



Novel stochastic methods to predict short-term solar radiation and photovoltaic power[★]

Jin Dong ^{a,*}, Mohammed M. Olama ^b, Teja Kuruganti ^b, Alexander M. Melin ^c,
Seddik M. Djouadi ^d, Yichen Zhang ^e, Yaosuo Xue ^c

^a Energy and Transportation Science Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831, USA

^b Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831, USA

^c Electrical and Electronics Systems Research Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831, USA

^d The Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, 37996, USA

^e Energy Systems Division, Argonne National Laboratory, Lemont, IL, 60439, USA

ARTICLE INFO

Article history:

Received 1 February 2019

Received in revised form

8 April 2019

Accepted 17 May 2019

Available online 23 May 2019

Keywords:

Renewable energy

Solar forecasting

Photovoltaics

Solar variability

Stochastic forecasting

Basis functions

ABSTRACT

Solar forecasting has evolved towards becoming a key component of economical realization of high penetration levels of photovoltaic (PV) systems. This paper presents two novel stochastic forecasting models for solar PV by utilizing historical measurement data to outline a short-term high-resolution probabilistic behavior of solar. First, an uncertain basis functions method is used to forecast both solar radiation and PV power. Three possible distributions are considered for the uncertain basis functions - Gaussian, Laplace, and Uniform distributions. Second, stochastic state-space models are applied to characterize the behaviors of solar radiation and PV power output. A filter-based expectation-maximization and Kalman filtering mechanism is employed to recursively estimate the system parameters and state variables. This enables the system to accurately forecast small as well as large fluctuations of the solar signals. The introduced forecasting models are suitable for real-time tertiary dispatch controllers and optimal power controllers. The PV forecasting models are tested using solar radiation and PV power measurement data collected from a 13.5 kW PV panel installed on the rooftop of our laboratory. The results are compared with standard time series forecasting mechanisms and show a substantial improvement in the forecasting accuracy of the total energy produced.

© 2019 Published by Elsevier Ltd.

1. Introduction

Among different forms of clean energies, solar energy has attracted a lot of attention because it is sustainable and renewable, but the potential of using this form of renewable energy depends on its effective integration into power distribution systems without impacting voltage and frequency limits. However, the application of

solar photovoltaic (PV) to the electric grid also has drawbacks, in particular the variability of power output adds stress and uncertainty to system. A major part of solar PV production is characterized by a large degree of intermittency driven by the natural variability of climate factors such as air temperature, wind velocity, precipitation, and evaporation [1]. The variable nature of solar PV could hamper further deployment due to the increase in reserves needed on the electric grid to compensate for fluctuations in the power output. Therefore, solar forecasting has evolved towards becoming a key component of economical realization of high penetration levels of PV systems [2].

Due to the intermittency of the power output of some renewable sources, micro-grid systems have become one of the main forms for the increasing share of distributed renewable systems [3]. As the amount of solar generation increases in relation to bulk generation, real-time controllers for dispatch will be needed to minimize the impact of short-term fluctuations in solar output. The performance of these real-time controllers will be improved by

[★] This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan <http://energy.gov/downloads/doe-public-access-plan>.

* Corresponding author.

E-mail address: dongj@ornl.gov (J. Dong).

short-term solar output forecasting. Moreover, the recently opened Energy Imbalance Market (EIM) allows for generation and demand balancing across Balancing Authority Areas (BAA) on 15 min and 5 min timescales with California Independent System Operator (CAISO) oversight [4]. One of the main challenges is therefore to perform forecasts as accurately as possible to overcome potential variability and uncertainty of renewable energy generation [5,6].

This paper examines the stochastic modeling and short-term high-resolution forecasting of solar radiation and PV power output. We introduce two stochastic state-space models to characterize the behaviors of solar radiation and PV power output. The short-term high-resolution forecasting models are suitable for the development of optimal power controllers for PV sources. They can also be utilized for automatic generation control (AGC), better energy management system, unit commitment, power scheduling and dispatching [7,8].

1.1. Overview of solar forecast methods

Various algorithms have been introduced to forecast solar radiation and PV generation [9] reviewed the state-of-the-art methods for short term solar/PV forecasting. Solar radiation in spatially distributed locations can be forecasted using neighboring spatial location information. Sky camera and satellite data can help reveal spatio-temporal characteristics of the solar radiation. Solar forecasting with satellite imaging or sky cameras has become a hot topic recently for data acquisition and cloud movement [2,10–15]. Analytical models have been developed for arbitrary cloud configuration and can also predict diffuse irradiance based on the information about cloud coverage [16]. Later, similar numerical solution has been extended to account for cloud movement [17].

Besides using ground-based images, there also exist approaches that only rely on a sensor network. For instance Ref. [18], established intra-hour forecasts using a network of 80 rooftop PV systems spread over a large area. More sensor network based forecasting methods have been proposed in Refs. [19–21], which rely on information from a collection of neighboring sites. Very short-term sub-5-min forecasting has been conducted in Ref. [12] using emerging sky camera technology.

Similarly, multiple linear regression methods including support vector machine (SVM) [22] were used to study the solar power output characteristics, combined with weather data and solar radiation data. Moreover, hybrid forecasting techniques have been investigated recently, including k-means algorithm [23], Artificial Neural Networks (ANNs) model [5], k-nearest neighbors (kNN) [24], as well as SVM with Wavelet Transform (WT) algorithm [25,26], to increase the forecast accuracy. However, most of the aforementioned methods required large amount of data samples, and the fitting results were sensitive to pathological data [27]. Besides many assumptions, such as linear cloud edge, that have to be made, various types of error will be embedded in different phases of such methods, especially during the conversion from cloud condition to ground-level radiation [28].

The solar/PV forecasts have attracted substantial attention from stochastic modeling and forecasting studies. The works in Refs. [29–31] represent solar radiation using hidden Markov models (HMMs). The works in Refs. [32–34] represent solar radiation using autoregressive integrated moving average (ARIMA) model. Lastly, the work in Ref. [35] presents a stochastic solar radiation forecasting framework based on conditional random fields (CRFs). These works provide accurate models and forecasting that capture the hourly, daily, and seasonal solar PV trends. However, they lack accurate high-resolution modeling and forecasting of fast and large fluctuations of solar radiation/PV power that often occur as illustrated in Fig. 1 for a typical summer day.

Moreover, a method based on a combination of an ARMA model and a Kalman filter has been introduced for the scope of solar radiation and temperature forecasting. The ARMA model is built with a high order parameters (p, q) in order to retrieve the state-space parameters needed to run the Kalman filter. [36] has presented a work which extends previous studies [37,38], to solve the problem of bias removal from the Kalman forecasts.

The closest related prior work to the present paper is the state-space modeling framework recently published in Ref. [39], where the linear Kalman filter is used to perform forecasting for different time horizons, between one minute and one hour ahead. They considered two sets of inputs, one that consisted only of past PV measurements and another which also incorporated cloud cover and air temperature values from a nearby meteorological station. They used a Kalman filter, where parameter tuning was addressed via two methods: an AR model and an expectation maximization (EM) algorithm.

1.2. Motivation and research objective

Most of the aforementioned techniques require additional measurements or sensor networks other than purely local solar radiation/PV power generation as in our paper. This increases the deployment complexity in both hardware and software, not to mention some unrealistic assumptions, such as linear cloud edge and various types of error embedded in different phases of these methods [28]. The advantage of our algorithm is that it is 'data-driven in real-time and adaptive', i.e., directly uses the collected data from local interested site to generate solar forecasting. While the methods mentioned above are offline needing training sets and cross validation on the top of using more information. This motivates us to develop novel low data demanding algorithms to improve the efficiency and accuracy of short-term high-resolution solar radiation/PV power forecasting. A new data-driven stochastic framework for short-term forecasting of solar radiation and PV production is laid out.

In this paper, we focus on designing a robust stochastic forecasting model which can accommodate the factors affecting PV output for higher forecast accuracy. In order to avoid requiring exogenous influential features namely the cloud cover and the temperature, we introduce a unique stochastic framework to forecast the solar radiation and the PV production. For this scope, a methodology based on statistical inference techniques is built, specifically, the uncertain basis functions method and Kalman filtering based EM algorithm. Both of the two introduced algorithms work adaptively and recursively to track the latest dynamics of the solar radiation and PV generation.

We begin with the uncertain basis functions technique, which takes into account the uncertain physical properties of the system of solar and PV power. In contrast with traditional works, where basis functions (e.g., radial basis function [40], and wavelets [41]) are fixed, we assume that the basis functions are random, according to three different probability distributions, i.e., *Uniform*, *Gaussian* and *Laplace* distributions.

The state-space model is dedicated to describing the system with all the parameters needed to capture the statistical properties and system dynamics. Thereafter, as a second phase, we talk about the methodology to estimate the needed parameters which quantify the dynamics of system. A filter-based EM mechanism is employed to estimate the parameters and state variables in the state-space model. The mechanism results in a finite dimensional filter which only uses the first and second order statistics. As mentioned in Ref. [39], the advantage of using linear discrete-time Kalman filter is mainly related to its ability to do future forecasting in real-time online in a memory-less way.

With both the uncertain basis and state-space model approaches, we can build a model from which we derive (estimate) an a priori unknown hidden signal and re-utilize it for forecasting w.r.t. the measured data. This mechanism is recursive, which allows the solar radiation and PV power to be forecasted online from measurements. Preliminary results based on this idea were initially presented in Refs. [42,43]. However, this paper includes more derivation details and case studies, which provide a concrete solar forecasting framework. Moreover, both Laplace distribution uncertain basis functions and Prediction Error Minimization (PEM) algorithm are presented for the first time.

To demonstrate the effectiveness of the introduced approach, simulation-based experimental analysis is carried out based on a set of PV generation data collected from PV cells located on the rooftop of the Distributed Energy Communication & Control (DECC) laboratory at Oak Ridge National Laboratory (ORNL) for the year of 2015. Although preliminary study for solar forecasting using uncertain basis functions has been investigated in Ref. [42], we will compare it with the stochastic state-space method as well as the traditional time series techniques.

The modeling and forecasting strategy of this paper starts in Section 2 with the uncertain basis algorithm to forecast solar radiation and solar power generation. A stochastic state-space algorithm is then presented in Section 3, based on Kalman filter and EM. A revisit to traditional time series technique is cast in Section 4, which is well-suited to forecast the solar radiation/PV generation. In Section 5, we present the simulation study regarding the forecasting performance using the introduced algorithms, and provide a comparison with time series methods, which act as a benchmark. Conclusions are provided in Section 6.

2. Modeling and prediction with uncertain basis functions

In this section, we introduce the uncertain basis functions technique employed to model and forecast solar radiation/PV power. The modeling and forecasting of signals using uncertain basis functions were proposed recently in Ref. [44]. Since the traditional ARIMA method is unable to yield good performance under highly dynamic and uncertain disturbances, the uncertain basis function technique has been utilized in forecasting frequency fluctuations in power grid [45,46]. Considering the uncertain characteristic of

cloud cover movement and the ambient temperature change, it is promising to apply this technique for solar forecasting (preliminary result presented in Ref. [42]). It should be remarked that most of the derivations in Sections 2 and 3 follow the literature, but to the best of our knowledge we are the first to apply these models for solar forecasting.

2.1. Uncertain basis functions

Let $\{y_k\}$ denote sampled measurements. The model for the measurements $\{y_k\}$ assumes the following form [44]:

$$y_k = \sum_{i=1}^p A_i \phi_{i,k}, \quad k = 1, 2, \dots, N, \quad (1)$$

where $\{\phi_{1,k}, \dots, \phi_{p,k}\}$ are basis functions, $p < N$ (for avoiding linearly dependency), and A_i is the corresponding coefficient of each basis function.

After obtaining a batch of measurements $\{y_1, \dots, y_N\}$, the amplitude coefficients can be estimated using least squares (minimizing the estimation error below) if the basis functions are known:

$$J(\mathbf{A}) = \sum_{k=1}^N \left(y_k - \sum_{i=1}^p A_i \phi_{i,k} \right)^2, \quad (2)$$

where $\mathbf{A} = [A_1, \dots, A_p]^T$ is the amplitude vector to be computed.

However, this assumption for known basis functions is not realistic since they are not given in practice. To tackle this difficulty, one possible solution is to assume that each basis function depends on some unknown parameter vectors θ_i , where only some statistics of the distribution of θ_i are known. Then, the corresponding coefficients can be estimated by minimizing an expected cost function following the technique developed in Ref. [44].

In the following discussion, we will review the method developed in Ref. [44]. First, assume that θ_i are independent. The basis functions are further represented by $\phi_{i,k}(\theta_i)$. The expected cost function $\hat{J}(\mathbf{A})$ is given as:

$$\hat{J}(\mathbf{A}) = \mathbb{E}_{\theta} \left[\sum_{k=1}^N \left| y_k - \sum_{i=1}^p A_i \phi_{i,k}(\theta_i) \right|^2 \right], \quad (3)$$

where \mathbb{E}_{θ} is the expectation with respect to θ_i . The measured values are real, so in order to estimate the coefficients \mathbf{A} , let $\mathbf{y} = [y_1, \dots, y_N]^T$, $\theta_i = [\theta_{i,1}, \dots, \theta_{i,p}]^T$, and define an $N \times p$ matrix \mathbf{H} as below:

$$\mathbf{H}(\theta) = \begin{bmatrix} \phi_{1,1}(\theta_1) & \phi_{2,1}(\theta_2) & \cdots & \phi_{p,1}(\theta_p) \\ \phi_{1,2}(\theta_1) & \phi_{2,2}(\theta_2) & \cdots & \phi_{p,2}(\theta_p) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{1,N}(\theta_1) & \phi_{2,N}(\theta_2) & \cdots & \phi_{p,N}(\theta_p) \end{bmatrix}. \quad (4)$$

In order to calculate \mathbf{A} , we define

$$\mathbf{G} = \mathbb{E}_{\theta} [\mathbf{H}(\theta)]. \quad (5)$$

The matrix \mathbf{G} is the expectation of $\mathbf{H}(\theta)$, and

$$\begin{aligned} \mathbf{C}_{\mathbf{H}} &= \mathbb{E}_{\theta} \left[(\mathbf{H}(\theta) - \mathbb{E}_{\theta}[\mathbf{H}(\theta)])^T (\mathbf{H}(\theta) - \mathbb{E}_{\theta}[\mathbf{H}(\theta)]) \right] \\ &= \mathbb{E}_{\theta} [\mathbf{H}(\theta)^T \mathbf{H}(\theta)] - \mathbf{G}^T \mathbf{G}, \end{aligned} \quad (6)$$

where $\mathbf{C}_{\mathbf{H}}$ is the covariation matrix.

Then, \mathbf{A} can be computed from the following equation [44]:

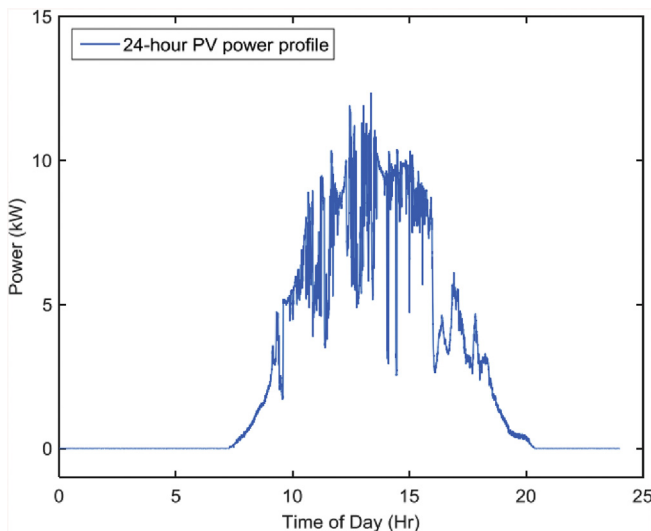


Fig. 1. The solar PV power profile for a typical summer day collected from a 13.5 kW PV panel that shows the fast and large fluctuations of PV power output.

$$\hat{\mathbf{A}} = (\mathbf{G}^T \mathbf{G} + \mathbf{C}_H)^{-1} \mathbf{G}^T \mathbf{y}. \quad (7)$$

The term \mathbf{C}_H compensates for the uncertainty in the basis functions [44], and due to its independence, it can be written as a $p \times p$ diagonal matrix as

$$\mathbf{C}_H = \begin{bmatrix} \sum_{k=1}^N \text{var}(\phi_{1,k}(\theta_1)) & 0 & \cdots & 0 \\ 0 & \sum_{k=1}^N \text{var}(\phi_{2,k}(\theta_2)) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sum_{k=1}^N \text{var}(\phi_{p,k}(\theta_p)) \end{bmatrix} \quad (8)$$

In our application, we wish to forecast a future solar PV power using the average basis function:

$$\hat{y}_k = \sum_{i=1}^p \hat{A}_i \mathbb{E}_\theta[\phi_{i,k}(\theta_i)], \quad (9)$$

where \hat{A}_i 's are estimated by (7). This means that we can forecast radiation values with p random basis functions.

In the next section, basis functions are assumed to be governed by three most popular distributions - *Gaussian*, *uniform*, and *Laplace* probability distributions; and we compute \mathbf{G} and \mathbf{C}_H for each case.

2.2. Applications with Gaussian, uniform, and Laplace probability distributions

In this section, we compute matrices \mathbf{G} and \mathbf{C}_H which will be used in the forecasting. Let s_i be a special case of θ_i , the basis functions are chosen in the form of

$$\phi_{i,k}(\theta_i) = (k+1)^{s_i-1/N}, \quad (10)$$

which is one of the most generally used bases [47] (at the end of Page 2). As claimed on [47], this basis function is chosen according to the “universal approximation” argument, when specific prior knowledge is not available. It characterizes both time variations and different bases.

Assume that s_i is a random unknown variable that follows some distributions. If s_i 's are fixed, then the bases become traditional basis functions. This choice introduces time dependence k as well as randomness. Detailed derivation of \mathbf{C}_H for the three different distributions can be found in Appendix A. Based on these results, we can explicitly calculate the estimation of \mathbf{A}_i 's.

It should be mentioned that in our introduced algorithm, the mean value of s_i for each distribution is chosen to be m_i/N . Therefore, it's clear that the traditional fixed basis functions method is a special case, where s_i is set as m_i/N in (10). The uncertain method can thus be explained as modeling solar PV power by random bases with assumed statistics.

3. Stochastic state-space model

The state-space model has been widely employed in control systems and signal processing since it can be used online and updated after receiving the most recent observation. It consists of a state (or system) equation (11) and a measurement (or output) equation (12). y_k denotes the solar radiation/PV power measurement at time k . The state-space equation in discrete time can be written in the following form:

$$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k + \mathbf{B}_k \mathbf{w}_k \quad (11)$$

$$\mathbf{y}_k = \mathbf{C}_k \mathbf{x}_k + \mathbf{D}_k \mathbf{v}_k \quad (12)$$

where $\mathbf{x}_k \in \mathbb{R}^n$ (\mathbb{R}^n denotes the space of real vectors of dimension n) is the state that characterizes the solar radiation/PV power; it is described by the previous state \mathbf{x}_k and the noise term $\mathbf{w}_k \in \mathbb{R}^{m \times 1}$ introduced at each k ; $\mathbf{A}_k \in \mathbb{R}^{n \times n}$ and $\mathbf{B}_k \in \mathbb{R}^{n \times m}$ are the corresponding coefficients. It should be mentioned that the order n can be randomly chosen by the user following a simple principle, which is to increase the order for more complicated and highly dynamic dataset. While, we may suffer computation challenge with higher order. Therefore, we need to pick and try to find a balance between accuracy and computational burden.

A 4th order stochastic model means we need 4 state-variables, i.e., the dimension of the state-space is 4, to characterize the solar radiation/PV power. Basically, the dimension of state matrix \mathbf{A}_k (in Eq. (11)) is 4 by 4 since \mathbf{x}_k is 4 by 1. Note we have investigated lower order state-space models, but their performance were not satisfactory. The solar radiation/PV power measurement sample y_k is described by the previous state \mathbf{x}_k and the measurement noise process $\mathbf{v}_k \in \mathbb{R}^{m \times 1}$, while $\mathbf{C}_k \in \mathbb{R}^{1 \times n}$ and $\mathbf{D}_k \in \mathbb{R}^{1 \times m}$ are the corresponding coefficients.

The time varying property of the parameters renders the state-space model able to adapt dynamically to a variety of radiation/PV power values. The noise terms \mathbf{w}_k and \mathbf{v}_k can capture small perturbations or uncertainties introduced at each time k which improve the flexibility of the model. In this paper, the noise terms are assumed to be independent zero-mean and unit-variance Gaussian processes.

Due to these properties, the state-space model is used to track and forecast the behaviors of solar radiation/PV power. The unknown system parameters $\beta_k = \{\mathbf{A}_k, \mathbf{B}_k, \mathbf{C}_k, \mathbf{D}_k\}$ and states \mathbf{x}_k can be estimated through a finite set of received signal measurement data $\mathbf{Y}_N = \{y_1, y_2, \dots, y_N\}$. Future radiation/PV power values can be forecasted recursively based on these estimated parameters. The parameters are identified using a filter-based EM algorithm [48,49] and the system states are estimated using the Kalman filter [50]. It has been previously proven that the Kalman filter could optimally estimate the system state in the mean square sense and that the filter-based EM algorithm yields a maximum likelihood (ML) parameter estimate [51].

3.1. State estimation: The Kalman filter

The Kalman filter estimates the system states \mathbf{x}_k for given system parameter β_k and measurements \mathbf{Y}_k . It should be mentioned that, hereafter, all the variables denoted with “ $\hat{\cdot}$ ” refer to estimated values. It is described by the following equations [50]:

$$\begin{aligned} \hat{\mathbf{x}}_{k/k} &= \mathbf{A}_{k-1} \hat{\mathbf{x}}_{k-1/k-1} + \mathbf{P}_{k/k} \mathbf{C}_{k-1}^T \mathbf{D}_{k-1}^{-2} (\mathbf{y}_k - \mathbf{C}_{k-1} \mathbf{A}_{k-1} \hat{\mathbf{x}}_{k-1/k-1}) \\ \hat{\mathbf{x}}_{k/k-1} &= \mathbf{A}_{k-1} \hat{\mathbf{x}}_{k-1/k-1} \end{aligned} \quad (13)$$

where $k = 1, 2, \dots, N$, and $\mathbf{P}_{k/k}$ is given by

$$\begin{aligned} \bar{\mathbf{P}}_{k/k}^{-1} &= \mathbf{P}_{k-1/k-1}^{-1} + \mathbf{A}_{k-1}^T \mathbf{B}_{k-1}^{-2} \mathbf{A}_{k-1} \\ \mathbf{P}_{k/k}^{-1} &= \mathbf{C}_{k-1}^T \mathbf{D}_{k-1}^{-2} \mathbf{C}_{k-1} + \mathbf{B}_{k-1}^{-2} - \mathbf{B}_{k-1}^{-2} \bar{\mathbf{P}}_{k/k} \mathbf{A}_{k-1}^T \mathbf{B}_{k-1}^{-2} \\ \mathbf{P}_{k/k-1} &= \mathbf{A}_{k-1} \mathbf{P}_{k-1/k-1} \mathbf{A}_{k-1}^T + \mathbf{B}_{k-1}^2 \end{aligned} \quad (14)$$

where $\mathbf{B}_{k-1}^2 = \mathbf{B}_{k-1} \mathbf{B}_{k-1}^T$ and $\mathbf{D}_{k-1}^2 = \mathbf{D}_{k-1} \mathbf{D}_{k-1}^T$. The system parameters $\beta_k = \{\mathbf{A}_k, \mathbf{B}_k, \mathbf{C}_k, \mathbf{D}_k\}$ are estimated using the EM algorithm

which is introduced next.

3.2. Parameter estimation: the EM algorithm

The filter-based EM algorithm uses a bank of filters to yield a ML parameter estimate of the Gaussian state space model. The EM algorithm is an iterative numerical algorithm for computing the ML estimate (MLE). Each iteration consists of two steps: the expectation and the maximization steps. The filtered expectation step only uses filters for the first and second order statistics. The memory costs are modest and the filters are decoupled and hence easy to implement in parallel on a multi-processor system. The EM algorithm is described by Ref. [49].

$$\begin{aligned}\hat{A}_k &= E\left(\sum_{i=1}^k x_i x_{i-1}^T \middle| Y_k\right) \times \left[E\left(\sum_{i=1}^k x_i x_i^T \middle| Y_k\right)\right]^{-1} \\ \hat{B}_k^2 &= \frac{1}{k} E\left(\sum_{i=1}^k \left((x_i - A_i x_{i-1})(x_i - A_i x_{i-1})^T\right) \middle| Y_k\right) \\ &= \frac{1}{k} E\left(\sum_{i=1}^k \left(x_i x_i^T - A_i (x_i x_{i-1}^T)^T - (x_i x_{i-1}^T) A_i^T + A_i (x_i x_{i-1}^T) A_i^T\right) \middle| Y_k\right) \\ \hat{C}_k &= E\left(\sum_{i=1}^k y_i x_i^T \middle| Y_k\right) \times \left[E\left(\sum_{i=1}^k x_i x_i^T \middle| Y_k\right)\right]^{-1} \\ \hat{D}_k^2 &= \frac{1}{k} E\left(\sum_{i=1}^k \left((y_i - C_i x_i)(y_i - C_i x_i)^T\right) \middle| Y_k\right) \\ &= \frac{1}{k} E\left(\sum_{i=1}^k \left(y_i y_i^T - C_i (y_i x_i^T)^T - (y_i x_i^T) C_i^T + C_i (y_i x_i^T) C_i^T\right) \middle| Y_k\right) \quad (15)\end{aligned}$$

where $E(x)$ denotes the expectation operator to variable x , $E(\cdot | Y_k)$ is the conditional expectation given the measurable set Y_k .

Thus the filter-based EM algorithm for computing the MLE β_k can be summarized as follows: Choose an initial parameter estimate β_0 , then compute $\hat{\beta}_k$ at each iteration according to (15).

The detailed computation for system parameters β_k is based on four conditional expectations [48], $L_k^i (i = 1, 2, 3, 4)$, and derived in Appendix B. Using the filters for $L_k^i (i = 1, 2, 3, 4)$ and the Kalman filter described earlier, the system parameters $\beta_k = \{A_k, B_k, C_k, D_k\}$ can be estimated through the EM algorithm described in (15).

After estimating all the parameters in the state space model, we can achieve one-step forecast of the solar radiation/PV power by Ref. [51]:

$$\hat{y}_{k+1} = \hat{C}_k (\hat{A}_k \hat{x}_{k/k} + \hat{K}_k (y_k - \hat{C}_k \hat{x}_{k/k})) \quad (16)$$

where \hat{y}_{k+1} denotes the forecasted solar radiation/PV power at $k+1$ and \hat{K}_k is the Kalman gain given by

$$\hat{K}_k = (\hat{A}_{k-1} P_{k/k} \hat{C}_{k-1}^T) (\hat{C}_{k-1} P_{k/k} \hat{C}_{k-1}^T + \hat{D}_{k-1}^2)^{-1} \quad (17)$$

This algorithm yields parameter estimates with non-decreasing values of the likelihood function, and converges under mild assumptions [52].

It should be noted that there is no exact requirement for the number of measurement data for the EM method since it is recursive. However, in order to fully track the model, the system parameter β_k may grow rapidly, which may cause an unsatisfying forecasting performance.

The forecast probability distribution is Gaussian with mean and

covariance given by those of the predicted state-variables. Note, since the Kalman filter is optimal for any Gaussian noise, the stochastic state space can also quantify explicitly the statistics of forecast probability distribution (mean and covariance) for the Gaussian noise scenario. For non-Gaussian noise, it is more complicated and challenging, and is out of the scope of this paper.

3.3. PEM algorithm

Another important parameter estimation algorithm is the Prediction Error Minimization (PEM) method [53], where the objective is to minimize forecasting errors. PEM updates the measurement set every time when a new measurement comes in, then, the model is updated to keep up with time variations.

Given a finite number of measurements, PEM algorithm employs N stored measurements for the next forecasting where the integer N can be chosen as the smallest number for which the algorithms have solutions. PEM estimates the parameters by minimizing a least square cost function:

$$\min J_N = \min \frac{1}{N} \sum_{k=0}^{N-1} \|y_k - \hat{y}_k\|_2^2 \quad (18)$$

where \hat{y}_k is one-step-ahead forecast at time k using the measurement data up to time $k-1$. More details about the PEM algorithm can be retrieved from Ref. [53].

Next, we will briefly introduce the time series method (ARIMA), which works as a benchmark in our study. Finally, different techniques of forecasting based on the introduced theories of uncertain basis functions, EM, and PEM algorithms will be executed and a comparison of the forecasting accuracy is held between the three introduced methods and traditional time series forecasting.

4. Comparable forecasting methods - time series forecasting

Time series forecasting has been applied for different applications with higher degree of success since the last decade. A time series is a collection of time ordered observations x_t , each one being recorded at a specific time t (period) [54]. Time series can appear in a wide set of domains such as finance, production or control, just to name a few. As a first approximation, a time series model (\hat{x}_t) assumes that past patterns will occur in the future. In fact, a time series model could be used to provide synthetic time series statistically similar to the original one. The modeling of time series begins with the selection of a suitable mathematical model (or class of models) for the data. Then, it is possible to forecast future values of measurements [55]. Moreover, time series methods have been applied to forecast solar PV power in Ref. [56]. In statistical terms, variables are used as model inputs having correlation with outputs and predictors. Several studies have been published for time series modeling [32,57,58].

4.1. AR model

An Auto-Regression (AR) model is a kind of random process which is often used to model and forecast various types of natural phenomena [59]. A typical AR process can be written as:

$$X(k) = \sum_{i=1}^p \alpha_i X(k-i) + \varepsilon(k), \quad (19)$$

where $X(k)$ is a time series, $\alpha_1, \dots, \alpha_p$ are parameters of the model corresponding to $X(k)$, and $\varepsilon(k)$ represents the uncertain term or fluctuations modeled as white noise at time k . p is the order of the

AR model representing how many past measurements the current one depends on. If we use $X(k)$ to represent the solar PV power metrics, then the future metrics can be forecasted through (19). Before forecasting, the parameters in (19) need to be estimated based on past measurements.

Then, one-step forecasting of the metrics can be computed by:

$$\hat{X}(k+1) = \sum_{i=1}^p \alpha_i X(k+1-i). \quad (20)$$

4.2. ARMA model

The ARMA model is a combination of two basic models which are autoregressive (AR) and moving average (MA) model. ARMA is considered as a good forecasting model to track the future values of time series with stable input variables. The ARMA model is referred as (p, q) , where p represents the order of the AR model and q represents the order of the MA model. A typical ARMA (1,1) model can be expressed as:

$$z_t = \alpha + \phi z_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t \quad (21)$$

The (1,1) in the equation stand for the auto-regressive (z_t) and moving average (ε_t) lag orders, respectively. The intuitive understanding of Eq. (21) is pretty straightforward. The current value of the time series z_t will depend on the past value of the series z_{t-1} and will correct itself to the error made in the last time period ε_{t-1} .

4.3. ARIMA model

The ARIMA time series model is an extension of the ARMA model, that consists of the AR model, integrating part I , and MA model. ARIMA models are used when the time series data show some evidence of non-stationarity, where an initial differencing step (corresponding to the “integrated” part of the model) can be applied one or more times to eliminate the non-stationarity [60]. The authors of [61,62] found that ARIMA models produced lower solar radiation and solar power errors compared to other time-series short-range forecasting.

The standard statistical methodology to construct an ARIMA model includes the following steps listed in Table 1 the detailed explanation for each step can be found in Ref. [60].

Since ARIMA is an extension to the other two time series techniques (AR and ARMA), it usually generates better results than its two predecessors, we will only list the forecasting results using the ARIMA method.

5. Numerical studies and discussion

To illustrate the effectiveness of the introduced solar PV power forecasting techniques, we evaluate their performances for different scenarios in this section. According to a set of solar radiation/power data from our local PV panel, both the state-space model and uncertain basis functions model are applied. For the state-space model, any initial values can be chosen since all these values will converge. For the basis-functions model, we chose to get a tradeoff between the computations while keeping the performance satisfactory. In addition, the number of measurement data points of the basis functions was chosen as the smallest number that can give us a solution. We tuned the parameters for three different distributions to have good prediction performance. To verify that the performance of uncertain basis functions is not sensitive to specific values chosen, three different parameter sets

are chosen as in Table 3 for uniform uncertain basis functions. Notice that, the mean values of all distributions should be smaller than 1 due to the exponential form of the basis functions. We consider 5-min prediction of the solar radiation/power and compute the forecasting error.

5.1. System description and experimental data

We have tested the introduced forecasting mechanisms on 12-month measurement of solar radiation and PV power output data recorded at one sample per second and collected from a 13.5 kW PV panel located on the rooftop of the Distributed Energy Communication & Control (DECC) laboratory (SMA Sunnyboy inverters SB7000-12 and the panels are Hanwha HSL 72 PPA 0280K) at Oak Ridge National Laboratory (ORNL) in Oak Ridge, Tennessee. We selected a “partly cloudy” day in this section. A partly cloudy day represents the worst-case scenario with respect to solar forecasting as it provides frequent and large variations in solar radiation and PV power. It is much easier to forecast solar in “mostly sunny” and “mostly cloudy” days as there are just few variations in solar in such days. It is shown in Ref. [63] that there is minimal energy delivered by PV systems in variations faster than one hertz. So, the collected solar radiation (in W/m^2) and power (in W) data was recorded at 1 Hz sampling rate.

It should be mentioned that we follow a common practice by removing night hours, i.e., the actual measurement only records daytime data which initiates by the first non-zero solar radiation (PV power) measurement. Night values are not included, nor are they used in the calculations for forecasting accuracy.

We will evaluate the forecasting performance under three different time scales - 1 min, 5 min and 50 min, respectively. All the following experiments are conducted based on the same segment of historical solar radiation/PV power measurements.

5.2. Error metrics

In order to evaluate the overall forecasting accuracy of the developed framework, the normalized Root Mean Square Error (nRMSE) and Mean Absolute Percentage Error (MAPE) are utilized as a criterion to measure the differences between values forecasted by a model and the values actually observed. The nRMSE and MAPE are defined by Ref. [64].

$$nRMSE = \frac{1}{y_{avg}} \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}, \quad (22)$$

$$MAPE = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (23)$$

where \hat{y}_i is the forecasted value of solar radiation (PV power); y_i is the actual measurement; y_{avg} is the average actual value of solar radiation (PV power) over the day; N represents the sample size, which is 1420.

5.3. Forecast skill

It is worth mentioning that it may be interesting to include a new metric: the forecast skill score (SS) parameter, which is defined by Ref. [65],

$$SS = 1 - \frac{nRMSE_{forecast}}{nRMSE_{persistence}} \quad (24)$$

Table 1

Methodology to construct an ARIMA model.

Step 0	A class of models is formulated assuming certain hypotheses.
Step 1	A model is identified for the series considered.
Step 2	The parameters of the model are estimated.
Step 3	If the hypotheses of the model are validated, the procedure continues in Step 4; otherwise, the procedure continues in Step 1 to refine the model.
Step 4	The model is used to forecast.

Table 2

Comparison of prediction accuracy, where UBU, UBG and UBL represent Uncertain Basis (Uniform), Uncertain Basis (Gaussian) and Uncertain Basis (Laplace), respectively. nRMSE and nMAPE with subindex R (P) represent forecasting error for solar radiation (PV power).

Algorithms	nRMSE _R (%)	nRMSE _P (%)	MAPE _R (%)	MAPE _P (%)	Time horizon (min)
UBU	7.43	8.11	5.72	5.96	1
UBG	7.60	8.29	6.08	6.21	1
UBL	7.92	8.43	5.86	6.04	1
EM-KF	7.79	8.20	5.52	5.81	1
PEM	12.40	12.14	9.38	10.61	1
ARIMA	8.59	8.82	6.06	6.27	1
UBU	18.13	18.58	16.85	17.26	5
UBG	17.82	18.34	16.32	16.86	5
UBL	18.21	18.62	16.94	17.28	5
EM-KF	19.64	20.03	18.75	19.18	5
PEM	19.35	19.94	17.04	17.57	5
ARIMA	23.53	24.61	21.38	22.19	5
UBU	26.13	26.82	25.30	25.68	50
UBG	27.72	27.81	25.73	25.89	50
UBL	26.04	27.13	25.32	25.80	50
EM-KF	28.43	30.74	26.06	26.74	50
PEM	25.56	27.94	25.58	25.62	50
ARIMA	29.18	32.26	26.52	29.60	50

where nRMSE_{persistence} is the nRMSE for a clear-sky persistence forecast which is defined as follows:

$$y(t + FH) = \frac{y^{clr}(t + FH)}{y^{clr}(t)} \times y(t). \quad (25)$$

This method takes into account the diurnal cycle of radiation, but it does require a clear-sky expectation, $y^{clr}(t)$, be known or modeled appropriately [13]. One way to calculate it is to use thekasten clear sky model [66], which accounts for atmospheric turbidity and solar elevation angle. However, it does require three exogenous inputs including air mass, Linke Turbidity, and elevation. Since we only have solar radiation and PV power measurement for our locally installed PV panel, we are not able to estimate $y^{clr}(t)$ due to missing the three exogenous inputs. The skill score gives the fractional improvement in the mean square error of the proposed forecasting model over the reference model (here the “persistence” model): a skill score of 1 indicates a perfect forecast, a score of 0 indicates no improvement against the reference, and a negative skill means that the forecast model tested performs worse than reference [39].

Table 3

Comparison of 5-min forecasting errors corresponding to three different parameter sets of Uncertain Basis (Uniform).

Algorithms	Parameter set I	Parameter set II	Parameter set III
μ_1	0.12	0.83	0.48
μ_2	0.69	0.11	0.36
μ_3	0.63	0.48	0.12
nRMSE _R (%)	18.74	18.83	18.60
nRMSE _P (%)	18.42	18.58	18.13

5.4. Experimental results

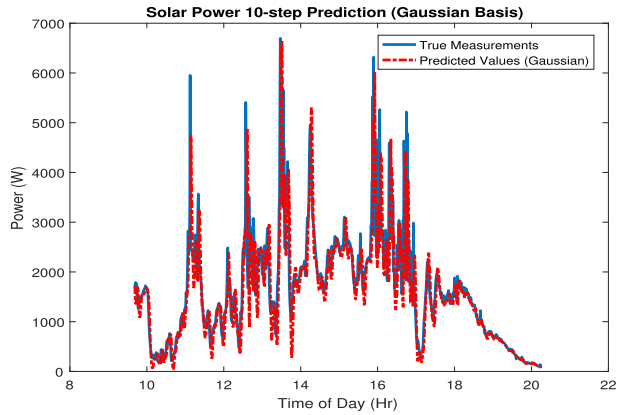
First, we will do a one-min ahead forecasting test when online measurements are available. In the second scenario, we show a relatively longer scenario, where the forecasting horizon has been extended to 5 min. Finally, the introduced forecasting algorithms have been tested with 50 min solar forecasting. This explores the possibility to extend each forecasting technique for longer forecasting cases.

Due to space limit, we only show figures for the 5 min forecasting horizon. In addition, 4th order models are used in the introduced forecasting techniques for comparison. For the other time horizons, similar performance were achieved (with numerical accuracy listed in Table 2).

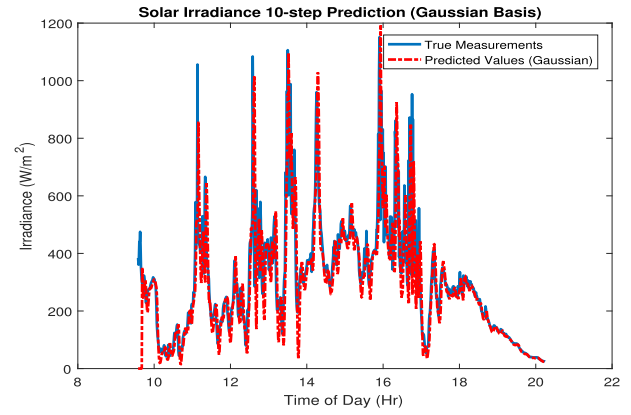
Fig. 2a – 2c depict solar power forecasting results from basis functions using *Gaussian*, *Laplace* and *Uniform* distributions, respectively. Notice that the forecasting results (red dashed lines) follow very closely with the true measurements (blue solid lines) in any of the figures of Fig. 2. Hence, it is claimed that satisfying forecasting results have been achieved utilizing any of the three different distributions. This basically validates that the introduced uncertain basis functions method is flexible to specific distribution chosen, though slightly different performance is achieved. However, *Gaussian* distribution does generate the best performance for this case study as shown in Fig. 2a, or from the numerical results shown in Table 2.

Similarly, forecasting results for solar radiation using uncertain basis functions method are shown in Fig. 3. The overall performance follows very closely to the basis functions for the solar power. Moreover, the numerical results in Table 3 verify that the performance of uncertain basis functions is not sensitive to specific values chosen.

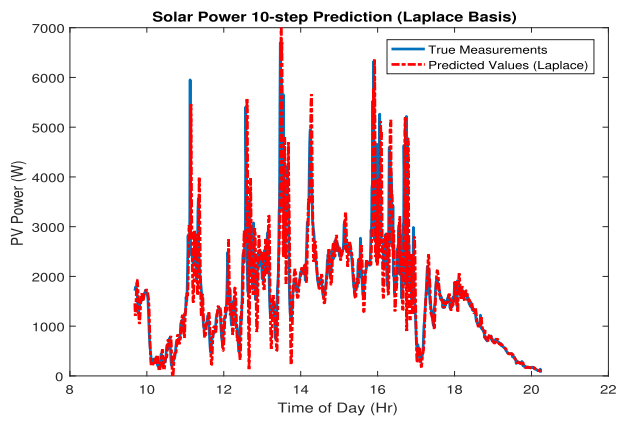
Fig. 4a and b demonstrate the performance of the introduced forecasting algorithm acting on real solar radiation and PV power



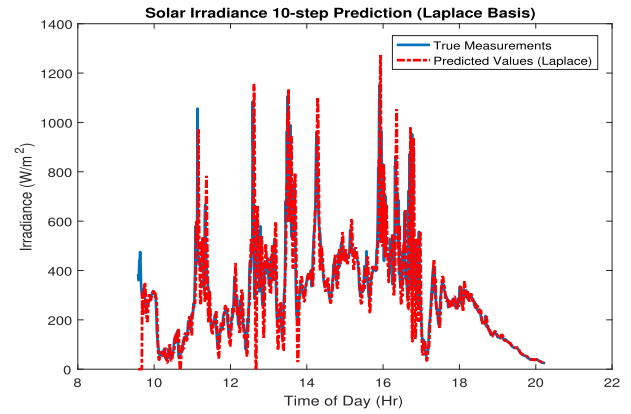
(a) Gaussian distribution



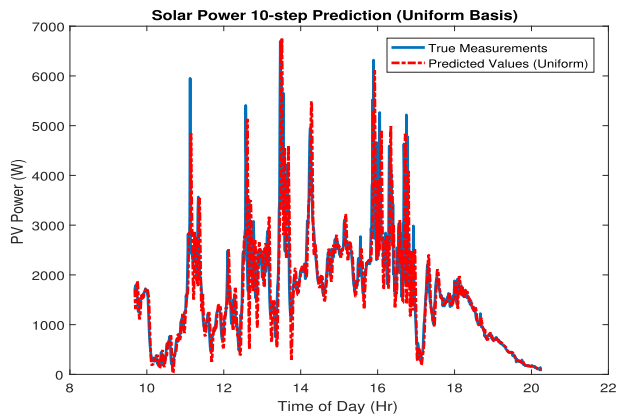
(a) Gaussian distribution



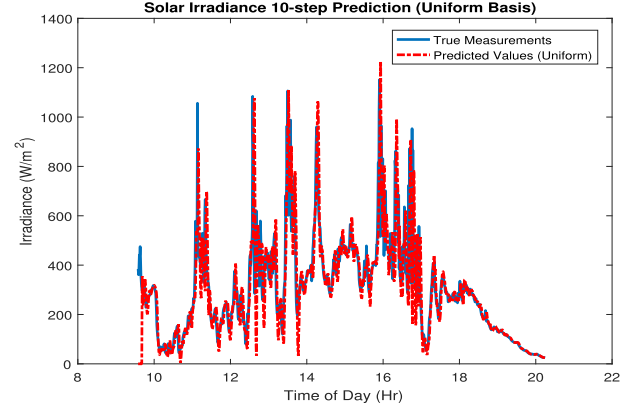
(b) Laplace distribution



(b) Laplace distribution



(c) Uniform distribution



(c) Uniform distribution

Fig. 2. 5 min solar power forecasting using uncertain basis functions.

Fig. 3. 5 min solar radiation forecasting using uncertain basis functions.

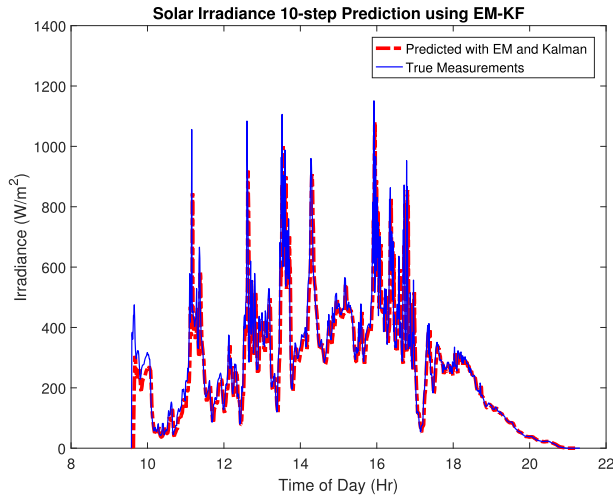
data, respectively using a 4th order *state-space model*. Observe that the radiation has been forecasting with very high accuracy. It takes only few iterations (about 5 iterations) for the forecasting algorithm to converge. The forecasting errors for each scenario are illustrated in Table 2.

Based on these observations, it can be concluded that the EM algorithm combined with the Kalman filter can dynamically track the short-term stochastic behavior of future solar radiation and PV power measurements. In addition, it can be concluded that a 4th

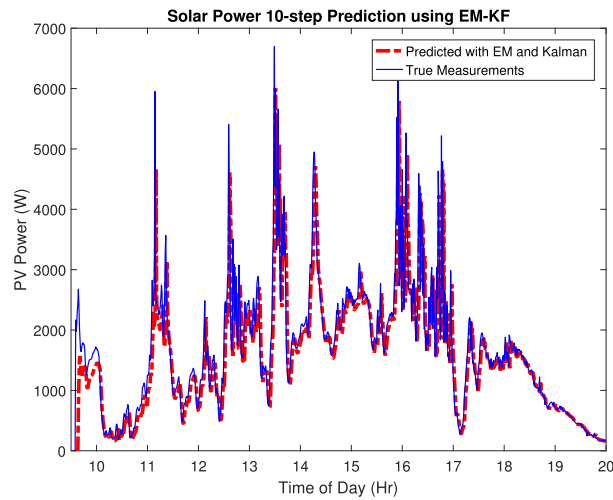
order stochastic model is sufficient to capture the dynamical fluctuations of PV measurements.

Moreover, forecasting results using the sibling PEM technique are provided in Fig. 5. Obviously, the EM-KF state-space technique outperforms the PEM technique.

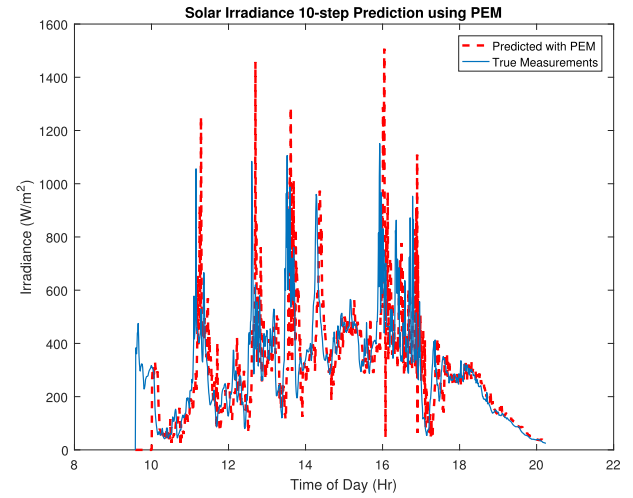
Finally, we show the baseline forecasting performance using the ARIMA method in Fig. 6. Observing that the ARIMA method misses most of the peaks for solar radiation (and PV power) in Fig. 6a and b. This indicates that the popular time series method is not able to



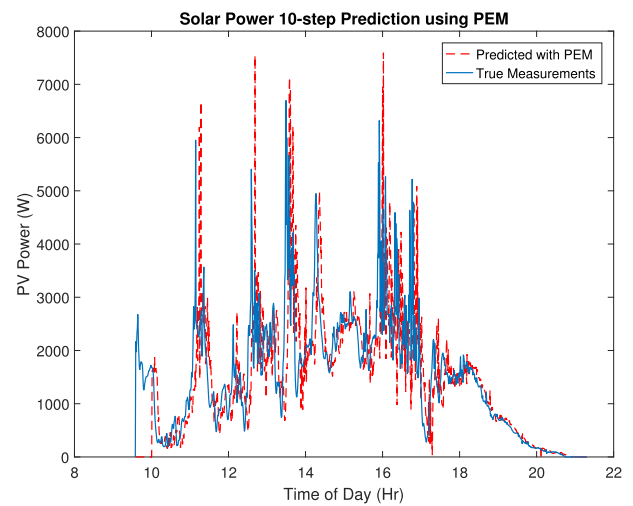
(a) Solar irradiance



(b) Solar power

Fig. 4. 5 min solar radiation and power forecasting using EM-KF.

(a) Solar radiation



(b) Solar power

Fig. 5. 5 min solar radiation and power forecasting using PEM.

track the highly dynamic/uncertain PV signal, which justifies our motivation to introduce different stochastic modeling techniques to overcome this challenge.

5.5. Summary of the results

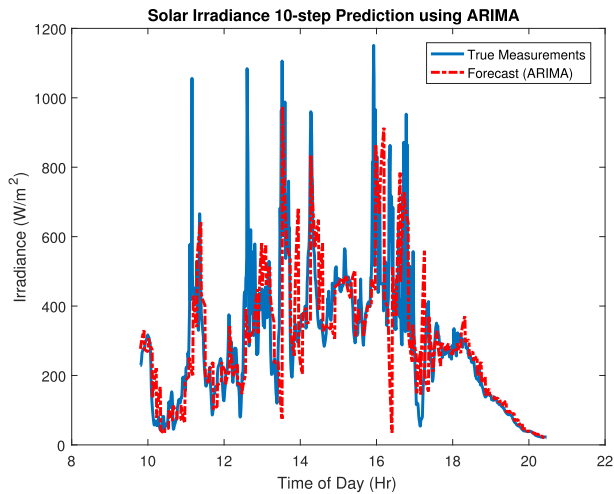
To compare the performance, Table 2 shows the nRMSEs obtained using the different forecasting techniques under different horizons. It can be seen, both the uncertain basis functions and state-space methods present high-accuracy prediction results for both solar radiation and PV power. All three introduced stochastic techniques achieve better forecasting results than the traditional time series method.

Among our three introduced techniques, the EM/Kalman method outperforms PEM method, while the uncertain basis model shows the best performance. For the uncertain basis model, similar forecasting accuracies are achieved utilizing three different probability distributions. There is not an absolute conclusive claim which probability distribution always performs the best. As shown in Table 4, depending on different conditions, the algorithms

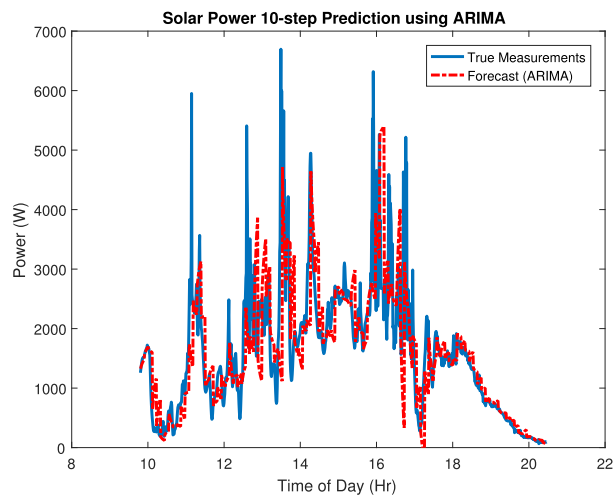
delivering the best and worst performance have been marked. For 1 min prediction horizon, EM-KF performs the best; for 5 min prediction horizon, uncertain basis with gaussian noise achieves the best performance; while for the longer 50 min horizon, uncertain basis with uniform noise and PEM methods deliver better results. Moreover, from the scatter plots in Figs. 7–9, it can be observed that the largest deviations for the proposed forecasting methods occur at peak times, especially when the output power is higher than 3 kW.

The models introduced are more flexible than the traditional ARIMA models, for e.g., it can be shown that the ARIMA model is a special case of the stochastic state-space used in the Kalman with EM, PEM. In the latter there are two noise sources in the state and measurement equations that capture uncertainties in the system, while in the ARIMA there is only one noise source. Therefore, the Kalman with EM, PEM offer more flexibility.

Shorter predictions are closer in time to the last measurements obtained, while longer predictions are farther from them. And therefore, resulting in degraded performance. If the measurements do not change rapidly then it is expected that longer predictions



(a) Solar radiation



(b) Solar power

Fig. 6. 5 min solar radiation and power forecasting using the ARIMA.

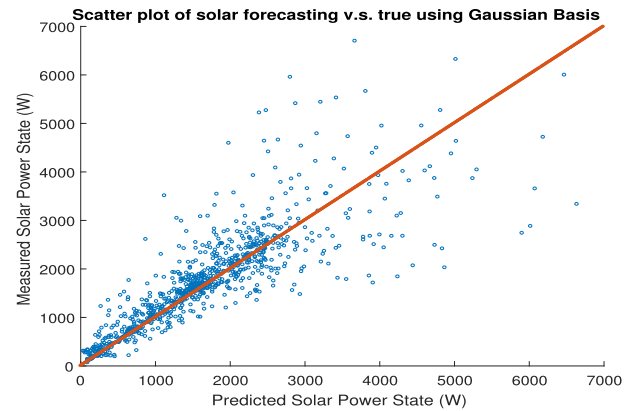
will perform well. However, this is not the case for fast varying processes since the estimated model used for the predictions becomes outdated.

It should be remarked that we do not perform cross validation, since the power of the introduced algorithms is that training and validation are performed in real-time. That is, if the data change the models will change accordingly online so as to adapt. In other words, we iteratively update our uncertain basis and EM-Kalman models based on the most recent data points, then provide future solar forecasting based on the latest model.

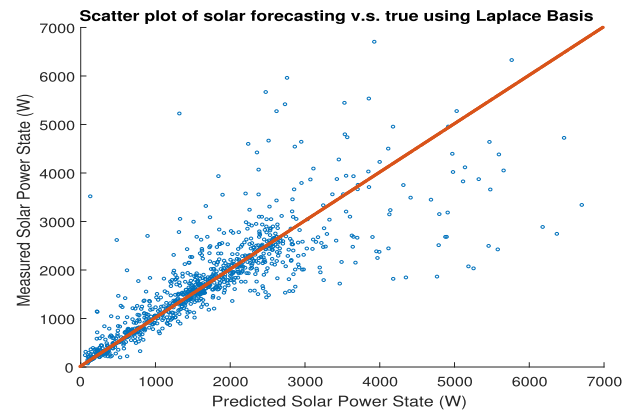
Table 4

: List of best prediction algorithm for each time horizon. nRMSE and nMAPE with subindex R (P) represent forecasting error for solar radiation (PV power).

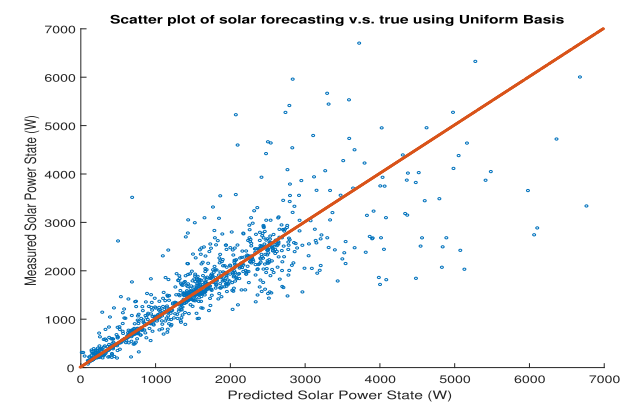
Time horizon (min)	nRMSE _R (%)	nRMSE _P (%)	MAPE _R (%)	MAPE _P (%)
1	UBU	EM-KF	EM-KF	EM-KF
5	UBG	UBG	UBG	UBG
50	PEM	UBU	UBU	PEM



(a) Gaussian distribution



(b) Laplace distribution

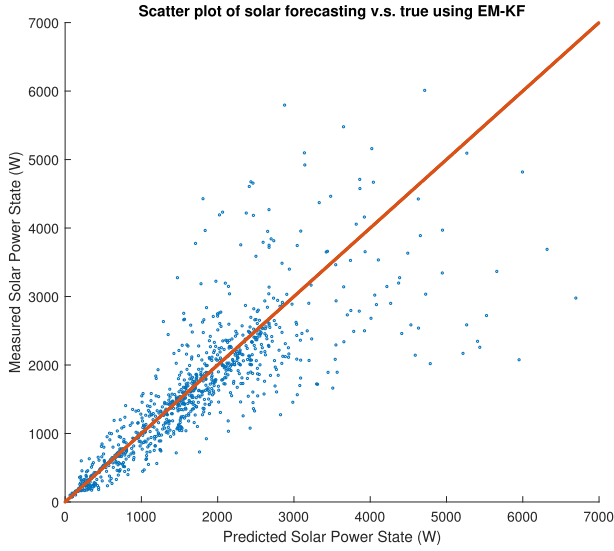


(c) Uniform distribution

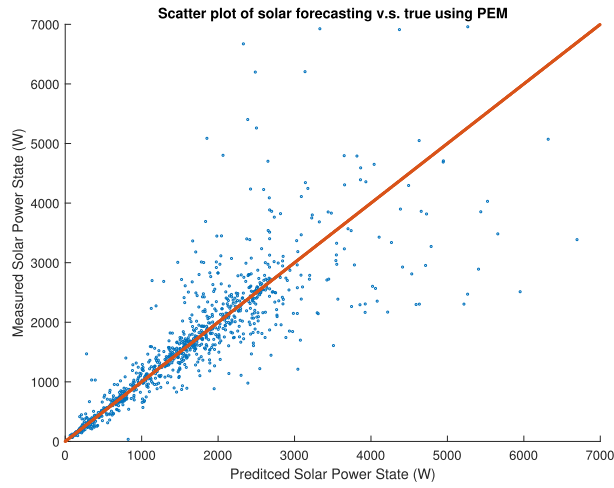
Fig. 7. Scatter plots of 5 min solar power forecasting using uncertain basis functions.

6. Conclusions

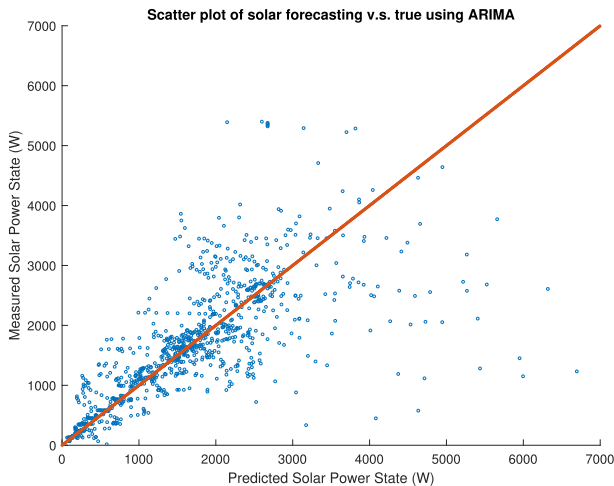
In this paper, we have considered the modeling and forecasting of short-term high-resolution solar radiation and PV power output using two stochastic methodologies. We validated the stochastic models using real measurement data. Given total solar PV generation measurements, the introduced methodologies can forecast short-term high-resolution trends of solar PV power with different time horizons. The uncertain basis model provides an excellent method for forecasting solar PV power. Since the state-space model



(a) EM-KF



(b) PEM

Fig. 8. Scatter plots of 5 min solar power forecasting using EM/PEM.**Fig. 9.** Scatter plots of 5 min solar power forecasting using the ARIMA.

approach is recursive, it can be implemented on-line to provide accurate forecasting of solar radiation and PV power output signals. The introduced methods are further compared with the ARIMA time series method. Both the uncertain basis model and stochastic state-space model can serve as a valuable tool in assessing the feasibility of solar energy collection systems.

For the future work, we will incorporate new factors other than solar PV power measurements into the developed forecasting techniques such as cloud cover, temperature, satellite data, and physical location to further enhance the forecasting accuracy and extend the forecasting horizon.

Acknowledgment

This material is based upon work supported by the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy under contract number DE-AC05-00OR22725..

Appendix A. Computation of Uncertain Basis Functions

Appendix A.1. Uniform distribution

Assume that s_i is uniformly distributed over the interval $[m_i/N - B/2, m_i/N + B/2]$, i.e.

$$s_i \sim \mathcal{U}[m_i/N - B/2, m_i/N + B/2], \quad (\text{A.1})$$

where m_i is a known integer with $1 \leq m_i \leq N$, and $0 < B \leq 1/N$. Note that if we choose $B = 0$, then it reduces to the deterministic case.

Under these statistics, we can calculate $\mathbf{G} = \mathbb{E}[\mathbf{H}(\theta)]$ for uniform distribution as follows:

$$\begin{aligned} [\mathbf{G}]_{ki} &= \int_{m_i/N - B/2}^{m_i/N + B/2} (k+1)^{s_i - 1/N} \frac{1}{B} ds_i \\ &= \frac{(k+1)^{-1/N}}{B} \int_{m_i/N - B/2}^{m_i/N + B/2} e^{s_i \log(k+1)} ds_i \\ &= \frac{(k+1)^{-1/N}}{B \cdot \log(k+1)} e^{\frac{m_i \log(k+1)}{N}} \left(e^{\frac{B}{2} \log(k+1)} - e^{-\frac{B}{2} \log(k+1)} \right), \end{aligned} \quad (\text{A.2})$$

where $k = 1, \dots, N$ and $i = 1, \dots, p$, and $[\mathbf{G}]_{ki}$ represents the k th row and i th column of \mathbf{G} . For \mathbf{C}_H , we need to compute $\text{var}(\phi_{i,k}(\theta_i))$. Note from probability theory that

$$\text{var}(\phi_{i,k}(\theta_i)) = \mathbb{E}((k+1)^{2s_i - 2/N}) - [\mathbf{G}]_{ki}^2. \quad (\text{A.3})$$

After some calculations, we get

$$\begin{aligned} \mathbb{E}((k+1)^{2s_i - 2/N}) &= \int_{m_i/N - B/2}^{m_i/N + B/2} (k+1)^{2s_i - 2/N} \frac{1}{B} ds_i \\ &= \frac{(k+1)^{-2/N}}{2B \cdot \log(k+1)} e^{\frac{2m_i \log(k+1)}{N}} \left(e^{B \cdot \log(k+1)} - e^{-B \cdot \log(k+1)} \right). \end{aligned} \quad (\text{A.4})$$

Then, with (8), (A.3) and (A.4), \mathbf{C}_H is completely determined.

Appendix A.2. Gaussian distribution

In this section, s_i is assumed to be a Gaussian random variable

following the distribution $\mathcal{N}(m_i/N, \sigma_i^2)$. Then, similarly to the previous section, we have

$$\begin{aligned} [\mathbf{G}]_{ki} &= \int_{-\infty}^{+\infty} (k+1)^{s_i-1/N} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(s_i-m_i/N)^2}{2\sigma_i^2}} ds_i \\ &= \frac{(k+1)^{-1/N}}{\sqrt{2\pi\sigma_i^2}} \int_{-\infty}^{+\infty} e^{s_i \log(k+1)} e^{-\frac{(s_i-m_i/N)^2}{2\sigma_i^2}} ds_i \\ &= \frac{(k+1)^{-1/N}}{\sqrt{2\pi\sigma_i^2}} \int_{-\infty}^{+\infty} e^{-\frac{(s_i-m_i/N - \sigma_i^2 \log(k+1))^2}{2\sigma_i^2}} ds_i \\ &\quad \times e^{\frac{1}{2}(\sigma_i^2 \log^2(k+1) + 2m_i \log(k+1)/N)} \\ &= (k+1)^{-1/N} e^{\frac{1}{2}(\sigma_i^2 \log^2(k+1) + 2m_i \log(k+1)/N)}, \end{aligned} \quad (\text{A.5})$$

and to calculate \mathbf{C}_H , we only need to compute the variable below as shown in (A.3)

$$\mathbb{E}\left((k+1)^{2s_i-2/N}\right) = (k+1)^{-2/N} \times e^{2(\sigma_i^2 \log^2(k+1) + m_i \log(k+1)/N)}. \quad (\text{A.6})$$

Appendix A.3 Laplace distribution

In this section, s_i is assumed to be a random variable with a Laplace distribution $\mathcal{L}(k_i/N, b_i)$, whose probability density function is given by $f(x|k_i/N, b_i) = \frac{1}{2b_i} e^{-|x-k_i/N|/b_i}$.

Then, we can compute

$$\begin{aligned} [\mathbf{G}]_{ki} &= \int_{-\infty}^{+\infty} (k+1)^{s_i-1/N} \frac{1}{2b_i} e^{-|s_i-k_i/N|/b_i} ds_i \\ &= \frac{(k+1)^{-1/N} e^{-k_i/(Nb_i)}}{2b_i} \int_{-\infty}^{k_i/N} e^{s_i \log(k+1) + s_i/b_i} ds_i \\ &\quad + \frac{(k+1)^{-1/N} e^{k_i/(Nb_i)}}{2b_i} \int_{k_i/N}^{+\infty} e^{s_i \log(k+1) - s_i/b_i} ds_i \\ &= \frac{(k+1)^{-1/N} e^{k_i \log(k+1)/N}}{2b_i \log(k+1) + 2} + \frac{(k+1)^{-1/N} e^{k_i \log(k+1)/N}}{-2b_i \log(k+1) + 2}, \end{aligned} \quad (\text{A.7})$$

and to calculate \mathbf{C}_H , we only need to compute the variable below as shown in (A.3)

$$\begin{aligned} \mathbb{E}\left((k+1)^{2s_i-2/N}\right) &= \frac{(k+1)^{-2/N} e^{2k_i \log(k+1)/N}}{4b_i \log(k+1) + 2} \\ &\quad + \frac{(k+1)^{-2/N} e^{2k_i \log(k+1)/N}}{-4b_i \log(k+1) + 2}. \end{aligned} \quad (\text{A.8})$$

Appendix B. Computation of Conditional Expectations [48,67]

$$\begin{aligned} L_k^{(1)} &= E\left\{\sum_{i=1}^k x_i^T x_i \middle| Y_k\right\} \\ L_k^{(2)} &= E\left\{\sum_{i=1}^k x_{i-1}^T x_{i-1} \middle| Y_k\right\} \\ L_k^{(3)} &= E\left\{\sum_{i=1}^k [x_i^T x_{i-1} + x_{i-1}^T x_i] \middle| Y_k\right\} \\ L_k^{(4)} &= E\left\{\sum_{i=1}^k [x_i^T y_i + y_i^T x_i] \middle| Y_k\right\} \end{aligned} \quad (\text{B.1})$$

The conditional expectations $\{L_k^{(1)}, L_k^{(2)}, L_k^{(3)}, L_k^{(4)}\}$ are estimated from measurements Y_k as defined follows, which follows [67].

Appendix B.1. Filter estimate of $L_k^{(1)}$

$$\begin{aligned} L_k^{(1)} &= E\left\{\sum_{i=1}^k x_i^T Q x_i \middle| Y_k\right\} = -\frac{1}{2} \text{Tr}\left(N_k^{(1)} P_{k|k}\right) - \frac{1}{2} \sum_{i=1}^k \text{Tr}\left(N_{i-1}^{(1)} \bar{P}_{i|i}\right) \\ &\quad - \frac{1}{2} \sum_{i=1}^k \left(-2x_{i|i-1}^T P_{i|i-1}^{-1} r_i^{(1)} + 2x_{i|i-1}^T P_{i|i-1}^{-1} r_{i|i-1}^{(1)} - x_{i|i}^T N_i^{(1)} x_{i|i} \right. \\ &\quad \left. + x_{i|i-1}^T B_{i-1}^{-2} A_{i-1} \bar{P}_{i|i} N_{i-1}^{(1)} \bar{P}_{i|i} A_{i-1}^T B_{i-1}^{-2} x_{i|i-1} \right) \end{aligned} \quad (\text{B.2})$$

where $\text{Tr}(\cdot)$ denotes the matrix trace. In (B.2), $r_i^{(1)}$ and $N_i^{(1)}$ satisfy the following recursions

$$\begin{cases} r_i^{(1)} = (A_{i-1} - P_{i|i} C_{i-1}^T D_{i-1}^{-2} C_{i-1} A_{i-1}) r_{i-1}^{(1)} + 2P_{i|i} Q x_{i|i-1} \\ \quad - P_{i|i} N_i^{(1)} P_{i|i} C_{i-1}^T D_{i-1}^{-2} (y_i - C_{i-1} x_{i|i-1}) \\ r_{i|i-1}^{(1)} = A_{i-1} r_i^{(1)} \\ r_0^{(1)} = \mathbf{0}_{2n \times 1} \end{cases} \quad (\text{B.3})$$

$$\begin{cases} N_i^{(1)} = B_{i-1}^{-2} A_{i-1} \bar{P}_{i|i} N_{i-1}^{(1)} \bar{P}_{i|i} A_{i-1}^T B_{i-1}^{-2} - 2Q \\ N_0^{(1)} = \mathbf{0}_{2n \times 2n} \end{cases} \quad (\text{B.4})$$

Appendix B.2. Filter estimate of $L_k^{(2)}$

$$\begin{aligned} L_k^{(2)} &= E\left\{\sum_{i=1}^k x_{i-1}^T Q x_{i-1} \middle| Y_k\right\} = E_\beta\{x_0^T Q x_0 \middle| Y_k\} + E_\beta\left\{\sum_{i=1}^k x_i^T Q x_i \middle| Y_k\right\} \\ &\quad - E_\beta\{x_k^T Q x_k \middle| Y_k\} \end{aligned} \quad (\text{B.5})$$

Therefore, $L_k^{(2)}$ can be obtained from $L_k^{(1)}$.

Appendix B.3. Filter estimate of $L_k^{(3)}$

$$\begin{aligned}
L_k^{(3)} &= E \left\{ \sum_{i=1}^k \left(x_i^T R x_{i-1} + x_{i-1}^T R^T x_i \right) \middle| Y_k \right\} \\
&= -\frac{1}{2} \text{Tr} \left(N_k^{(3)} P_{k|k} \right) - \frac{1}{2} \sum_{i=1}^k \text{Tr} \left(N_{i-1}^{(3)} \bar{P}_{i|i} \right) \\
&\quad - \frac{1}{2} \sum_{i=1}^k \left(-2x_{i|i}^T P_{i|i}^{-1} r_i^{(3)} + -2x_{i|i-1}^T P_{i|i-1}^{-1} r_{i-1}^{(3)} - x_{i|i}^T N_i^{(3)} x_{i|i} \right. \\
&\quad \left. + x_{i|i-1}^T B_{i-1}^{-2} A_{i-1} \bar{P}_{i|i} N_{i-1}^{(3)} \bar{P}_{i|i} A_{i-1}^T B_{i-1}^{-2} x_{i|i-1} \right)
\end{aligned} \tag{B.6}$$

In this case, $r_i^{(3)}$ and $N_i^{(3)}$ satisfy the following recursions

$$\begin{cases} r_i^{(3)} = \left(A_{i-1} - P_{i|i} C_{i-1}^T D_{i-1}^{-2} C_{i-1} A_{i-1} \right) r_{i-1}^{(3)} - P_{i|i} N_i^{(3)} P_{i|i} C_{i-1}^T D_{i-1}^{-2} (y_i - C_{i-1} x_{i|i-1}) \\ \quad + \left(2P_{i|i} R + 2P_{i|i} B_{i-1}^{-2} A_{i-1} \bar{P}_{i|i} R^T A_{i-1} \right) \\ \quad r_{i|i-1}^{(3)} = A_{i-1} r_i^{(3)} \\ \quad r_0^{(3)} = 0_{2n \times 2n} \end{cases} \tag{B.7}$$

$$\begin{cases} N_i^{(3)} = B_{i-1}^{-2} A_{i-1} \bar{P}_{i|i} N_{i-1}^{(3)} \bar{P}_{i|i} A_{i-1}^T B_{i-1}^{-2} - 2R \bar{P}_{i|i} A_{i-1}^T B_{i-1}^{-2} - 2B_{i-1}^{-2} A_{i-1} \bar{P}_{i|i} R^T \\ \quad N_0^{(3)} = 0_{2n \times 2n} \end{cases} \tag{B.8}$$

Appendix B.4. Filter estimate of $L_k^{(4)}$

$$\begin{aligned}
L_k^{(4)} &= E \left\{ \sum_{i=1}^k \left(x_i^T S y_i + y_i^T S^T x_i \right) \middle| Y_k \right\} \\
&= \sum_{i=1}^k \left(x_{i|i}^T P_{i|i}^T r_i^{(4)} - x_{i|i-1}^T P_{i|i-1}^{-1} r_{i|i-1}^{(4)} \right)
\end{aligned} \tag{B.9}$$

where $r_i^{(4)}$ satisfy the following recursions

$$\begin{cases} r_i^{(4)} = \left(A_{i-1} - P_{i|i} C_{i-1}^T D_{i-1}^{-2} A_{i-1} \right) r_{i-1}^{(4)} + 2P_{i|i} S y_i \\ \quad r_{i|i-1}^{(4)} = A_{i-1} r_i^{(4)} \\ \quad r_0^{(4)} = 0_{2n \times 1} \end{cases} \tag{B.10}$$

References

- [1] K. Engeland, M. Borga, J.-D. Creutin, B. François, M.-H. Ramos, J.-P. Vidal, Space-time variability of climate variables and intermittent renewable electricity production—a review, *Renew. Sustain. Energy Rev.* 79 (2017) 600–617.
- [2] B. Elsinga, W.G. van Sark, Short-term peer-to-peer solar forecasting in a network of photovoltaic systems, *Appl. Energy* 206 (2017) 1464–1483.
- [3] J. Yan, Y. Zhai, P. Wijayatunga, A.M. Mohamed, P.E. Campana, Renewable energy integration with mini/micro-grids, *Appl. Energy* 201 (2017) 241–244.
- [4] A. Kaur, L. Nonnenmacher, H.T. Pedro, C.F. Coimbra, Benefits of solar forecasting for energy imbalance markets, *Renew. Energy* 86 (2016) 819–830.
- [5] F. Almonacid, P. Pérez-Higueras, E.F. Fernández, L. Hontoria, A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a pv generator, *Energy Convers. Manag.* 85 (2014) 389–398.
- [6] A. Tuohy, J. Zack, S.E. Haupt, J. Sharp, M. Ahlstrom, S. Dise, E. Gritmit, C. Mohrlen, M. Lange, M.G. Casado, et al., Solar forecasting: methods, challenges, and performance, *IEEE Power Energy Mag.* 13 (6) (2015) 50–59.
- [7] M.Q. Raza, M. Nadarajah, C. Ekanayake, On recent advances in pv output power forecast, *Sol. Energy* 136 (2016) 125–144.
- [8] H.T. Pedro, E. Lim, C.F. Coimbra, A database infrastructure to implement real-time solar and wind power generation intra-hour forecasts, *Renew. Energy* 123 (2018) 513–525.
- [9] R.H. Inman, H.T. Pedro, C.F. Coimbra, Solar forecasting methods for renewable energy integration, *Prog. Energy Combust. Sci.* 39 (6) (2013) 535–576.
- [10] R. Dambreville, P. Blanc, J. Chanussot, D. Boldo, Very short term forecasting of the global horizontal irradiance using a spatio-temporal autoregressive model, *Renew. Energy* 72 (2014) 291–300.
- [11] L.M. Aguiar, B. Pereira, P. Lauret, F. Díaz, M. David, Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting, *Renew. Energy* 97 (2016) 599–610.
- [12] J. Alonso-Montesinos, F. Batlle, C. Portillo, Solar irradiance forecasting at one-minute intervals for different sky conditions using sky camera images, *Energy Convers. Manag.* 105 (2015) 1166–1177.
- [13] A.T. Lorenzo, W.F. Holmgren, A.D. Cronin, Irradiance forecasts based on an irradiance monitoring network, cloud motion, and spatial averaging, *Sol. Energy* 122 (2015) 1158–1169.
- [14] F. Barbieri, S. Rajakaruna, A. Ghosh, Very short-term photovoltaic power forecasting with cloud modeling: a review, *Renew. Sustain. Energy Rev.* 75 (2017) 242–263.
- [15] S. Monjoly, M. André, R. Calif, T. Soubdhan, Hourly forecasting of global solar radiation based on multiscale decomposition methods: a hybrid approach, *Energy* 119 (2017) 288–298.
- [16] M. Kocifaj, Angular distribution of scattered radiation under broken cloud arrays: an approximation of successive orders of scattering, *Sol. Energy* 86 (12) (2012) 3575–3586.
- [17] M. Kocifaj, Unified model of radiance patterns under arbitrary sky conditions, *Sol. Energy* 115 (2015) 40–51.
- [18] V.P. Lonij, A.E. Brooks, A.D. Cronin, M. Leuthold, K. Koch, Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors, *Sol. Energy* 97 (2013) 58–66.
- [19] J.L. Bosch, J. Kleissl, Cloud motion vectors from a network of ground sensors in

- a solar power plant, *Sol. Energy* 95 (2013) 13–20.
- [20] I. Jayawardene, G.K. Venayagamoorthy, Spatial predictions of solar irradiance for photovoltaic plants, in: *Photovoltaic Specialists Conference (PVSC)*, 2016 IEEE 43rd, IEEE, 2016, pp. 0267–0272.
 - [21] J. Domke, N. Engerer, A. Menon, C. Webers, Distributed solar prediction with wind velocity, in: *Photovoltaic Specialists Conference (PVSC)*, 2016 IEEE 43rd, IEEE, 2016, pp. 1218–1223.
 - [22] J. Li, J.K. Ward, J. Tong, L. Collins, G. Platt, Machine learning for solar irradiance forecasting of photovoltaic system, *Renew. Energy* 90 (2016) 542–553.
 - [23] R. Azimi, M. Ghayekhloo, M. Ghofrani, A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting, *Energy Convers. Manag.* 118 (2016) 331–344.
 - [24] Y. Chu, C.F. Coimbra, Short-term probabilistic forecasts for direct normal irradiance, *Renew. Energy* 101 (2017) 526–536.
 - [25] K. Mohammadi, S. Shamsirband, C.W. Tong, M. Arif, D. Petković, S. Ch, A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation, *Energy Convers. Manag.* 92 (2015) 162–171.
 - [26] J. Cao, X. Lin, Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks, *Energy Convers. Manag.* 49 (6) (2008) 1396–1406.
 - [27] J. Liu, W. Fang, X. Zhang, C. Yang, An improved photovoltaic power forecasting model with the assistance of aerosol index data, *IEEE Trans. Sustain. Energy* 6 (2) (2015) 434–442.
 - [28] D. Yang, Z. Ye, L.H.I. Lim, Z. Dong, Very short term irradiance forecasting using the lasso, *Sol. Energy* 114 (2015) 314–326.
 - [29] F.O. Hocaoglu, Stochastic approach for daily solar radiation modeling, *Sol. Energy* 85 (2) (2011) 278–287.
 - [30] H. Morf, Sunshine and cloud cover prediction based on markov processes, *Sol. Energy* 110 (2014) 615–626.
 - [31] M.J. Sanjari, H.B. Gooi, Probabilistic forecast of pv power generation based on higher-order Markov chain, *IEEE Trans. Power Syst.* 32 (4) (2017) 2942–2952.
 - [32] E. Ranganai, M.B. Nzuza, A comparative study of the stochastic models and harmonically coupled stochastic models in the analysis and forecasting of solar radiation data, *J. Energy South. Afr.* 26 (1) (2015) 125–137.
 - [33] J. Huang, M. Korolkiewicz, M. Agrawal, J. Boland, Forecasting solar radiation on an hourly time scale using a coupled autoregressive and dynamical system (cards) model, *Sol. Energy* 87 (2013) 136–149.
 - [34] J. Boland, Spatial-temporal forecasting of solar radiation, *Renew. Energy* 75 (2015) 607–616.
 - [35] J. Xu, S. Yoo, D. Yu, H. Huang, D. Huang, J. Heiser, P. Kalb, A stochastic framework for solar irradiance forecasting using condition random field, in: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, 2015, pp. 511–524.
 - [36] M. Diagne, M. David, J. Boland, N. Schmutz, P. Lauret, Post-processing of solar irradiance forecasts from wrf model at reunion island, *Sol. Energy* 105 (2014) 99–108.
 - [37] G. Galanis, P. Louka, P. Katsafados, I. Pytharoulis, G. Kallos, Applications of kalman filters based on non-linear functions to numerical weather predictions, *Ann. Geophys.* 24 (2006) 2451–2460.
 - [38] S. Pelland, G. Galanis, G. Kallos, Solar and photovoltaic forecasting through post-processing of the global environmental multiscale numerical weather prediction model, *Prog. Photovoltaics Res. Appl.* 21 (3) (2013) 284–296.
 - [39] T. Soubdhan, J. Ndong, H. Ould-Baba, M.-T. Do, A robust forecasting framework based on the kalman filtering approach with a twofold parameter tuning procedure: application to solar and photovoltaic prediction, *Sol. Energy* 131 (2016) 246–259.
 - [40] J.-J. Guo, P.B. Luh, Selecting input factors for clusters of Gaussian radial basis function networks to improve market clearing price prediction, *IEEE Trans. Power Syst.* 18 (2) (2003) 665–672.
 - [41] Z. Bashir, M. El-Hawary, Applying wavelets to short-term load forecasting using pso-based neural networks, *IEEE Trans. Power Syst.* 24 (1) (2009) 20–27.
 - [42] J. Dong, T. Kuruganti, S.M. Djouadi, Very short-term photovoltaic power forecasting using uncertain basis function, in: *Information Sciences and Systems (CISS)*, 2017 51st Annual Conference on, IEEE, 2017, pp. 1–6.
 - [43] M. Olama, A. Melin, J. Dong, S.M. Djouadi, Y. Zhang, Stochastic short-term high-resolution prediction of solar irradiance and photovoltaic power output, in: *North American Power Symposium (NAPS)*, 2017, IEEE, 2017, pp. 1–6.
 - [44] S. Kay, Signal fitting with uncertain basis functions, *Signal Process. Lett. IEEE* 18 (6) (2011) 383–386.
 - [45] J. Dong, X. Ma, S.M. Djouadi, H. Li, T. Kuruganti, Real-time prediction of power system frequency in fnet: a state space approach, in: *Smart Grid Communications (SmartGridComm)*, 2013 IEEE International Conference on, IEEE, 2013, pp. 109–114.
 - [46] J. Dong, X. Ma, S.M. Djouadi, H. Li, Y. Liu, Frequency prediction of power systems in fnet based on state-space approach and uncertain basis functions, *IEEE Trans. Power Syst.* 29 (6) (2014) 2602–2612.
 - [47] M. Niedzwiecki, T. Klaput, Fast recursive basis function estimators for identification of time-varying processes, *IEEE Trans. Signal Process.* 50 (8) (2002) 1925–1934.
 - [48] C.D. Charalambous, A. Logothetis, Maximum likelihood parameter estimation from incomplete data via the sensitivity equations: the continuous-time case, *IEEE Trans. Autom. Control* 45 (5) (2000) 928–934.
 - [49] R.J. Elliott, V. Krishnamurthy, New finite-dimensional filters for parameter estimation of discrete-time linear Gaussian models, *IEEE Trans. Autom. Control* 44 (5) (1999) 938–951.
 - [50] G. Welch, G. Bishop, *An Introduction to the Kalman Filter*.
 - [51] M. Verhaegen, V. Verdult, *Filtering and System Identification: a Least Squares Approach*, Cambridge university press, 2007.
 - [52] C.J. Wu, On the convergence properties of the em algorithm, *Ann. Stat.* (1983) 95–103.
 - [53] L. Ljung, *System Identification*, Wiley Online Library, 1999.
 - [54] J. Faraway, C. Chatfield, *Time Series Forecasting with Neural Networks: A Case Study*, Research Report, University of Bath, Bath (United Kingdom), 1995, 95–06.
 - [55] P.J. Brockwell, R.A. Davis, *Time Series: Theory and Methods*, Springer Science & Business Media, 2013.
 - [56] W. Ji, K.C. Chee, Prediction of hourly solar radiation using a novel hybrid model of arma and tdn, *Sol. Energy* 85 (5) (2011) 808–817.
 - [57] C. Craggs, E. Conway, N. Pearsall, Stochastic modelling of solar irradiance on horizontal and vertical planes at a northerly location, *Renew. Energy* 18 (4) (1999) 445–463.
 - [58] P. Jain, E. Lungu, Stochastic models for sunshine duration and solar irradiation, *Renew. Energy* 27 (2) (2002) 197–209.
 - [59] T.C. Mills, *Time Series Techniques for Economists*, Cambridge University Press, 1991.
 - [60] A.J. Conejo, M.A. Plazas, R. Espinola, A.B. Molina, Day-ahead electricity price forecasting using the wavelet transform and arima models, *IEEE Trans. Power Syst.* 20 (2) (2005) 1035–1042.
 - [61] M. Hassanzadeh, M. Etezadi-Amoli, M. Fadali, Practical approach for sub-hourly and hourly prediction of pv power output, in: *North American Power Symposium (NAPS)*, 2010, IEEE, 2010, pp. 1–5.
 - [62] D. Yang, P. Jirutitijaroen, W.M. Walsh, Hourly solar irradiance time series forecasting using cloud cover index, *Sol. Energy* 86 (12) (2012) 3531–3543.
 - [63] J.A. Magerko, Y. Cao, P.T. Krein, Quantifying photovoltaic fluctuation with 5 khz data: implications for energy loss via maximum power point trackers, in: *Power and Energy Conference at Illinois (PECI)*, 2016 IEEE, IEEE, 2016, pp. 1–7.
 - [64] C. Feng, M. Cui, M. Lee, J. Zhang, B.-M. Hodge, S. Lu, H.F. Hamann, Short-term global horizontal irradiance forecasting based on sky imaging and pattern recognition, in: *2017 IEEE Power & Energy Society General Meeting*, IEEE, 2017, pp. 1–5.
 - [65] J. Kleissl, *Solar Energy Forecasting and Resource Assessment. Overview of Solar-Forecasting Methods and a Metric for Accuracy Evaluation*, Academic Press, 2013.
 - [66] F. Kasten, A simple parameterization of the pyrheliometric formula for determining the linke turbidity factor, *Meteorol. Rundsch.* 33 (1980) 124–127.
 - [67] M.M. Olama, S.M. Djouadi, C.D. Charalambous, Stochastic differential equations for modeling, estimation and identification of mobile-to-mobile communication channels, *IEEE Trans. Wirel. Commun.* 8 (4) (2009) 1754–1763.