sigmoidal activation function for the hidden layer, and linear activation function for the output layer were used. SVR with the quadratic and the Gaussian kernel functions were implemented and found to be more accurate than SVR with linear and cubic kernels. Similarly, GPR with Matern 5/2 kernel function, which was more accurate than the squared exponential, exponential, and rational quadratic kernels, was applied. The ETR with a minimum leaf size of 8 and the bag ensemble algorithm, which performed better than the boost ensemble algorithm, was modelled. DNN with two LSTM layers and two fully connected layers were implemented. The output layer of the DNN was the regression layer, which calculates the square of the difference

Table 7
RMSE performance comparison of basic and combined-forecasting models of plant 1.

Rank	Month 1		Month 2		Month 3		Month 4	
	Models	RMSE	Models	RMSE	Models	RMSE	Models	RMSE
1	MeanC	7.79	MeanC	9.8	SVRGC	12.11	SVRGC	9.97
2	LSTMC	7.83	MedianC	9.81	LSTMC	12.11	LSTMC	10.11
3	SVRGC	7.87	LSTMC	9.85	LRIC	12.13	SVRLC	10.14
4	LRC	7.89	SVRGC	9.95	GPRMaC	12.14	MeanC	10.15
5	SVRLC	7.92	LRC	9.98	MeanC	12.17	MedianC	10.19
6	MedianC	7.94	SVRLC	9.99	SVRLC	12.18	LRIC	10.22
7	GPRMaC	7.95	LRIC	10.06	LRC	12.24	GPRMaC	10.25
8	LRIC	8.04	GPRMaC	10.1	MedianC	12.26	LRC	10.28
9	SVRG	8.64	ETR	10.22	SVRQ	12.86	LSTM	10.64
10	FFNN	8.77	LSTM	10.67	LSTM	12.93	SVRQ	10.69
11	GPRMa	8.91	SVRQ	10.76	SVRG	13.13	SVRG	11.28
12	SVRQ	9.06	FFNN	10.84	ETR	13.14	FFNN	11.35
13	LRI	9.21	SVRG	11.11	LRI	13.16	LRI	11.43
14	ETR	9.37	GPRMa	11.18	FFNN	13.32	ETR	11.45
15	LSTM	9.78	LRI	11.24	GPRMa	13.78	GPRMa	12.66
16	Persis	18.43	Persis	18.71	Persis	20.63	Persis	20.24

Table 8
RMSE performance comparison of basic and combined-forecasting models of plant 2.

Rank	Month 1		Month 2		Month 3		Month 4	
	RMSE	Models	RMSE	Models	RMSE	Models	RMSE	Models
1	LRIC	15.36	SVRLC	7.37	SVRLC	12.23	MedianC	8.77
2	LRC	15.37	SVRGC	7.41	SVRGC	12.28	MeanC	8.88
3	LSTMC	15.37	MedianC	7.41	LSTMC	12.38	SVRGC	8.9
4	GPRMaC	15.4	MeanC	7.53	LRIC	12.39	SVRLC	8.9
5	MeanC	15.44	LSTMC	7.54	LRC	12.44	LSTMC	8.93
6	MedianC	15.5	GPRMaC	7.54	MeanC	12.45	LRC	9.09
7	GPR Ma	15.58	LRIC	7.55	GPRMaC	12.45	LRIC	9.16
8	SVRGC	15.72	LRC	7.56	MedianC	12.55	GPRMaC	9.16
9	SVRLC	15.79	ETR	8.07	ETR	12.73	SVRQ	9.75
10	SVRG	15.95	LSTM	8.68	LSTM	12.96	LSTM	9.97
11	ETR	16.58	SVRQ	8.75	SVRQ	13.21	SVRG	10.01
12	LSTM	16.7	SVRG	8.98	FFNN	13.29	GPR Ma	10.1
13	SVRQ	17.16	GPR Ma	9.1	LRI	13.66	ETR	10.13
14	LRI	17.79	LRI	9.38	SVRG	14.03	LRI	10.26
15	FFNN	17.97	FFNN	9.51	GPR Ma	14.43	FFNN	10.29
16	Persis	20.23	Persis	21.04	Persis	19.47	Persis	20.78

Table 9
RMSE performance comparison of basic and combined-forecasting models of plant 3.

Rank	Month 1		Month 2		Month 3		Month 4	
	Models	RMSE	Models	RMSE	Models	RMSE	Models	RMSE
1	LRC	8.71	LSTMC	9.19	LSTMC	10.72	GPRMaC	9.11
2	MeanC	8.75	LRIC	9.25	LRIC	10.76	SVRGC	9.14
3	LSTMC	8.76	LRC	9.31	MeanC	10.76	LRIC	9.18
4	LRIC	8.8	GPRMaC	9.33	LRC	10.79	SVRLC	9.22
5	SVRLC	8.85	MeanC	9.38	GPRMaC	10.84	MeanC	9.25
6	MedianC	8.85	MedianC	9.41	SVRGC	10.88	LRC	9.27
7	SVRGC	8.86	SVRGC	9.56	SVRLC	10.88	LSTMC	9.37
8	GPRMaC	8.87	SVRLC	9.62	MedianC	10.93	MedianC	9.41
9	LSTM	9.64	GPR Ma	10.14	FFNN	11.66	LSTM	9.93
10	SVRG	9.92	ETR	10.28	LRI	11.7	SVRQ	10.09
11	SVRQ	9.95	LSTM	10.41	ETR	11.71	FFNN	10.26
12	ETR	9.96	SVRG	10.65	SVRQ	11.78	SVRG	10.28
13	FFNN	10.22	LRI	10.71	LSTM	12.1	LRI	10.32
14	LRI	10.62	FFNN	10.86	SVRG	12.12	ETR	10.48
15	GPR Ma	10.84	SVRQ	10.91	GPR Ma	12.48	GPR Ma	10.71
16	Persis	21.12	Persis	19.95	Persis	19.57	Persis	19.41

Table 10
MAPE performance comparison of basic and combined-forecasting models of plant 2.

Rank	Month 1		Month 2		Month 3		Month 4	
	Models	MAPE	Models	MAPE	Models	MAPE	Models	MAPE
1	SVRGC	9.33	SVRGC	4.03	SVRGC	7.85	MedianC	5.7
2	MeanC	9.36	SVRLC	4.05	SVRLC	7.88	MeanC	5.81
3	SVRLC	9.39	MeanC	4.41	MeanC	8.02	SVRLC	5.82
4	LSTMC	9.41	MedianC	4.44	MedianC	8.05	SVRGC	5.83
5	GPRMaC	9.41	GPRMaC	4.53	LSTMC	8.09	LSTMC	5.92
6	LRC	9.42	LSTMC	4.54	LRC	8.21	LRC	6.02
7	LRIC	9.43	LRC	4.57	LRIC	8.21	GPRMaC	6.05
8	MedianC	9.43	LRIC	4.57	GPRMaC	8.22	LRIC	6.06
9	ETR	10.14	ETR	4.94	ETR	8.48	LSTM	6.43
10	GPR Ma	10.14	SVRQ	5.36	SVRQ	8.7	SVRQ	6.49
11	SVRG	10.17	LSTM	5.66	LSTM	8.8	SVRG	6.54
12	Persis	10.48	GPR Ma	5.87	FFNN	8.89	GPR Ma	6.61
13	SVRQ	10.86	SVRG	5.9	LRI	9.01	ETR	6.62
14	FFNN	10.96	LRI	5.95	SVRG	9.02	FFNN	6.76
15	LSTM	11.06	FFNN	6.21	GPR Ma	9.42	LRI	7.08
16	LRI	11.14	Persis	11.08	Persis	11.83	Persis	12.98

between actual and predicted values by model.

The analysis of factors affecting the performance of the forecasting models is discussed in the following subsections.

4.1. Difference between solar power and load demand forecasting

Solar power output from PV panels depends on solar irradiation and weather variables. Load demand depends on temperature, humidity and day of the week. Load demand is generally continuous whereas solar power is discontinuous because sunlight is unavailable during some part of the day. It is important to note that the dataset is only considered for training solar forecasting regression model when the solar radiation is non-zero (discard the data set while solar radiation is unavailable). Removing non-essential training data sets improves the performance of solar power forecasting models.

4.2. The outlier management of day-ahead solar power forecasting

The regression-based forecasting models may give a negative or very high value of solar power due to the noise in the inputs or the sub-optimal parameters of the forecasting models, which is called an outlier. The outliers in solar power forecasts can be eliminated by setting the upper and the lower limit of the predicted solar power. The lower limit is usually set to zero; which means there is no sunlight. The solar power of the clear sky gives the upper limit of the solar forecast. Clear sky solar power can be predicted using radiative models [49]. However, the dataset used here is of GEFCom2014 in which the exact plant locations were not shared. Hence, another method was applied to find the upper limit of forecasts of solar power.

To determine the upper limit of solar power forecasts for any hour, *t*, of the day, recent past 15 days in the same year, 15 days before and 15 days after the forecast date in the previous year can be used to create a vector of solar power for hour *t*. The maximum value of the solar power vector gives the upper limit of solar power forecasts for hour *t* of the day. The prevented outliers in the forecast outputs improved the overall prediction performance of the models.

4.3. Hourly and one-for-all-hours models

Solar power is available during some hours of the day. One forecasting model for all the sunny hours of the day or separate forecasting models for each sunny hour of the day can be prepared. Comparison of these two models were studied.

Tables 2 and 3 list the RMSE performance of the hourly and one-for-all-hours forecasting models based on LRI and SVR methods,

respectively. The forecasting performance of one for all hour models was found better than that of the hourly models. The computational burden of one-for-all-hours model is also less compared to the hourly models because only one forecasting model needs to be prepared.

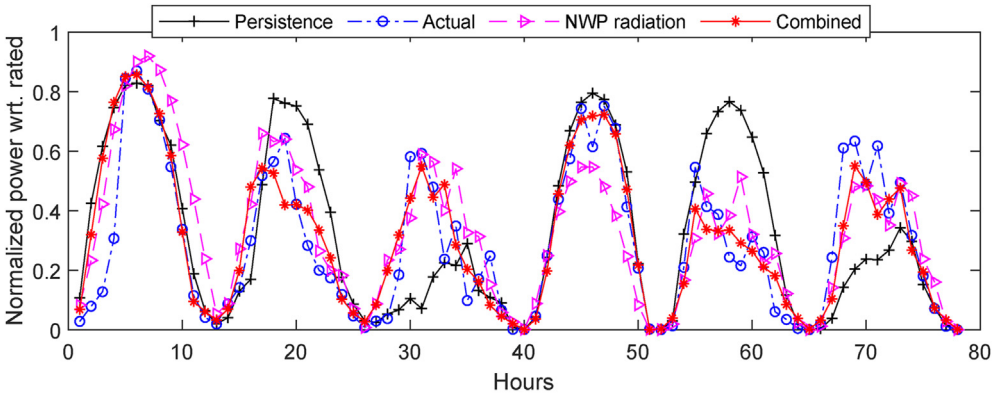
4.4. Effect of retraining frequency on the performance of the forecasting models

When the forecasting model is prepared and implemented. The model has to be retrained so that it can learn from recent data and maintain its performance. It can be retrained daily, weekly or monthly. The effect of different retraining frequencies on the linear regression (LRI) and non-linear regression (SVR) based forecasting models was analysed.

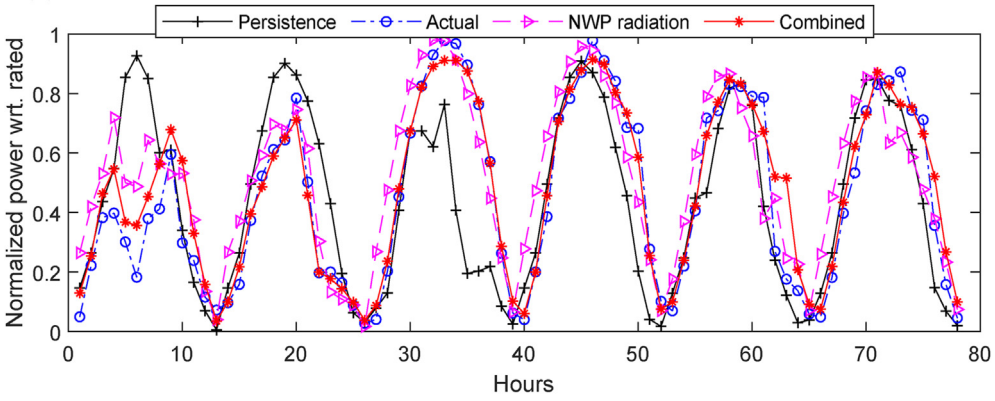
Tables 4 and 5 show the performance of the daily, weekly, and monthly retrained-forecasting models based on LRI and SVR methods, respectively. The performance of LRI based models improved slightly with increase in the retraining frequency. SVR model is a non-linear model that exhibhits stochastic nature for different retraining frequencies. The performance of the forecasting models for daily, weekly, and monthly was found similar. This implies that a monthly retraining frequency is sufficient, which in turn reduces the computational burden of retraining the forecasting models daily and weekly.

4.5. Combined-forecasts

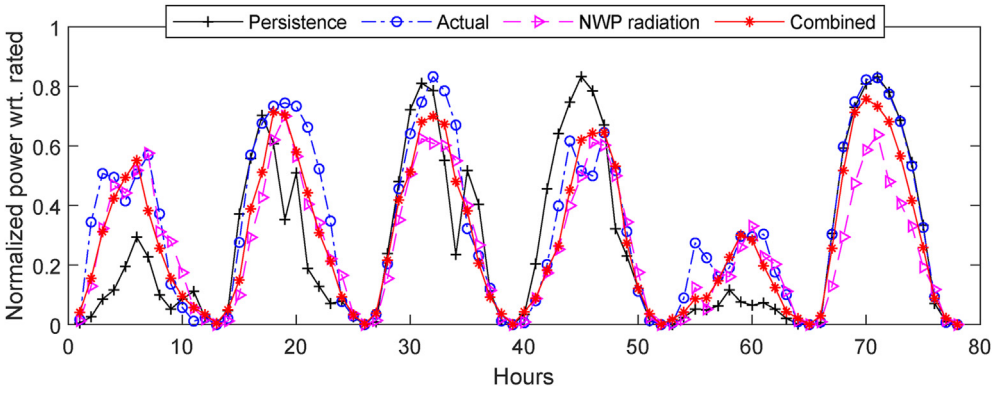
The basis of selection of candidate forecasts for combined-forecast was the RMSE performance during cross-validation. The



(a) Results of some days of plant 1



(b) Results of some days of plant 2



(c) Results of some days of plant 3

Fig. 6. Comparison of forecasting models performance of all three solar plants.

three best performing basic models out of eight basic models, were considered. Combined-forecasts were modelled using simple mean, simple median, linear regression, linear regression interacts, and nonlinear regression such as the Gaussian and linear kernel based SVR, Matern 5/2 kernel based GPR, and deep networks with LSTM layers. The output of the selected basic models were used as input and were supervised by the recorded value of solar power for training the combined-forecast models.

The list of the abbreviations used in the following tables is given in Table 6. 'C' suffix was added to the name of the combined algorithm to differentiate the basic and the combined models listed in the tables. The RMSE performance of the combined-forecasts and basic models of all the three solar plants are provided in Tables 7–9. Table 10 shows the MAPE performance comparison of plant 2. Basic and combined-forecasting models were tested for four months. The performance was ranked to identify the best performer from the sixteen forecasting models.

The average performance of basic forecasting models based on SVR with quadratic and Gaussian kernel function, ETR-bagging, and deep networks with LSTM layers were found better than the other basic forecasting models. It suggests that any of the four basic models should be preferred when a single forecasting model is affordable.

The average performance of the combined-forecasts was found better than the basic models. The benefit of the combined forecast is RMSE reduced by 0–2% (of the nominal power of plant) from basic forecasting models. The average performance of combined forecasting models based on SVR with Gaussian kernel function, deep networks with LSTM layers, and simple mean were found better than the other combined forecasting models. Bigger data sets are required for training the regression-based combined-forecasts than to prepare the single forecast models. When the data set is not big enough for regression based combined-forecast, mean based combined-forecast should be preferred.

The performance of the forecasting models is also illustrated in Fig. 6. The six days mix of sunny, rainy, and cloudy days of each plant were considered. Combined-forecast using SVR with Gaussian kernel was compared with recorded solar power and persistence forecast. The forecast for day-ahead solar radiation provided by NWP was also plotted to show their uncertainty. The performance graphs show that combined-forecast models performed better than the persistence models. The solar radiation forecast has high uncertainty during cloudy and rainy weather, due to which the error of day-ahead solar power forecast is also high. The combined-forecast models also performed well in uncertain weather such as cloudy and rainy weather.

5. Conclusions

The forecasting models based on a single algorithm such as persistence, LRI, FFNN, SVR, GPR, ETR, and deep networks were implemented. SVR with quadratic and Gaussian kernel function, ETR-bagging, and deep networks with LSTM layers were found good performers among all the basic forecasting models. The combined-forecast based on mean, median, linear regression, SVR, GPR, and deep networks were applied to all the three data sets of the solar plants. SVR with Gaussian kernel, deep networks with LSTM layers, and simple mean were found excellent in preparing the combined forecasting models.

The one model for all hours performed better than the separate models for each hour of the day. Therefore, the number of forecasting models required for day-ahead forecasting is only one. One-for-all-hours model reduces the computational burden of preparing many hourly models. The frequency of retraining of the forecasting models such as monthly, weekly, and daily was also analysed. A retraining frequency of one month was found to perform satisfactorily, which can reduce the computational burden of retraining the forecasting models at a higher frequency.

In this study, day-ahead solar power forecasting models were prepared, which is very important for large scale grid integration of solar PV. The proposed forecasting algorithms can also be applied to other forecasting time horizons by making some changes in the input vector. In future work, applications of solar power forecasting to power system operations and planning will be studied.

Credit author statement

Chaman Lal Dewangan, conceived of the presented idea, developed the theory and performed the computations, Writing-original and writing-review. S.N. Singh; conceived of the presented idea. S. Chakrabarti; developed the theory and performed the computations,Writing-original and writing-review. All authors discussed the results and contributed to the final manuscript.

Acknowledgment

The financial support for this research work is granted by Central Power Research Institute(CPRI), Bangalore, India (Project CPRI/EE/2016101: Development of Control Strategies for Grid-Connected PV System utilizing MPPT and Reactive Power Capability). The authors would also like to thank Dr Vipul Arora for his constructive discussion on machine learning algorithms.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.energy.2020.117743.

References

[1] Ssekulima EB, Anwar MB, Al Hinai A, El Moursi MS. Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: a review. IET Renew Power Gener 2016;10(7):885–989.
[2] Wang Y, Wu L. Improving economic values of day-ahead load forecasts to real-time power system operations. IET Gener Transm Distrib 2017;11(17):4238–47.
[3] Aziz Asma, Oo. Aman than and Stojcevski, Alex: 'Frequency regulation capabilities in wind power plant'. Sustainable Energy Technologies and Assessments 2018;26:47–76.
[4] Gupta Akshita, Kumar Arun, Khatod Dheeraj Kumar. Optimized scheduling of hydropower with increase in solar and wind installations. Energy 2019;183:716–32.
[5] Graabak Ingeborg, Korpås Magnus, Jaehnert Stefan, Belsnes M. Balancing future variable wind and solar power production in Central-West Europe with Norwegian hydropower. Energy 2019;168:870–82.
[6] Lei Ming, Zhang Jin, Dong Xiaodai, Jane J Ye. Modeling the bids of wind power producers in the day-ahead market with stochastic market clearing. Sustainable Energy Technologies and Assessments 2016;16:151–61.
[7] Wang Z, Negash A, Kirschen DS. Optimal scheduling of energy storage under forecast uncertainties. IET Gener Transm Distrib 2017;11(17):4220–6.
[8] Afrasiabi Mousa, Mohammadi Mohammad, Rastegar Mohammad, Kargarian Amin. Multi-agent microgrid energy management based on deep learning forecaster. Energy 2019;186:115873.
[9] Zhu Jian-hong, Pan Wen-xia, Li Xiaoqiang. Energy storage scheduling design on friendly grid wind power. Sustainable Energy Technologies and Assessments 2018;25:111–8.
[10] Dobschinski J, Bessa R, Du P, et al. Uncertainty forecasting in a nutshell: prediction models designed to prevent significant errors. IEEE Power Energy Mag 2017;15(16):40–9.
[11] Dewangan CL, Singh S, Chakrabarti S. Solar irradiance forecasting using wavelet neural network. In: Asia-pacific power and energy engineering conference (APPEEC). Bangalore, India: IEEE PES; 2017. p. 1–6. November 2017.
[12] Alonso-Montesinos J, Monterreal R, Fernández-Reche J, Ballestrín J, Carra E, Polo J, Barbero J, Batlles FJ, López G, Enrique R, others. Intra-hour energy potential forecasting in a central solar power plant receiver combining Meteosat images and atmospheric extinction. Energy 2019;188:116034.
[13] Tuohy A, Zack J, Haupt SE, et al. Solar forecasting: methods, challenges, and performance. IEEE Power Energy Mag 2015;13(6):50–9.
[14] Tahmasebifar R, Sheikh-El-Eslami MK, Kheirollahi R. 'Point and interval

forecasting of real-time and day-ahead electricity prices by a novel hybrid approach', *IET Gener*. Transm Distrib 2017;11(9):2173–83.
[15] Huang J, Perry M. A semi-empirical approach using gradient boosting and k-nearest neighbors regression for GEFCom2014 probabilistic solar power forecasting. Int J Forecast 2016;32(3):1081–6.
[16] Kim J-G, Kim D-H, Yoo W-S, Lee J-Y, Kim YB. Daily prediction of solar power generation based on weather forecast information in Korea. IET Renew Power Gener 2017;11(10):1268–73.
[17] Abedinia O, Raisz D, Amjady N. Effective prediction model for Hungarian small-scale solar power output. IET Renew Power Gener 2017;11(13):1648–58.
[18] de Freitas Viscondi Gabriel, Alves-Souza, Solange N. A Systematic Literature Review on big data for solar photovoltaic electricity generation forecasting. Sustainable Energy Technologies and Assessments 2019;31:54–63.
[19] Qing Xiangyun, Niu. Yugang: 'Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM'. Energy 2018;148:461–8.
[20] Mukhoty, Pratim Bhaskar, Maurya Vikas, Shukla Sandeep Kumar. Sequence to sequence deep learning models for solar irradiation forecasting. Milan, Italy: 2019 IEEE Milan PowerTech; June 2019. p. 1–6.
[21] Srivastava Shikhar, Lessmann Stefan. A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. Sol Energy 2018;162(3):232–47.
[22] Sheng H, Xiao J, Cheng Y, Ni Q, Wang S. Short-term solar power forecasting based on weighted Gaussian process regression. IEEE Trans Ind Electron 2018;65(1):300–8.
[23] Zeng J, Qiao W. Short-term solar power prediction using a support vector machine. Renew Energy 2013;52:118–27.
[24] Moghaddam AA, Seifi A. Study of forecasting renewable energies in smart grids using linear predictive filters and neural networks. IET Renew Power Gener 2011;5(6):470–80.
[25] Juban R, Ohlsson H, Maasoumy M, Poirier L, Kolter JZ. A multiple quantile regression approach to the wind, solar, and price tracks of GEFCom2014. Int J Forecast 2016;32(3):1094–102.
[26] Elliott G, Timmermann A. Economic forecasting. Princeton University Press; 2016.
[27] Liu J. 'Combining sister load forecasts', *Master thesis*. The University of North Carolina at Charlotte; 2012.
[28] Abuella Mohamed, Chowdhury. Badrul: 'improving combined solar power forecasts using estimated ramp rates: data-driven post-processing approach'. IET Renew Power Gener 2018;12(10):1127–35.
[29] Hong T, Pinson P, Fan S, Zareipour H, Troccoli A, Hyndman RJ. Probabilistic energy forecasting: global energy forecasting competition 2014 and beyond. Int J Forecast 2016;32(3):896–913.
[30] Marcano-Cedeno A, Quintanilla-Dominguez J, Cortina-Januchs MG, Andina D. Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network. In: IECON 2010-36th annual conference on IEEE industrial electronics society. Glendale, AZ: U.S.A.; November 2010. p. 2845–50.
[31] Zarbakhsh P, Demirel H. Fuzzy SVM for 3D facial expression classification using sequential forward feature selection. In: Computational intelligence and communication networks (CICN), 2017 9th international conference on. Cyprus: Girne; 2017. p. 131–4.
[32] Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003;3(Mar):1157–82.
[33] May R, Dandy G, Maier H. Review of input variable selection methods for artificial neural networks. In: Artificial neural networks-methodological advances and biomedical applications. InTech; 2011.
[34] Stalph P. Analysis and design of machine learning techniques: evolutionary solutions for regression, prediction, and control problems. Springer Science & Business Media; 2014. p. 72–89.
[35] Demuth HB, Beale MH, De Jess O, Hagan MT. Neural network design. Martin Hagan; 2014.
[36] Matalab 2018a Neural network tool box user's guide. accessed June 2018, https://www.mathworks.com/help/pdfdoc/nnet/nnetug.pdf.
[37] Lewis Nigel Da Costa. Deep time series forecasting with Python: an intuitive introduction to deep learning for applied time series modeling. CreateSpace Independent Publishing Platform; 2016.
[38] Long short-term memory (LSTM). accessed March 2020, https://en.wikipedia.org/wiki/Long_short-term_memory.
[39] Goodfellow Ian, Bengio Yoshua, Courville Aaron. Deep learning. MIT press; 2016.
[40] Smola AJ, Scholkopf B. A tutorial on support vector regression. Stat Comput 2004;14(3):199–222. Springer.
[41] Matalab 2018a Statistics and machine learning tool box user's guide. accessed June 2018, https://in.mathworks.com/help/pdfdoc/stats/stats.pdf.
[42] Sastry P. An introduction to support vector machines. Computing and information sciences: recent trends. New Delhi: Narosa Publishing House; 2003.
[43] Vapnik V. The nature of statistical learning theory. Springer science & business media; 2013.
[44] Christopher MB. Pattern recognition and machine learning. Springer-Verlag New York; 2016.
[45] Rasmussen CE, Williams CK. Gaussian process for machine learning. MIT press; 2006.
[46] James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. Springer; 2013.
[47] Breiman L. Classification and regression trees. Routledge; 2017.
[48] MATLAB. accessed June 2018, https://in.mathworks.com/products/matlab; 2017. version 9.2.0-564.
[49] Gueymard Christian A. 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 2012;86(8):2145–69.