






A Model for Joint Probabilistic Forecast of Solar Photovoltaic Power and Outdoor Temperature

Raksha Ramakrishna , *Student Member, IEEE*, Anna Scaglione , *Fellow, IEEE*, Vijay Vittal , *Fellow, IEEE*, Emiliano Dall'Anese , *Member, IEEE*, and Andrey Bernstein , *Member, IEEE*

Abstract—In this paper, a stochastic model is proposed for a joint statistical description of solar photovoltaic (PV) power and outdoor temperature. The underlying correlation emerges from solar irradiance that is responsible in part for both the variability in solar PV power and temperature. The proposed model can be used to capture the uncertainty in solar PV power and its correlation with the electric power consumption of thermostatically controlled loads. First, a model for solar PV power that explicitly incorporates the stochasticity due to clouds via a regime-switching process between the three classes of *sunny*, *overcast* and *partly cloudy* is proposed. Then, the relationship between temperature and solar power is postulated using a second-order Volterra model. This joint modeling is leveraged to develop a joint probabilistic forecasting method for solar PV power and temperature. Real-world datasets that include solar PV power and temperature measurements in California are analyzed and the effectiveness of the joint model in providing probabilistic forecasts is verified. The proposed forecasting methodology outperforms several reference methods thus portraying that the underlying correlation between temperature and solar PV power is well defined and only requires a simple lower-complexity sampling space.

Index Terms—Probabilistic forecast solar power, Roof-top solar panels, Dictionary learning, Hidden Markov Models, Volterra system, temperature forecast.

I. INTRODUCTION

SOLAR power generation, both from solar photovoltaic (PV) farms and roof-top solar panel installations is on the rise and their impact is increasingly felt in traditional energy markets. Being a weather-dependent natural resource, attached to this

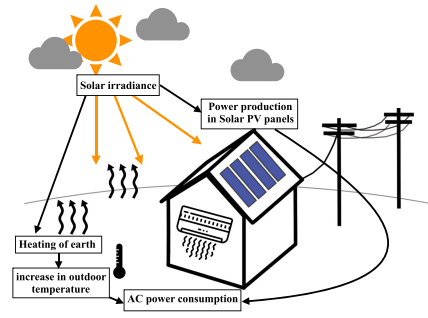


Fig. 1. Figure highlighting the relationship between solar PV power and thermostatically controlled loads.

renewable energy production is certain level of variability and uncertainty that cannot be controlled. This calls for greater amounts of reserves to be online to compensate for renewable generation shortage. On the other hand, electricity consumption is also affected by the weather. In fact, the so called thermostatically controlled loads (TCL), i.e. the appliances used for heating and/or cooling, have a consumption pattern that has direct dependence on outdoor temperature. Naturally, solar irradiation affects both solar power production as well as temperature, making supply of solar power and TCL demand statistically dependent. This is illustrated in Fig. 1. These underlying dependencies, however, are not leveraged in traditional power systems operations. Today, there is significant interest in introducing stochastic optimization tools in power systems operations [3], [4] and the challenge is the explosion of complexity associated with representing such uncertainty. Typically, accounting for uncertainty involves constructing scenario trees [5] i.e., quantization of multidimensional random processes with random variables being load, renewable generation and electricity prices. Treating solar production and TCL load, or demand response, as independent random processes is inaccurate and inefficient besides increasing the domain of the associated random variables. By capturing the relationship between temperature and solar power, using the model presented in this paper, one can relate renewable infeed (solar PV power) with TCLs (see Fig. 1). Such modeling leads to reduction in the size of the sample space by excluding the unlikely joint events of occurrence of a certain PV output power and TCL consumption. A smaller sample space neglects low-probability outcomes and limits the computational complexity while computing expected cost which is the objective function in a stochastic optimization problem. [3]

Manuscript received August 22, 2018; revised June 4, 2019 and November 3, 2019; accepted November 4, 2019. Date of publication November 22, 2019; date of current version December 11, 2019. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Pierre Borgnat. This work was supported in part by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award DE-AR0000696, in part by National Science Foundation under Grant CPS-1549923, and in part by Laboratory Directed Research and Development at the National Renewable Energy Laboratory. This paper was presented in part at the 50th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, November 2016 [1] and in part at the IEEE International Conference on Acoustics, Speech, and Signal Processing, Calgary, AB, April 2018 [2]. (*Corresponding author: Raksha Ramakrishna.*)

R. Ramakrishna, A. Scaglione, and V. Vittal are with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85281 USA (e-mail: raksha.ramakrishna@asu.edu; anna.scaglione@asu.edu; vijay.vittal@asu.edu).

E. Dall'Anese is with the Department of Electrical, Computer and Energy Engineering, University of Colorado Boulder, Boulder, CO 80309 USA (e-mail: emiliano.dallanese@colorado.edu).

A. Bernstein is with the National Renewable Energy Laboratory, Golden, CO 80401 USA (e-mail: andrey.bernstein@nrel.gov).

Digital Object Identifier 10.1109/TSP.2019.2954973

The main goal of this paper is to provide a mathematical model for the joint uncertainty in solar PV power and temperature which can also be used to produce probabilistic forecasts. The modeling approach is motivated by the physics of the problem which helps in understanding the underlying phenomenon and provides an easier interpretation of the results obtained. The proposed forecasting method provides *parametric* predictive densities which is useful for creating scenario trees for stochastic optimization to take decisions under uncertainty. The premise is that solar irradiation is the cause of both solar PV power and temperature and therefore acts as a confounding or lurking variable [6]. Solar PV power exclusively depends on solar irradiance which means that the random variable of solar power is a direct transformation of irradiance. The uncertainty in solar power or irradiance is dictated by the cloud cover. The residual uncertainty in temperature given solar power is then captured by the means of a conditional density function. A vast array of literature exists in the area of forecasts for solar power. On the other hand, there are very few papers that model the relationship that solar PV power has with TCLs. These are reviewed separately in the subsequent subsections.

A. Prior Works: Solar PV Power Forecasting

A majority of approaches taken to model solar power generation can be broadly classified as being physical, statistical or a hybrid of the two methods (see e.g. [7] for a review). Physical methods employ astronomical relationships [8], meteorological conditions and numerical weather predictions (NWP) for an improved forecast [9]–[12]. Other papers use static images of clouds recorded by a total sky imager (TSI) [13] recording cloud motion [14] to predict solar power. These models rely on a deterministic mapping given additional information to produce an estimate of the power generated by the panel.

One of the most prevalent statistical methods for solar power forecasting include time-series modeling such as using autoregressive (AR) models [15]–[17] that are designed to model stationary processes. They assume that some transformation such as dividing by the clear sky power time series makes the series stationary. Such assumptions may not be sufficient to fully capture the non-stationarity of solar power production. Alternatively, there are other approaches such as autoregressive integrated moving average (ARIMA) and autoregressive with exogenous input (ARX) [7], [18], [19] based non-stationary methods for solar power prediction. In the same class of statistical methods there exist other works that capture variability in solar PV power in a multitude of ways [20]–[22] including the use of black box methods like artificial neural networks (ANN) [23], [24]. Additionally, there are also methods based on the Markovian assumption on solar PV power [25], [26] for forecasting. The body of work on probabilistic forecasting of solar PV power is relatively small. In [27], a non-parametric kernel density estimation method is used to fit a probability distribution to PV power. In [28], a higher order Markov chain is used to characterize solar PV power and operating points based on temperature are also used to classify different PV systems and then Gaussian mixture models (GMM) are used for probabilistic

forecasts. In [29], a combination of beta distribution, quantile regression and Markov model is used for probabilistic forecasting. More recently, producing parametric predictive densities was undertaken in [30] where beta and two-sided power distributions were used to provide an aggregated forecast distribution. A review of approaches for probabilistic forecasting of solar PV power can be found in [31].

B. Prior Work: Joint Modeling

Empirical studies such as in [32], [33] and data driven approaches like in [34] demonstrated a relationship between temperature and PV power. In [35], ARMA models are utilized to model PV power generation, load and wind generation separately, and their profiles are used to investigate reliability of the resultant hybrid power system which consists of renewable infeed and flexible loads. In the area of stochastic optimization, clustering methods such as k -Nearest Neighbors (k -NN) along with auxiliary variables such as temperature are used in [36] to create k scenarios for PV power and TCL load. Correlation was accounted for in this case; however, equal probabilities were assigned to all k scenarios. Self-consumption of photo-voltaic power in buildings, i.e. usage of rooftop solar PV power production for satisfying the load requirements within a building has also been studied to some extent. See [37], for a literature review.

C. Contributions

The goal of this paper is to present a unified forecasting approach using a joint stochastic model for solar PV power and temperature that is physically inspired. To highlight this, Fig. 2 illustrates the block diagram of the joint stochastic model. The deterministic models for solar PV power and temperature were introduced in [1] and [2] respectively. In this paper, a stochastic model for solar PV power is introduced and a forecasting methodology is developed. In combination with the model for temperature [2], a joint stochastic model is presented.

The proposed model for solar PV power at a macro-level defines a regime switching process [38] that classifies the PV production under one of the following three settings: *sunny*, *partly cloudy*, *overcast*. The stochastic models for *sunny* and *overcast* are Gaussian distributions whereas for the *partly cloudy* regime, a hidden Markov model (HMM) is proposed. Such an approach simplifies the understanding of temporal variations in solar PV power by examining each regime separately. It is also important to note that no assumption of stationarity is made while describing the regime switching process and no attempt is made to estimate this time-varying transition probability. In this manner, the proposed method uniquely captures the non-stationarity in solar power, which is not just due to its diurnal structure, and maps that onto the non-stationarity of the temperature process.

Prior work in [1] by the authors briefly described in Section II involved the development of a parametric model that was proven to efficiently capture the effect of clouds on solar PV power while providing a compact representation. In this paper, the prior modeling technique is extended to fit a switching process to solar PV

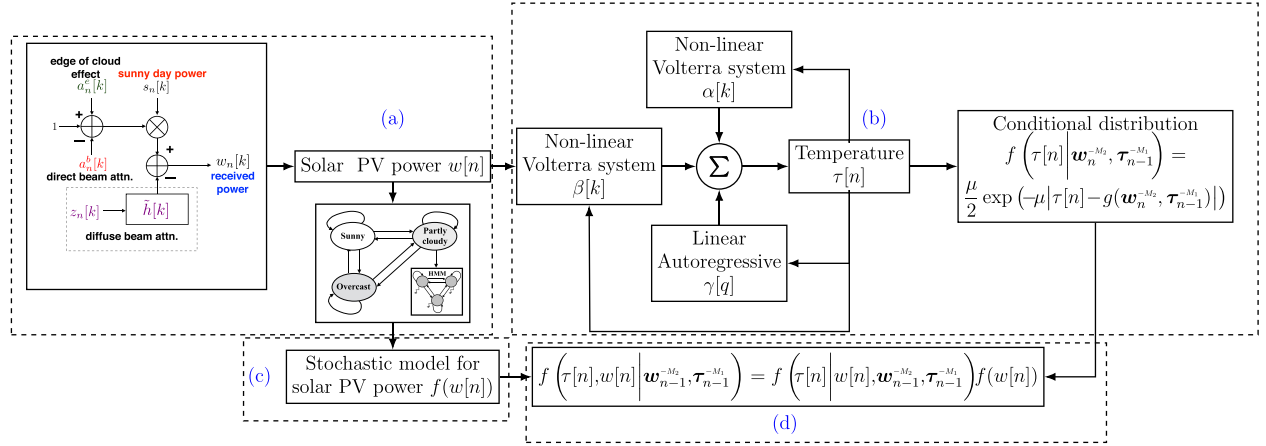


Fig. 2. The block diagram describing the overall joint system modeling approach in this paper. (a) A deterministic system model for solar PV power was introduced in [1]. (b) A joint system model for both solar PV power and temperature was introduced in [2]. (c) Stochastic model for solar PV power is introduced in this paper so that it can be used for (d) joint probabilistic forecasting methods.

power, using which a solar power prediction algorithm to provide short-term probabilistic forecasts is designed. The resulting low order model ensures reduced computational complexity for the proposed algorithm. The key contributions of this paper are:

- The proposition of a regime-switching process for solar PV power that consists of periods that can be classified as *sunny*, *overcast* and *partly cloudy* and development of stochastic models for the three regimes.
- A HMM for the *partly cloudy* regime whose latent states are the support of sparse parameters pertaining to attenuation of power.
- A classification algorithm to identify the present regime using solar power data and use it for solar power prediction. No other auxiliary data such as wind speed is used.
- A Volterra model to capture the correlation between temperature and solar PV power for a joint forecast of temperature and solar PV power in the probabilistic sense using the reduced size uncertainty model that emerges from the proposed modeling approach.

D. Paper Organization

In Section II, the deterministic modeling strategies for solar PV power and temperature from [1] and [2] respectively are reviewed. In Section III, the joint stochastic models is detailed. Section VI contains the numerical results by testing the forecasting methodology with real-world data and benchmarking with other reference methods.

II. DISCRETE-TIME MODEL FOR SOLAR PV POWER AND TEMPERATURE

In this section, the discrete time model for solar PV power output that was derived in detail in the authors' prior work in [1] and Volterra-type model relating temperature and solar irradiance developed in [2] are briefly reviewed as a precursor to the joint stochastic model. Instead of directly formulating a stochastic model, the deterministic model with parameters is

first constructed. Then, the stochasticity in parameters is characterized and further leveraged in Section III to define stochastic models for solar power output and temperature.

A. Solar PV Power Model

It is hypothesized that the panel sums solar irradiation by weighting each contribution with a bi-dimensional gain function that handles the scaling factors to obtain total electrical power. The solar irradiation is attenuated by clouds modeled as a random mask that subtracts a percentage of the light coming from the patch of sky it covers at a certain time. The motion of the clouds over the panel can be approximated to be moving at a constant speed in a certain direction throughout the day. This assumption is reasonable considering the size of the panel relative to that of the displacement of the clouds. It is known that the solar irradiation has two major components [39], *direct beam component* and *diffuse beam component*. Each of these components is attenuated by the cloud coverage in different ways. Let the power produced by the PV panel be $w_d[n]$ on day d and time instant $n \in \{0, \dots, N\}$, then:

$$w_d[n] = s_d[n] - \left(p_d^b[n] + p_d^{dif}[n] \right) + p_d^e[n] + \eta_d[n] \quad (1)$$

where $s_d[n]$ is the solar power if the d th day is sunny, $p_d^b[n]$ and $p_d^{dif}[n]$ are the components pertaining to direct and diffused beam component attenuation by the clouds respectively, $p_d^e[n]$ is attributed to edge of the cloud effect and $\eta_d[n]$ is Gaussian measurement and modeling noise.

Sunny days parametrization: Each cloudless day is modeled using a simple basis expansion model, whose expansion coefficients are periodically updated to reflect seasonal variations. Let \mathcal{S} denote the set of sunny days. For $d \in \mathcal{S}$ the solar PV power samples are modeled as:

$$w_d[n] = s_d[n] = \sum_{q=0}^Q s_{dq} b_q[n] + \eta_d[n] \quad (2)$$

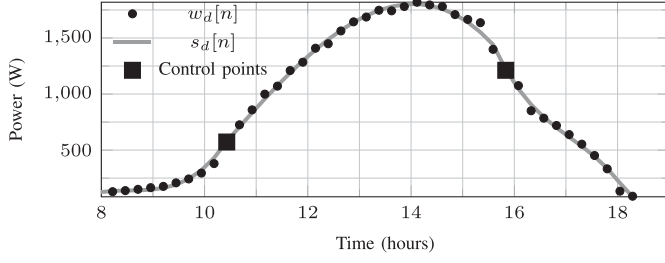


Fig. 3. Plot showing $s_d[n]$ and $w_d[n]$ for a sunny day on October 22, 2009.

where $\{b_q[n]\}$ are given basis functions; for example, one could use a set of non-overlapping cubic splines that cover three daylight periods delimited by two control points n_{k1}, n_{k2} chosen as time instants at which an abrupt change in the data (first and second order, i.e. C^1 and C^2 discontinuity) is observed. This particular basis is constructed using $Q = 9$ i.e. 10 functions derived from Bernstein polynomials [40], $B_{j,\nu}(t)$ of degree $\nu = 3$ as

$$b_{q=\nu i+j}[n] = B_{j,\nu}(t_i[n]), \quad i = 0, 1, 2 \quad (3)$$

$$B_{j,\nu}(t) = \binom{\nu}{j} t^j (1-t)^{\nu-j} \text{rect}(t), \quad j = 0, 1, 2, 3, \quad (4)$$

$\text{rect}(t)$ denotes the rectangular function between $[0, 1)$ and

$$t_0[n] = n/n_{k1}, \quad 0 \leq n < n_{k1} \quad (5)$$

$$t_1[n] = (n - n_{k1})/(n_{k2} - n_{k1}), \quad n_{k1} \leq n < n_{k2} \quad (6)$$

$$t_2[n] = (n - n_{k2})/(N - n_{k2}), \quad n_{k2} \leq n \leq N \quad (7)$$

Coefficients s_{dq} are found by linear regression with respect to power data $w_d[n]$. The residual error is minimal for sunny days, $w_d[n] = s_d[n]$. The approximated $s_d[n]$ for a sunny day is shown in Fig. 3. It highlights the very specific pattern obtained in October due to shading.

Cloudy days parametrization: By referring to the previous work in [1] where the expressions for $p_d^b[k]$ and $p_d^{dif}[n]$ were derived as,

$$p_d^b[n] \approx a_d^b[n] s_d[n], \quad p_d^{dif}[n] \approx \sum_q h[q] z_d[n - q] \quad (8)$$

and where $a_d^b[n]$ is the stochastic time series capturing the sudden power attenuations caused by clouds directly occluding the sun. The diffuse beam attenuation is smoother and, thus, is modeled as the convolution of a one-dimensional filter $h[k]$ with a stochastic input $z_d[n]$ that represents the cloud attenuation. Furthermore, to explain the increase of power even beyond the expected sunny day power $s_d[n]$, along the lines of direct beam attenuation, the term $p_d^e[n]$ is introduced to be present only when $w_d[n] > s_d[n]$,

$$p_d^e[n] \approx a_d^e[n] s_d[n], \quad (9)$$

This term captures the so called *edge of the cloud effect* that has been reported in literature [41], [42]. The edges of some clouds act like magnifying lenses when their paths intersect with that of the sun thereby boosting the power. It is important to note that the edge of cloud effect cannot occur simultaneously with cloud related attenuation and in general this term will be far sparser.

Complete model: From the cloudy day parametrization, the complete model for power on day d is written in vector form as

$$w_d = s_d - \mathbf{U} [\mathbf{S}_d \mathbf{a}_d^b + \mathcal{T}(\mathbf{h}) \mathbf{z}_d] + \tilde{\mathbf{U}} \mathbf{S}_d \mathbf{a}_d^e + \boldsymbol{\eta}_d \quad (10)$$

where $w_d, s_d, \mathbf{a}_d^b, \mathbf{a}_d^e \in \mathbb{R}_+^N$, $\mathbf{z}_d \in \mathbb{R}_+^{(N+M-1)}$ and $\mathbf{U} = \text{diag}(U(s_d - w_d))$, where $U(\cdot)$ is the Heaviside step function operating element-wise, $\tilde{\mathbf{U}} = \mathbf{I} - \mathbf{U}$ and \mathbf{I} is the identity matrix of size N . The convolution term in matrix-vector form with extended end conditions [43] as $\mathcal{T}(\mathbf{h}) \mathbf{z}_d$ where, $\mathbf{h} \in \mathbb{R}_+^M$ and $\mathcal{T}(\mathbf{h}) \in \mathbb{R}_+^{N \times (N+M-1)}$ is the Toeplitz matrix with first column $[h[M-1], \mathbf{0}^{1 \times N-1}]^T$ and first row $[h[M-1], \dots, h[0], \mathbf{0}^{1 \times N-1}]$. Cloud cover parameters $\mathbf{a}_d^b, \mathbf{a}_d^e$ are assumed to be sparse since direct cloud occlusions are rare events in time and \mathbf{z}_d is regularized to avoid overfitting. Here, the estimation of the cloud coverage parameters is seen as a blind deconvolution problem that falls in the class of sparse dictionary learning problems [44], [45]; this can be solved, for example, by alternating between the estimation of the vectors $\mathbf{z}_d, \mathbf{a}_d^b, \mathbf{a}_d^e$ by sparse coding [46] and the estimation of filter \mathbf{h} via least squares over multiple iterations. More specifically, as in a typical sparse coding problem formulation, estimates can be obtained by solving:

$$\begin{aligned} \min_{\mathbf{h}, \mathbf{z}_d, \mathbf{a}_d^b, \mathbf{a}_d^e} \quad & \sum_d \|\mathbf{U} [s_d - w_d - \mathbf{S}_d \mathbf{a}_d^b - \mathcal{T}(\mathbf{h}) \mathbf{z}_d] \\ & + \tilde{\mathbf{U}} [s_d - w_d + \mathbf{S}_d \mathbf{a}_d^e]\|_2^2 \\ & + \sum_d \lambda_1 (\mathbf{1}^T \mathbf{a}_d^e) + \lambda_2 (\mathbf{1}^T \mathbf{a}_d^b) + \lambda_3 (\mathbf{1}^T \mathbf{z}_d) \\ \text{subject to} \quad & \mathbf{a}_d^b \geq 0, \mathbf{a}_d^e \geq 0, \mathbf{z}_d \geq 0 \quad \forall d, \mathbf{h} \geq 0 \\ & \tilde{\mathbf{U}} [\mathbf{S}_d \mathbf{a}_d^b + \mathcal{T}(\mathbf{h}) \mathbf{z}_d] = \mathbf{0}, \mathbf{U} \mathbf{S}_d \mathbf{a}_d^e = \mathbf{0} \quad (11) \end{aligned}$$

The alternating algorithm is guaranteed to find only a locally optimal solution and it depends on the initialization [47]. Regularization constants are chosen such that $\lambda_1 \geq \lambda_2 \gg \lambda_3$ in order to force \mathbf{a}_d^b and \mathbf{a}_d^e to be more sparse than \mathbf{z}_d . Such a choice is made since sudden power attenuation due to direct occlusion of sun (\mathbf{a}_d^b) and the edge of the cloud effect (\mathbf{a}_d^e) are less likely events as compared to diffuse beam attenuation (\mathbf{z}_d). It was shown in [1] that the proposed model led to an excellent fit with the data. Even though the regression problem in [1] was solved in a completely deterministic fashion, such a model allowed the separation of the components and study a plausible stochastic model for them. This is explained in detail in the Section III.

B. Relationship Between Temperature and Solar PV Power

In [2], a Volterra model was derived to describe the relationship between temperature and solar irradiance. This model can be extended to solar PV power since the stochasticity in solar power is due to solar irradiance itself. Also, solar PV power is a linear transformation of solar irradiance [39]. Thus, the derivation in [2] can be directly applied to capture the dependency of PV power and temperatures.

Solar irradiance heats up landmass during the day that causes temperature increase in a certain geographical region. At night,

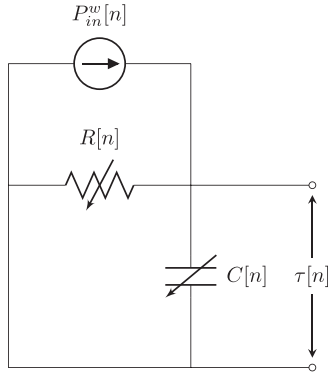


Fig. 4. Equivalent thermal circuit to describe the effect of solar PV power $w_d[n]$ on temperature $\tau[n]$.

heat absorbed by land is dissipated. This phenomenon is analogous to charging and discharging a capacitor in an RC circuit where voltage across the resistor can be thought of as temperature $\tau_d[n]$ at time instant n on day d . Current source is a function of solar irradiance and consequently solar PV power,

$$P_{in}^w[n] \approx \sum_{q=0}^{M_1-1} g[q]w_d[n-q]. \quad (12)$$

The resistor $R[n]$ and capacitor $C[n]$ themselves are quantities dependent on temperature [48] since they relate to thermal resistance of air and heat capacity¹ of the landmass respectively:

$$R[n] \approx \sum_{q=1}^{M_2} a[q]\tau[n-q], \quad C[n] = \frac{C_0 + \sum_{q=1}^{M_1} b[q]\tau[n-q]}{(\tau[n] - \tau[n-1])}. \quad (13)$$

Here $g[q]$, $a[q]$, $b[q]$ are coefficients describing $P_{in}^w[n]$, $R[n]$, $C[n]$ respectively at a certain location. Wind and other weather related phenomena are considered as noise in the model. Fig. 4 depicts the equivalent circuit describing the physics of the system. Writing the first order linear difference equation for the analogous circuit and performing further simplification as in [2] the model for temperature is:

$$\begin{aligned} \tau[n] \approx & \sum_{q=1}^{M_2} \gamma[q]\tau[n-q] + \sum_{k=0}^{M_1-1} \sum_{p=1}^{M_1-1-k} \alpha[k]\tau[n-p]\tau[n-p-k] \\ & + \sum_{k=0}^{M_1-1} \sum_{p=1}^{M_1-1-k} \beta[k]w_d[n-p+1]\tau[n-p-k] + \epsilon[n] \end{aligned} \quad (14)$$

where $\epsilon[n]$ is the modeling error. This model for temperature is interpreted as a Volterra series expansion [50], [51] up to second order with solar PV power as an input. In this way, the non-linearity in the system which is a result of temperature dependent thermal resistivity and heat capacity is captured. The parameters $\alpha[k]$, $\beta[k]$, $\gamma[q]$ can be estimated by solving a

¹In [49], heat capacity is defined as “the proportionality constant between the heat that the object absorbs or loses and the resulting temperature change of the object”. The heat absorbed by the surroundings to change the temperature from $\tau[n-1]$ to $\tau[n]$ can also be modeled as a linear relationship with temperature.

least-squares problem using training data of temperature and solar power output. Let a vector of parameters $\mathbf{x} \in \mathbb{R}^{M_2+2M_1}$ be defined as $\mathbf{x}^T = [\alpha^T \beta^T \gamma^T]$ where

$$[\alpha]_k = \alpha[k], [\beta]_k = \beta[k], [\gamma]_q = \gamma[q] \quad (15)$$

For time instant n , define

$$\begin{aligned} \mathbf{q}_n^T &= [\mathbf{r}_{\tau\tau}^T \mathbf{r}_{w\tau}^T \mathbf{r}_\tau^T], \text{ where} \\ [\mathbf{r}_{\tau\tau}]_k &= \sum_{p=1}^{M_1-k} \tau[n-p]\tau[n-p-k], \\ [\mathbf{r}_{w\tau}]_k &= \sum_{p=1}^{M_1-k} w_d[n-p+1]\tau[n-p-k], \\ k &= 0, \dots, M_1-1, \mathbf{r}_{\tau\tau} \in \mathbb{R}^{M_1}, \mathbf{r}_{w\tau} \in \mathbb{R}^{M_1} \\ [\mathbf{r}_\tau]_q &= \tau[n-q], q = 1, \dots, M_2, \mathbf{r}_\tau \in \mathbb{R}^{M_2}. \end{aligned}$$

Let $\mathbf{Q} \triangleq [\mathbf{q}_1 \mathbf{q}_2 \dots \mathbf{q}_N]$, $\mathbf{Q} \in \mathbb{R}^{N \times M_2+2M_1}$. Then, the estimate of $\hat{\mathbf{x}}$ is given by solving

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{Q}\mathbf{x} - \boldsymbol{\tau}\|_2^2, \quad (16)$$

where $\boldsymbol{\tau} \in \mathbb{R}^N$, $[\boldsymbol{\tau}]_n = \tau[n]$. If \mathbf{Q} is full rank, $\hat{\mathbf{x}}$ can be written in terms of pseudo-inverse as $\hat{\mathbf{x}} = \mathbf{Q}^\dagger \boldsymbol{\tau}$.

C. Spatio-Temporal Model for Solar Power and Temperature

After modeling solar PV power and temperature for a single location, it is desirable to exploit spatial correlation among solar PV panel installations that are in close proximity. For example, installations that belong to a single ZIP code which means that the geographical area is 6–7 km in diameter share similar power patterns as shown in [52]. Let us consider multiple residential solar PV installations within a ZIP code. It can be assumed that all the locations under consideration are seeing the same cloud cover which means that the sparse parameters \mathbf{a}_d^b , \mathbf{a}_d^e , \mathbf{z}_d are *approximately* same for all the locations on a day $d = 1, 2, \dots, D$. This assumption gives rise to the following model for solar PV power (in vector form) for locations indexed by $\hat{\ell} = 1, 2, \dots, L$ similar to (10):

$$\mathbf{w}_d^{\hat{\ell}} = \mathbf{s}_d^{\hat{\ell}} - \mathbf{U} [\mathbf{S}_d^{\hat{\ell}} \mathbf{a}_d^b + \mathcal{T}(\mathbf{h}^{\hat{\ell}}) \mathbf{z}_d] + \tilde{\mathbf{U}} \mathbf{S}_d^{\hat{\ell}} \mathbf{a}_d^e + \boldsymbol{\eta}_d^{\hat{\ell}} \quad (17)$$

The goal is to estimate the sparse parameters \mathbf{a}_d^b , \mathbf{a}_d^e , \mathbf{z}_d for each day d and filter $\mathbf{h}^{\hat{\ell}}$ for each location $\hat{\ell}$. Then, the regression problem can be written as in (11),

$$\begin{aligned} \min_{\mathbf{h}^{\hat{\ell}}, \forall \hat{\ell}, \mathbf{z}_d, \mathbf{a}_d^b, \mathbf{a}_d^e} \quad & \sum_{\hat{\ell}} \sum_d \left\| \mathbf{U} [\mathbf{s}_d^{\hat{\ell}} - \mathbf{w}_d^{\hat{\ell}} - \mathbf{S}_d^{\hat{\ell}} \mathbf{a}_d^b - \mathcal{T}(\mathbf{h}^{\hat{\ell}}) \mathbf{z}_d] \right. \\ & \left. + \tilde{\mathbf{U}} [\mathbf{s}_d^{\hat{\ell}} - \mathbf{w}_d^{\hat{\ell}} + \mathbf{S}_d^{\hat{\ell}} \mathbf{a}_d^e] \right\|_2^2 \\ & + \sum_d \lambda_1 (\mathbf{1}^T \mathbf{a}_d^e) + \lambda_2 (\mathbf{1}^T \mathbf{a}_d^b) + \lambda_3 (\mathbf{1}^T \mathbf{z}_d) \\ \text{subject to} \quad & \mathbf{a}_d^b \geq 0, \mathbf{a}_d^e \geq 0, \mathbf{z}_d \geq 0 \quad \forall d, \mathbf{h}^{\hat{\ell}} \geq 0, \forall \hat{\ell} \\ & \tilde{\mathbf{U}} [\mathbf{S}_d^{\hat{\ell}} \mathbf{a}_d^b + \mathcal{T}(\mathbf{h}^{\hat{\ell}}) \mathbf{z}_d] = \mathbf{0}, \mathbf{U} \mathbf{S}_d^{\hat{\ell}} \mathbf{a}_d^e = \mathbf{0} \end{aligned} \quad (18)$$

It can be solved similarly as the single location problem using alternating minimization. Outdoor temperature can be assumed

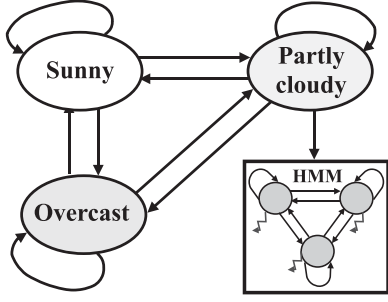


Fig. 5. Block diagram highlighting the proposed switching process between stochastic models.

to be almost constant for all locations in a ZIP code. As a proxy to solar irradiance, sum of solar power values from all locations $\hat{\ell}$ serves as the current source,

$$P_{in}^{w,\hat{\ell}}[n] \approx \sum_{q=0}^{M_1-1} g[q] \sum_{\hat{\ell}=1}^L w_d^{\hat{\ell}}[n-q]. \quad (19)$$

Thus, with different values for the parameters, temperature for an entire ZIP code, $\tilde{\tau}[n]$ can be modeled as in (14) with different parameters $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}$:

$$\begin{aligned} \tilde{\tau}[n] \approx & \sum_{q=1}^{M_2} \tilde{\gamma}[q] \tilde{\tau}[n-q] + \sum_{k=0}^{M_1-1} \sum_{p=1}^{M_1-k} \tilde{\alpha}[k] \tilde{\tau}[n-p] \tilde{\tau}[n-p-k] \\ & + \sum_{k=0}^{M_1-1} \sum_{p=1}^{M_1-k} \tilde{\beta}[k] \sum_{\hat{\ell}=1}^L w_d^{\hat{\ell}}[n-p+1] \tilde{\tau}[n-p-k] + \epsilon[n] \end{aligned} \quad (20)$$

III. STOCHASTIC MODELS FOR CLASSIFICATION AND FORECAST OF SOLAR POWER DATA

In spite of the fact that the switching regime of solar PV power is not intrinsically part of the model discussed in section II and that parameters were considered deterministic, as reported in [1], the results of the least-squares fit after solving (11) highlighted such nature for solar irradiation/power. The solar PV power produced in a period of time can be broadly classified as coming from *sunny*, *overcast* or *partly cloudy* models. The model switches between the three classes as shown in Fig. 5 due to weather changes. In this section, a stochastic model for each of the three classes is proposed. The first application of this model is to classify power data as coming from one of the three models to provide a probabilistic short-term forecast by assuming that the model persists.

1) *Stochastic Model for Sunny Period*: For sunny periods, it is hypothesized that the solar power is the deterministic solar power pattern i.e.,

$$w_d[n] = s_d[n] + \eta_d[n] \quad (21)$$

The modeling error is given by $\eta_d[n] \sim \mathcal{N}(0, \sigma_s^2) \forall n$ and $d \in \mathcal{S}$. The variance σ_s^2 is estimated using the error values after fitting the sunny day pattern from Section. II to the sunny days, $d \in \mathcal{S}$.

2) *Stochastic Model for Overcast Period*: During overcast periods, the attenuation of solar power is mostly from the diffuse beam component [39] which is why $z_d[n]$ accounts for the

relevant attenuation. Also, there is an average component in the overcast days for $z_d[n]$ that mimics a scaled version of sunny day pattern $s_d[n]$ [1]. Given the smoothness of this attenuation, the model postulated for the overcast period is:

$$w_d[n] \approx \alpha_d s_d[n] + \eta_n[k], \quad (22)$$

The parameter α_d is analogous to *clear sky index* defined as $w_d[n]/s_d[n]$ that many papers use to model solar PV power [53]. However, all the samples in the overcast period are used to estimate α_d unlike the determination of clear sky index. This leads to robustness with respect to noise. Noise $\eta_n[k]$ follows a truncated Gaussian distribution with mean 0 and noise variance σ_{oc}^2 for overcast period so as to limit the values of power $w_d[n]$ to lie in the interval $\mathcal{I}_{oc} \triangleq [0, s_d[n]]$. The distribution considered for $w_d[n]$ during overcast period is,

$$f_{oc}(w_d[n]) = \begin{cases} \frac{(2\pi\sigma_{oc}^2)^{-\frac{1}{2}} e^{-\frac{(w_d[n]-\alpha_d s_d[n])^2}{2\sigma_{oc}^2}}}{\Phi\left(\frac{s_d[n]-\alpha_d s_d[n]}{\sigma_{oc}}\right) - \Phi\left(\frac{-\alpha_d s_d[n]}{\sigma_{oc}}\right)}, & w_d[n] \in \mathcal{I}_{oc} \\ 0 & \text{otherwise} \end{cases} \quad (23)$$

where $\Phi(\cdot)$ denotes the CDF of a standard normal distribution.

3) *Stochastic Model for Partly Cloudy Period*: The model for the partly cloudy period includes all the three parameters in (10). However, a hidden Markov model (HMM) can capture the underlying on-off process that characterizes the sparse parameters in periods with fast moving clouds that cause sharp fluctuations in solar PV power. The observed solar PV power data $w_d[n]$ is modeled as coming from underlying hidden states that are Markovian in nature. Let the state/latent vector, \mathbf{q}_n be a coordinate vector that enumerates all the possible support combinations of sparse parameters for time instant n , i.e. of the vector

$$\mathbf{x}_n = [z_d[k-M+1] \quad \dots \quad z_d[n] \quad a_d^b[k] \quad a_d^e[k]]^T \quad (24)$$

The relationship between observations and latent state is:

$$w_d[n] = \alpha_d s_d[n] - \mathbf{P} \text{diag}(\Phi \mathbf{q}_n) \mathbf{x}_n \quad (25)$$

$$\mathbb{E}(\mathbf{q}_{n+1} | \mathbf{q}_n) = \mathbf{A}^T \mathbf{q}_n \quad (26)$$

$$\mathbf{P} = [h[M-1] \quad \dots \quad h[0] \quad s_d[n] \quad -s_d[n]] \quad (27)$$

Parameter α_d is included to model any constant attenuation over the time period considered unlike the other sparse parameters \mathbf{x}_n . Let the total number of states be N_s . Then, $\mathbf{A} \in \mathbb{R}^{N_s \times N_s}$ is the state transition matrix where $\mathbf{A}(i, j)$ is the probability of going from state i to state j . The state vector $\mathbf{q}_n \in \mathbb{R}^{(M+2)}$ is a binary vector taking values from the set of coordinate vectors $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{N_s}\}$ where $\mathbf{e}_i \in \mathcal{R}^{N_s}$ has a 1 at position i and zero elsewhere. Each column in matrix $\Phi \in \mathbb{R}^{(M+2) \times N_s}$ consists of a possible support of vector \mathbf{x}_n .

Certain assumptions are made to decrease the number of states. Firstly, the vector $[z_n[k-M+1] \quad \dots \quad z_d[n]]$ is restricted to have $\ell < M$ non-zero entries. Secondly, $a_d^e[k]$ cannot co-exist with the other parameters due to the fact that edge of cloud effect is indicative of the absence of attenuation. As a simplification, it is also assumed that direct beam and diffuse beam attenuations do not occur together which means that the total number of

states is:

$$N_s = \sum_{\tilde{\ell}=0}^{\ell} \binom{M}{\tilde{\ell}} + 2 \quad (28)$$

Noise term can be ignored in the observation (25) since it is small in amplitude relative to the ‘noisy’ nature of the solar power data that is caused by the fast movement of clouds. The simplest case of choosing $\ell = 1$ and having $N_s = M + 3$ states is considered. All non-zero coefficients in \mathbf{x}_n are hypothesized to come from independent exponential distributions with different parameters since they have different levels of sparsity. While in state i a certain $w_d[n]$ is observed:

$$w_d[n] = \begin{cases} \alpha_d s_d[n], & i = 1 \\ \alpha_d s_d[n] - h[i-2]z_d[n-i+2], & i = 2, \dots, M+1, \\ \alpha_d s_d[n] - s_d[n]a_d^b[n], & i = M+2 \\ \alpha_d s_d[n] + s_d[n]a_d^e[n], & i = N_s \end{cases} \quad (29)$$

where the following statistical model is postulated for the parameters

$$z_d[n] \sim \exp(\lambda_z), a_d^b[n] \sim \exp(\lambda_{a^b}), a_d^e[n] \sim \exp(\lambda_{a^e}) \quad (30)$$

The corresponding conditional probability distribution given the state i is denoted as $f_i(w_d[n]) \triangleq f(w_d[n]|\mathbf{q}_n = \mathbf{e}_i)$ and is equal to,

$$f_i(w_d[n]) = \begin{cases} \delta(\alpha_d s_d[n] - w_d[n]), & i = 1 \\ \frac{C_i \lambda_z}{h[i-2]} e^{-\frac{\lambda_z(\alpha_d s_d[n] - w_d[n])}{h[i-2]}}, & i = 2, \dots, M+1 \\ \frac{C_i \lambda_{a^b}}{s_d[n]} e^{-\frac{\lambda_{a^b}}{s_d[n]}(\alpha_d s_d[n] - w_d[n])}, & i = M+2 \\ \frac{\lambda_{a^e}}{s_d[n]} e^{-\frac{\lambda_{a^e}}{s_d[n]}(w_d[n] - \alpha_d s_d[n])}, & i = N_s \end{cases} \quad (31)$$

where C_i is the normalizing constant for the probability distribution given by

$$C_i^{-1} = \begin{cases} 1 - e^{-\lambda_z s_d[n]/h[i-2]}, & i = 2, 3, \dots, M+1 \\ 1 - e^{-\lambda_{a^b}}, & i = M+2 \end{cases} \quad (32)$$

The normalization is done so that $w_d[n] \in [0, s_d[n]]$.

Learning the parameters of the HMM: Viterbi training

The models for sunny and the overcast periods are such that the only the mean of the process can be predicted in both cases. The mean during sunny period, $s_d[n]$ is assumed to be known. The mean for overcast periods can be determined by estimating the scaling α_d . Hence, the problem of learning the stochastic parameters of the model to perform predictions is non-trivial only during partly cloudy periods. To do so, it is assumed that the values of the parameters $\lambda_z, \lambda_{a^b}, \lambda_{a^e}$ of conditional probability distributions are known. It was seen that the algorithm is not very sensitive to the exact values of these parameters as long as they follow $\lambda_z \leq \lambda_{a^b} \leq \lambda_{a^e}$ which is consistent with the regularization performed in the regression problem. The probability of starting from a state i denoted by $\pi_i = 1/N_s$ is also assumed to be known. In order to learn the the state transition matrix \mathbf{A} , Viterbi training [54] or segmental

k-means [55] approach was adopted. Let

$$\xi = \{\mathbf{A}(i, j) | i, j \in \{1, \dots, N_s\}\} \quad (33)$$

be the set of unknown parameters to be estimated. Let

$$\mathbf{Q} = [\mathbf{q}_1 \dots \mathbf{q}_N] \text{ and } \mathbf{w} = [w_d[1] \dots w_d[N]]$$

denote the sequence of hidden states and solar power observations respectively. In the Viterbi training algorithm, instead of maximizing the likelihood over all possible state sequences, \mathbf{Q} , the likelihood is maximized only over the most probable state sequence to find the estimates of parameters in ξ . The algorithm starts with an initial estimate for all the unknown parameters $\xi_0 = \{\mathbf{A}_0(i, j) | i, j \in \{1, \dots, N_s\}\}$ and performs this maximization iteratively [55],

$$\hat{\xi}_k = \arg \max_{\xi} \left(\max_{\mathbf{Q}} f(\mathbf{w}, \mathbf{Q} | \hat{\xi}_{k-1}) \right) \quad (34)$$

where k is the iteration number and

$$f(\mathbf{w}, \mathbf{Q} | \xi) = f(\mathbf{q}_1) \prod_{n=1}^N f(w_d[n] | \mathbf{q}_n, \xi) \prod_{n=1}^{N-1} f(\mathbf{q}_{n+1} | \mathbf{q}_n, \xi) \quad (35)$$

The inner maximization is performed by using Viterbi algorithm [56]. As a result of this maximization,

$$\hat{\mathbf{Q}}_k = \arg \max_{\mathbf{Q}} f(\mathbf{w}, \mathbf{Q} | \hat{\xi}_{k-1}) = \hat{\mathbf{q}}_1^k, \hat{\mathbf{q}}_2^k, \dots, \hat{\mathbf{q}}_N^k, \quad (36)$$

is the most likely state sequence at iteration k which best describes the observed data. Then maximum-likelihood (ML) estimates, $\hat{\xi}_k$, are

$$\hat{\xi}_k = \arg \max_{\xi} f(\mathbf{w}, \hat{\mathbf{Q}}_k | \hat{\xi}_{k-1}) \quad (37)$$

Maximizing $\log f(\mathbf{w}, \hat{\mathbf{Q}}_k | \hat{\xi}_{k-1})$ with respect to $\mathbf{A}_k(i, j)$ under the constraint that \mathbf{A}_k is a column-stochastic matrix gives

$$\hat{\mathbf{A}}_k(i, j) = \hat{N}_{ij} / \sum_{j=1}^{N_s} \hat{N}_{ij} \quad (38)$$

where \hat{N}_{ij} is the number of times the transition from state i to state j occurs within the state sequence $\hat{\mathbf{Q}}_k$. Following from Section III-A3 wherein the number of active coefficients at time instant n in \mathbf{x}_n is restricted to 1, only a limited number of transitions from state i are possible. Also, since z_n is the input to a filter with memory M , it means that $M - 1$ components need to be retained and shifted while a new one comes in. All of the above reasons give the state transition matrix \mathbf{A} a sparse and specific structure as shown in Fig. 6 which is forced on \mathbf{A}_0 during the initialization. As a result, only $(M - 1) + 4 \times 3$ entries of the matrix need to be estimated when $\ell = 1$ instead of $(N_s)^2$.

IV. REGIME CLASSIFICATION FOR PREDICTION

The premise for prediction is the persistence in the weather condition for the time horizon over which a forecast of solar power is provided. Therefore, the proposed prediction algorithm has two steps:

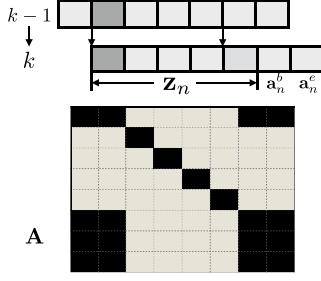


Fig. 6. The specific way in which state transition from time instant $k - 1$ to k takes places determines the structure of the state transition matrix \mathbf{A} .

- Classification of the solar power from a given period as coming from one of the three classes of models: *sunny*, *overcast*, *partly cloudy*
- Assuming that this weather condition *persists* for the duration of the prediction horizon and provide with a point forecast corresponding to the class decided in the classification step.

Such a scheme captures the inherent switching behavior that solar power exhibits, i.e. that of going from one model to another, while persisting for a certain duration in each of these. Note that the classification step can be skipped if prior knowledge in the form of weather prediction is available.

A. Classification Algorithm for Solar Power

The classification algorithm uses the stochastic models for the solar power data as detailed in III. The underlying principle behind classification leverages the fact that it is easier to classify a segment of data as coming from *sunny* or *overcast* regime by computing the negative log-likelihood since they both follow Gaussian distributions with different mean and variance. If the negative log-likelihood is low, then the means are compared to distinguish between the two classes. If neither of the negative log-likelihoods is lower than a certain threshold, then it is classified to be *partly cloudy*. The thresholds for classification are determined empirically using training data. Let $\mathbf{w}_n^{-M_1} \triangleq [w_d[n - M_1] \dots w_d[n]]$ be the M_1 solar power samples that have to be classified into one of the three regimes. Here, M_1 is the length of the filter $\beta[q]$ in (14). The algorithm begins by computing the error made by assuming *sunny* period and comparing with some threshold. If

$$\sum_{k=n-M_1}^n (w_d[k] - s_d[k])^2 \leq \tau_{\text{sunny}}, \quad (39)$$

then it is classified as a *sunny* period, where the threshold is chosen so that $\tau_{\text{sunny}} = \mu\sigma_s$, $\mu > 1$. If that is not the case, then the hypothesis that it is *overcast* is tested by calculating

$$\hat{\alpha}_n = \arg \min_{\alpha_n} \sum_{k=n-M_1}^n (w_d[k] - \alpha_n s_d[k])^2 \quad (40)$$

and checking if both the value of $\hat{\alpha}_n$ as well as the estimation error is less than some threshold,

$$\hat{\alpha}_n \leq \tau_\alpha, \text{ and } \sum_{k=n-M_1}^n (w_d[k] - \hat{\alpha}_n s_d[k])^2 \leq \tau_{\text{overcast}}. \quad (41)$$

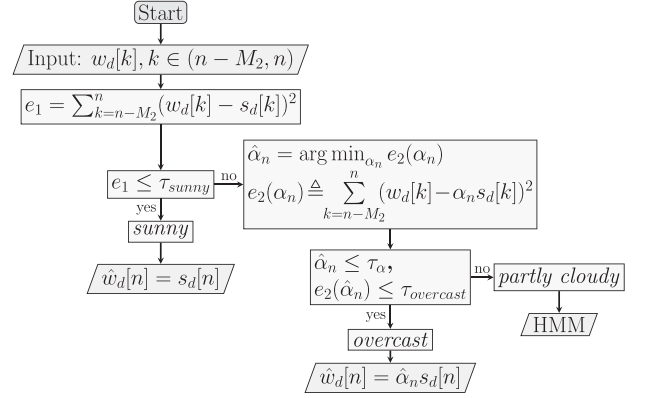


Fig. 7. Flowchart of the solar power prediction algorithm.

If these conditions are also not satisfied, then the regime is classified as *partly cloudy*. If the decision is in favor of *partly cloudy*, the most likely state sequence \mathbf{Q}_{M_1} that generated the power observations $\mathbf{w}_n^{-M_1}$ is determined using the Viterbi algorithm with state transition matrix $\hat{\mathbf{A}}$. Fig. 7 shows the flowchart describing the classification algorithm.

B. Prediction for Each Class of Models

Based on the classification results on $\mathbf{w}_n^{-M_1}$, a probabilistic solar power forecast for m time steps ahead is provided. It is an m dimensional probability density function, $f(\mathbf{w}_n^{+m} | \mathbf{w}_n^{-M_1})$ where $\mathbf{w}_n^{+m} \triangleq [w_d[n + 1] \dots w_d[n + m]]$.

1) *Prediction Using Sunny Model*: When the classification algorithm chooses the hypothesis that the present solar power data is from a *sunny* model, then:

$$f_s(\mathbf{w}_n^{+m} | \mathbf{w}_n^{-M_1}) = (2\pi\sigma_s)^{-\frac{m}{2}} e^{-\|\mathbf{w}_n^{+m} - \mathbf{s}_n^{+m}\|_2^2 / 2\sigma_s^2} \quad (42)$$

where $\mathbf{s}_n^{+m} \triangleq [s_d[n + 1] \dots s_d[n + m]]$. Note that the deterministic sequence of the sunny day solar power pattern is known beforehand and it is updated at a very slow pace on days that are classified as being sunny, to adjust for seasonal variations.

2) *Prediction using Overcast Model*: If it is overcast, the predicted distribution is

$$f_{oc}(\mathbf{w}_n^{+m} | \mathbf{w}_n^{-M_1}) = \frac{(2\pi\sigma_{oc})^{-\frac{m}{2}} e^{-\|\mathbf{w}_n^{+m} - \hat{\alpha}_n \mathbf{s}_n^{+m}\|_2^2 / 2\sigma_{oc}^2}}{\prod_{i=1}^m \left(\Phi\left(\frac{s_d[n+i] - \hat{\alpha}_n s_d[n+i]}{\sigma_{oc}}\right) - \Phi\left(\frac{-\hat{\alpha}_n s_d[n+i]}{\sigma_{oc}}\right) \right)}, \quad (43)$$

when $w_d[n + i] \in [0, s_d[n + i]]$, $\forall i$ and $\hat{\alpha}_n$ is estimated using (40).

3) *Prediction using Partly Cloudy Model*: Since solar PV power on a partly cloudy day has an underlying Markov model, the estimated state transition matrix $\hat{\mathbf{A}}$ and state at time n is used to compute a forecast distribution for m steps ahead as

$$f_{pc}(\mathbf{w}_n^{+m} | \mathbf{w}_n^{-M_1}) = \prod_{\ell=1}^m \sum_{j=1}^{N_s} \mathbf{A}^\ell(\hat{i}, j) f_j(w_d[n]) \quad (44)$$

where \hat{i} is the most probable state at time n . This is estimated by Viterbi algorithm [57, see Section III.B] using solar power values $\mathbf{w}_n^{-M_1}$. Therefore, the predictive distribution of solar PV

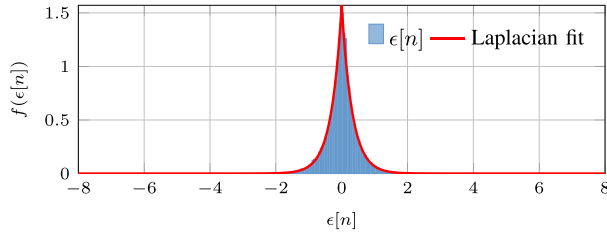


Fig. 8. Laplacian fit for $\epsilon[n]$ with $\mu = 3.1422$.

power can be summarized as

$$f(\mathbf{w}_n^{+m} | \mathbf{w}_n^{-M_1}) = \begin{cases} f_s(\mathbf{w}_n^{+m} | \mathbf{w}_n^{-M_1}), & \text{sunny} \\ f_{oc}(\mathbf{w}_n^{+m} | \mathbf{w}_n^{-M_1}), & \text{overcast} \\ f_{pc}(\mathbf{w}_n^{+m} | \mathbf{w}_n^{-M_1}), & \text{partly cloudy} \end{cases} \quad (45)$$

where the expressions of $f_s(\mathbf{w}_n^{+m} | \mathbf{w}_n^{-M_1})$, $f_{oc}(\mathbf{w}_n^{+m} | \mathbf{w}_n^{-M_1})$ and $f_{pc}(\mathbf{w}_n^{+m} | \mathbf{w}_n^{-M_1})$ are in (42), (43) and (44) respectively.

V. JOINT PROBABILISTIC FORECASTS

Since the relationship between temperature and solar PV power was established in section II, joint description of future values, i.e. probabilistic forecasts of temperature and solar power can be provided.

A. Conditional Distribution of Temperature Given Solar Power

After solving the problem in (16) in Section II-B, residual $\epsilon[n]$ from (14) was analyzed and found that it can be approximated to be drawn from a Laplacian distribution with parameter μ . This is illustrated in Fig. 8. To write the conditional distribution of temperature at time n given all the past values of temperature and solar PV power, define the vectors

$$\begin{aligned} \boldsymbol{\tau}_n^{-M_2} &= [\tau[n] \quad \tau[n-1] \quad \dots \quad \tau[n-M_2]]^T \\ \mathbf{s}_n^{-M_1} &= [s_d[n] \quad s_d[n-1] \quad \dots \quad s_d[n-M_1]]^T \end{aligned}$$

Then the conditional distribution from (14) is

$$f(\tau[n] | \mathbf{w}_n^{-M_1}, \boldsymbol{\tau}_n^{-M_2}) = \frac{\mu}{2} e^{-\mu |\tau[n] - g(\mathbf{w}_n^{-M_1}, \boldsymbol{\tau}_n^{-M_2})|}, \quad (46)$$

$$\begin{aligned} g(\mathbf{w}_n^{-M_1}, \boldsymbol{\tau}_n^{-M_2}) &= \sum_{q=1}^{M_2} \gamma[q] \tau[n-q] \\ &+ \sum_{k=0}^{M_1-1} \sum_{p=1}^{M_1-k} \alpha[k] \tau[n-p] \tau[n-p-k] \\ &+ \sum_{k=0}^{M_1-1} \sum_{p=1}^{M_1-k} \beta[k] w_d[n-p+1] \tau[n-p-k] \end{aligned} \quad (47)$$

B. Joint Distribution and Probabilistic Forecast

From the conditional probability formula, the joint distribution of $\tau[n]$, $w_d[n]$ at time n given $\mathbf{w}_{n-1}^{-M_1}$ and $\boldsymbol{\tau}_{n-1}^{-M_2}$ is

$$f(\tau[n], w[n] | \mathbf{w}_{n-1}^{-M_1}, \boldsymbol{\tau}_{n-1}^{-M_2}) = \quad (48)$$

$$f(\tau[n] | w[n], \mathbf{w}_{n-1}^{-M_1}, \boldsymbol{\tau}_{n-1}^{-M_2}) f(w_d[n]) \quad (49)$$

Let the vectors of m steps ahead forecasts of $\tau[n]$ be

$$\boldsymbol{\tau}_n^{+m} = [\tau[n+m] \quad \dots \quad \tau[n+1]]^T, \quad \boldsymbol{\tau}_n^{+0} = \boldsymbol{\tau}_n^{-M_2}$$

Probabilistic forecasts i.e. the joint distribution which following (48) and the probability chain rule can be written as,

$$\begin{aligned} f(\mathbf{w}_n^{+m}, \boldsymbol{\tau}_n^{+m} | \mathbf{w}_n^{-M_1}, \boldsymbol{\tau}_n^{-M_2}) &= \\ \prod_{\ell=1}^m f(\tau[n+\ell] | \mathbf{w}_n^{+m}, \boldsymbol{\tau}_n^{+\ell-1}, \mathbf{w}_n^{-M_1}, \boldsymbol{\tau}_n^{-M_2}) &f(\mathbf{w}_n^{+m} | \mathbf{w}_n^{-M_1}) \end{aligned} \quad (50)$$

Each component in the product over ℓ is the conditional distribution of temperature $\tau[n+\ell]$ given current and all past values of solar power, $\mathbf{w}_n^{+\ell}$, $\mathbf{w}_n^{-M_1}$ and past values of temperature $\boldsymbol{\tau}_n^{+\ell-1}$, $\boldsymbol{\tau}_n^{-M_2}$. This can be written using (46). The probabilistic forecast of solar PV power for m steps ahead, $f(\mathbf{w}_n^{+m} | \mathbf{w}_n^{-M_1})$ is given by (45).

Spatio-temporal model for solar power and temperature:

Since the stochastic parameters for solar power remain the same for a ZIP code based on the modeling assumption, installations within a ZIP code have solar power from the same regime for a certain period of time. Given the regime, the predicted distributions at different locations come from the same family of distributions with the parameters of the distribution different due to solar sunny day power pattern $s_d^{\hat{\ell}}$ and the filter $\mathbf{h}^{\hat{\ell}}$ which are not the same across locations. Analogous to the single location case, let $\mathbf{w}_{n,\hat{\ell}}^{-M_1}$, $\mathbf{w}_{n,\hat{\ell}}^{+m}$ be the power at $\hat{\ell}$ -th location with sunny day solar power pattern $\mathbf{s}_{n,\hat{\ell}}^{-M_1}$. Then, the predictive distribution of solar PV power for all locations can be written as

$$\begin{aligned} f\left(\left\{\mathbf{w}_{n,\hat{\ell}}^{+m}\right\}_{\hat{\ell}=1}^L \mid \left\{\mathbf{w}_{n,\hat{\ell}}^{-M_1}\right\}_{\hat{\ell}=1}^L\right) &= \\ \begin{cases} \prod_{\hat{\ell}=1}^L f_s(\mathbf{w}_{n,\hat{\ell}}^{+m} | \mathbf{w}_{n,\hat{\ell}}^{-M_1}), & \text{sunny} \\ \prod_{\hat{\ell}=1}^L f_{oc}(\mathbf{w}_{n,\hat{\ell}}^{+m} | \mathbf{w}_{n,\hat{\ell}}^{-M_1}), & \text{overcast} \\ \prod_{\hat{\ell}=1}^L f_{pc}(\mathbf{w}_{n,\hat{\ell}}^{+m} | \mathbf{w}_{n,\hat{\ell}}^{-M_1}), & \text{partly cloudy} \end{cases} \end{aligned} \quad (51)$$

The conditional distribution of temperature from (20) given solar power at all locations is

$$f(\tau[n] | \left\{\mathbf{w}_{n,\hat{\ell}}^{+m}\right\}_{\hat{\ell}=1}^L, \tilde{\boldsymbol{\tau}}_{n-1}^{-M_2}) = \frac{\tilde{\mu}}{2} e^{-\tilde{\mu} |\tau[n] - g(\left\{\mathbf{w}_{n,\hat{\ell}}^{+m}\right\}_{\hat{\ell}=1}^L, \tilde{\boldsymbol{\tau}}_{n-1}^{-M_2})|},$$

$$\text{where } g\left(\left\{\mathbf{w}_{n,\hat{\ell}}^{+m}\right\}_{\hat{\ell}=1}^L, \tilde{\boldsymbol{\tau}}_{n-1}^{-M_2}\right)$$

$$\begin{aligned} &= \sum_{q=1}^{M_2} \tilde{\gamma}[q] \tilde{\tau}[n-q] + \sum_{k=0}^{M_1-1} \sum_{p=1}^{M_1-k} \tilde{\alpha}[k] \tilde{\tau}[n-p] \tilde{\tau}[n-p-k] \\ &+ \sum_{k=0}^{M_1-1} \sum_{p=1}^{M_1-k} \tilde{\beta}[k] \sum_{\hat{\ell}=1}^L w_d^{\hat{\ell}}[n-p+1] \tilde{\tau}[n-p-k] \end{aligned} \quad (52)$$

The joint probabilistic distribution for m steps ahead can be written as,

$$\begin{aligned} & f\left(\left\{w_{n,\hat{\ell}}^{+m}\right\}_{\hat{\ell}=1}^L, \tilde{\tau}_n^{+m} \middle| \left\{w_{n,\hat{\ell}}^{-M_1}\right\}_{\hat{\ell}=1}^L, \tilde{\tau}_n^{-M_2}\right) \\ &= \prod_{\ell=1}^m f\left(\tilde{\tau}[n+\ell], \left\{w_{n,\hat{\ell}}^{+m}\right\}_{\hat{\ell}=1}^L, \tilde{\tau}_n^{+\ell-1}, \left\{w_{n,\hat{\ell}}^{-M_1}\right\}_{\hat{\ell}=1}^L, \tilde{\tau}_n^{-M_2}\right) \\ & f\left(\left\{w_{n,\hat{\ell}}^{+m}\right\}_{\hat{\ell}=1}^L \middle| \left\{w_{n,\hat{\ell}}^{-M_1}\right\}_{\hat{\ell}=1}^L\right) \end{aligned} \quad (53)$$

VI. NUMERICAL RESULTS

In this section, firstly the results for solar PV power forecasting alone are presented. A dataset of solar PV power from a rooftop installation in California is used for this purpose. Then, the results for joint forecasting are presented using datasets from multiple weather stations in California. In this dataset, solar irradiance is available and not solar power. Since the source of stochasticity remains the same, the effect of the algorithm does not change.

A. Solar PV Power Forecasting

The dataset used for this work was acquired from a rooftop panel installation in Antioch, California and was provided by SolarCity. This dataset was also used in authors' prior work in [1]. The format of this solar power data consisted of current (in A), voltage measurements (in V) and timestamps (in Hours) at the inverter approximately every 15 minutes recorded for a duration of two years. Each panel had a rating of 170 W and there were a total of 22 panels. Therefore, the nameplate rating of all panels combined was $170 \times 22 = 3740$ W.

1) *Algorithm Settings*: To solve (11), the alternating minimization algorithm is initialized with the filter h being a scaled Hamming window of length M , $h[q] = g \times (0.54 - 0.46 \cos(2\pi q/(M-1)))$. To address the scale ambiguity inherent in blind deconvolution problems, the scale g is chosen such that $h[q]$ and $s_d[n]$ have similar amplitudes. Prediction horizon was $m = 12$ i.e. 3 hours and a filter, $h[n]$, of length $M = 5$ was used for *partly cloudy* conditions. The algorithm started with $M_1 = 6$ samples (1.5 hour) for each day and predicted for the next m samples. Then, the window was moved by one sample.

2) *Metrics for Evaluation*: The results are presented using probabilistic forecast metrics. In the probabilistic setting at each m -step prediction there is a cumulative distribution function (CDF) $F_m(w[n]) \triangleq F(w[n+m] | w_n^{-M_1})$ instead of a point forecast. The metrics used for evaluation are continuous rank probability score (CRPS) [58], reliability metric and score [59]. CRPS is defined for each m -step prediction as an average over all the samples,

$$\text{CRPS}(m) = (\tilde{P})^{-1} \sum_{n,d} \int_0^\infty (F_m(y) - u(y - w_d[n]))^2 dy \quad (54)$$

where $u(\cdot)$ is the Heaviside step function and \tilde{P} is the total number of samples. The CRPS evaluates to mean absolute error (MAE) when it is a point forecast.

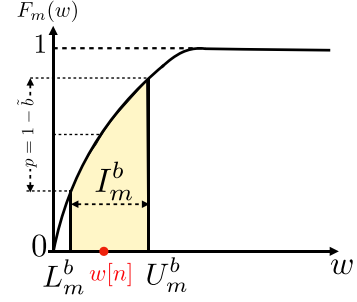


Fig. 9. Figure highlighting predictive density and the corresponding probability interval.

Reliability of a probabilistic forecasting method is a useful metric in understanding the proximity of the estimated CDF to the actual CDF of the data. Let a probability interval (PI), I_m^b , be defined such that the interval covers the observed value $w_d[n]$ with probability $(1 - b)$. Then, to calculate the reliability, define

$$R_b(m) = (\tilde{P})^{-1} \sum_{n,d} \mathbb{I}_{\{w_d[n] \in I_m^b\}} \quad (55)$$

as the estimated probability of coverage where $\mathbb{I}_{\{\cdot\}}$ is an indicator function with value 1 if the observed sample belongs to the probability interval. Now, the probabilistic forecast is more reliable if the quantity ΔR_b is small,

$$\Delta R_b(m) \triangleq R_b(m) - (1 - b). \quad (56)$$

Another metric used for evaluation is the interval score [60]. This metric is helpful in determining the sharpness of the forecast probability interval by imposing a penalty when an observation is outside the interval, by a value proportional to the size of the interval. If the upper and lower bounds of the PI are denoted as U_m^b and L_m^b respectively then interval score is defined as

$$\text{Score}_b^m[k] = \begin{cases} D_m^b - 4(L_m^b - w_d[n]), & w_d[n] < L_m^b \\ D_m^b, & w_d[n] \in I_m^b \\ D_m^b - 4(w_d[n] - U_m^b), & w_d[n] > U_m^b \end{cases} \quad (57)$$

$$\text{where } D_m^b \triangleq -2b(U_m^b - L_m^b) \quad (58)$$

The average interval score is,

$$\text{Score}_b(m) = (\tilde{P})^{-1} \sum_{n,d} \text{Score}_b^m[k] \quad (59)$$

Lower values of this score indicate sharper and more reliable forecasts. Fig. 9 illustrates the predictive density $F_m(w)$ and the corresponding probability interval I_m^b . Performance of the prediction methods is also analyzed using average reliability and score defined as

$$R_b^{\text{avg}} = \sum_{m=1}^X R_b(m)/X, \quad \text{Score}_b^{\text{avg}} = \sum_{m=1}^X \text{Score}_b(m)/X \quad (60)$$

As representative examples, Fig. 10, 11 and 12 show the actual and predicted power for different days with a multitude of weather conditions. This predicted power is one-step prediction,

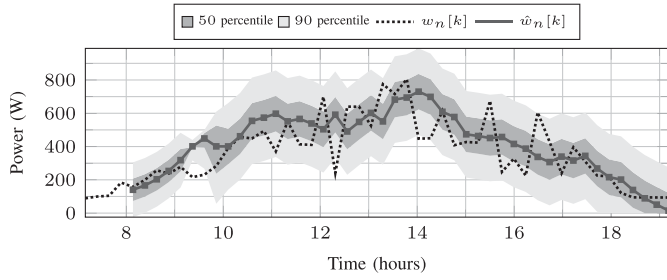


Fig. 10. Plot of actual and predicted value with one-step prediction for day that is *overcast*.

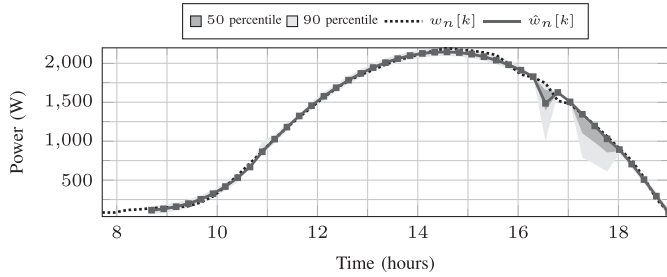


Fig. 11. Plot of actual and predicted value with one-step prediction for day that is *sunny* day.

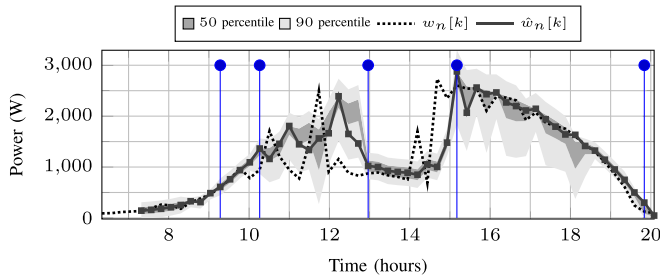


Fig. 12. Plot of actual and predicted value with one-step prediction for day with variety of weather conditions. Time instants at which algorithm detected regime change is shown in blue.

$m = 1$. For days that are entirely *overcast* or *sunny*, predictions have little error as can be seen in Fig. 10 and 11. These results highlight that the stochasticity of power in both these regimes is minimal leading to better predictions if the weather condition persists. However, there is higher error whenever there is a change in regime, for example going from *partly cloudy* condition to *overcast* around 12 PM as seen in Fig. 12. It can be attributed to the delay in detecting the change in model. This uncertainty cannot be avoided in days with sudden change in weather unless there is some additional information in the form of cloud motion information or accurate weather forecasts. To summarize, prediction during *partly cloudy* conditions is prone to larger errors than during *overcast* or *sunny*. This is in accordance with the associated uncertainty in solar PV power for those periods.

3) Comparison with Reference Methods:

Probabilistic smart persistence: It is assumed that the mean of predictive density for the next m steps is given as the continued

fraction of clear sky component at the current time step,

$$\hat{w}_d^{per}[n+m] = s_d[n+m] (w_d[n]/s_d[n]), m = 1, \dots, \chi \quad (61)$$

Variance is given by the sample variance of the power values in $w_n^{-M_1}$. The distribution is assumed to be Gaussian.

AR model: The power is expressed as the sum of a clear sky component and a stochastic component, and it is assumed that the stochastic component follows an autoregressive model:

$$w_d^{AR}[n+m] = s_d[n+m] - x_d^{AR}[n+m],$$

$$x_d^{AR}[n+m] = \sum_{i=1}^{M_{AR}} a[i] x_d^{AR}[n+m-i] + \epsilon_{AR}[n+m] \quad (62)$$

AR model with regime switching: It is assumed that each of the classes *sunny*, *partly cloudy* and *overcast* have stochastic components with different coefficients for the AR model:

$$x_d^{AR}[n+m] = \begin{cases} \sum_{i=1}^{M_{ARs}} a_s[i] x_d^{AR}[n+m-i] + \epsilon_{ARs}[n+m], & \text{sunny} \\ \sum_{i=1}^{M_{ARpc}} a_{pc}[i] x_d^{AR}[n+m-i] + \epsilon_{ARpc}[n+m], & \text{partly cloudy} \\ \sum_{i=1}^{M_{ARoc}} a_{oc}[i] x_d^{AR}[n+m-i] + \epsilon_{ARoc}[n+m], & \text{overcast} \end{cases} \quad (63)$$

Classification into different regimes for the AR model is done using the proposed approach in Subsection IV-A. In essence, it is the closest method to the proposed one by design.

In both the AR model and the regime switching AR models, the point forecast value is the mean and the variance of Gaussian noise, ϵ_{AR} is estimated along with the coefficients.

Quantile regression (QR) Linear quantile regression (QR) [61] is used. b^{th} quantile of solar power, $Q_m(b)$ for predictive time step m , where $b = 0.01, 0.02, \dots, 0.99$ is expressed as:

$$Q_m(b) = x^{QR} s_d[n+m], x^{QR} = \beta_{m,b}^T \mathbf{x}, m = 1, \dots, \chi \quad (64)$$

$$[\mathbf{x}]_i = w_d[n - M_1 + i] / s_d[n - M_1 + i], i = 1, \dots, M_1 \quad (65)$$

where $\beta_{m,b}$ are vectors of parameters that were estimated using the procedure in [61].

4) **Results: Probabilistic Forecast of Solar Power:** All the simulations were performed using one year of training data and one year of testing data for validation. Most of the training for estimation of parameters of HMM is done apriori making computational time of the proposed method very short since the Viterbi algorithm, which is efficient [56] was used. The computational time specifically depends on the acquisition time of samples in a real-time setting.

Comparison with CRPS as the metric is shown in Fig. 13. The AR models beat the smart persistence. For the QR method, CRPS cannot be calculated since the predictive density is not continuous. CRPS is inversely proportional to the parameters² [62] which makes larger values of parameters better.

²Closed form expressions for CRPS in (54) can be computed. For Gaussian densities, it is inversely proportional to σ . For exponential densities, it is inversely proportional to λ .

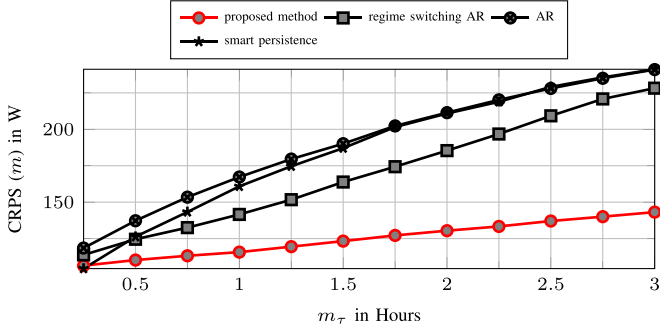


Fig. 13. CRPS for m -step prediction using proposed model and other reference models.

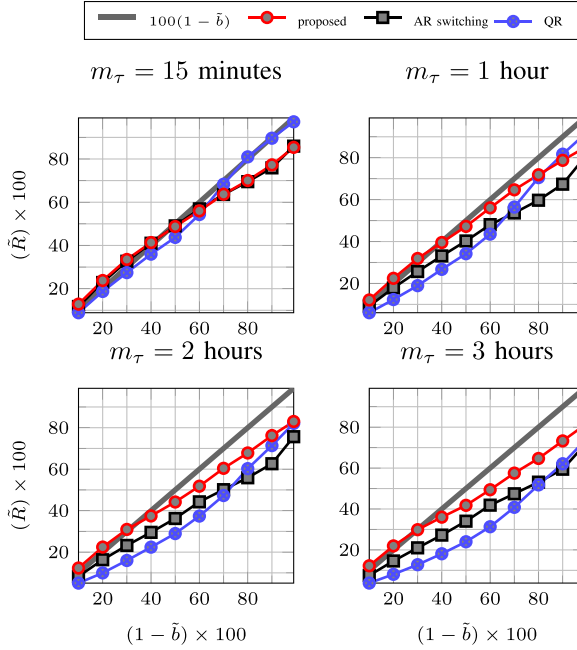


Fig. 14. Reliability prediction horizons using proposed model and few better performing reference models.

Fig. 14 shows the reliability metric for prediction horizons $m_\tau = 0.25, 1, 2, 3$ hours only for certain better performing methods whereas Fig. 15 shows the average reliability metric, R_b^{avg} and comparison with the curve $1 - b$ for all benchmarks. Both figures are indicative of the probability of the actual occurrences of power being well covered by the predicted intervals. The closer the metric to the curve $1 - b$, the more reliable the method. From Fig. 14, it is seen that at $m_\tau = 0.25$, quantile regression performs the best. However, as the prediction horizon increases, the performance of QR method deteriorates. This can be attributed to fewer samples of data available for training as the prediction horizon increases. AR method with regime-switching performs the closest to the proposed method. It even matches the proposed method for lower quantiles across all prediction horizons but does not fare well in higher quantiles. The key takeaway is that the proposed method consistently (in all prediction horizons) does better than the benchmark methods

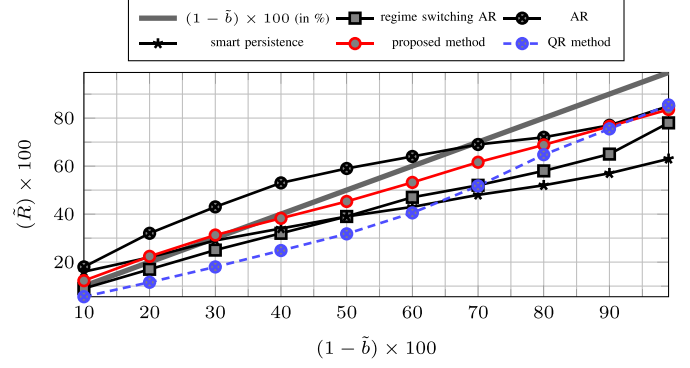


Fig. 15. Average reliability (over all prediction horizons) using proposed model and other reference models.

TABLE I
TABLE WITH AVERAGE SCORE NORMALIZED BY THE NAMEPLATE CAPACITY

$1 - b$	Proposed	QR	AR (switching)	Persistence	AR
0.1	-0.214	-0.216	-0.229	-0.242	-0.223
0.2	-0.202	-0.209	-0.218	-0.232	-0.211
0.3	-0.196	-0.204	-0.212	-0.228	-0.208
0.4	-0.195	-0.202	-0.212	-0.231	-0.212
0.5	-0.199	-0.204	-0.218	-0.24	-0.224
0.6	-0.210	-0.215	-0.23	-0.256	-0.245
0.7	-0.232	-0.253	-0.25	-0.282	-0.276
0.8	-0.268	-0.349	-0.281	-0.32	-0.322
0.9	-0.334	-0.511	-0.331	-0.383	-0.394
0.99	-0.482	-0.994	-0.455	-0.503	-0.566

considered. The trend of under performance in higher quantiles by the proposed method can be attributed to the choice of the values of λ_z , λ_a^b , λ_a^e and to classification error in the regime i.e. error due to assumption of persistence in regime of *sunny, partly cloudy, overcast*.

Using the same set of parameters, Table I shows the average interval score, $\text{Score}_b^{\text{avg}}$ normalized by the nameplate capacity. In terms of interval score, the proposed method outperforms all the other methods considered. This is indicative of the fact that the proposed method of forecasting is sharp [58].

B. Joint Probabilistic Forecast

Experiments were run on a dataset of solar irradiance (in W/m^2) and temperature (in $^\circ\text{C}$) obtained from 3 NOAA weather stations: Fallbrook, Redding and Merced in California (CA). Data is sampled at 15-minute intervals. For estimating the parameters of temperature model from (14), $\alpha[k]$, $\beta[k]$, $\gamma[q]$, it was assumed that $M_2 = 12$ i.e. 3 hours and $M_1 = 6$ i.e. 1.5 hour. Temperature and irradiance data for 500 contiguous days were used for estimation of the parameter μ of the conditional distribution in (46).

As an example, the fit of the model to the temperature data along with the corresponding irradiance in Merced, CA is highlighted in Fig. 16 for a week in May 2014. In Fig. 17, the estimated parameters $\hat{\alpha}[k]$, $\hat{\beta}[k]$, $\hat{\gamma}[q]$ for all the weather stations are plotted.

To highlight the efficacy of the probabilistic forecasts, $m = 2$ step ahead forecast, i.e. 30 minutes ahead, was performed. The

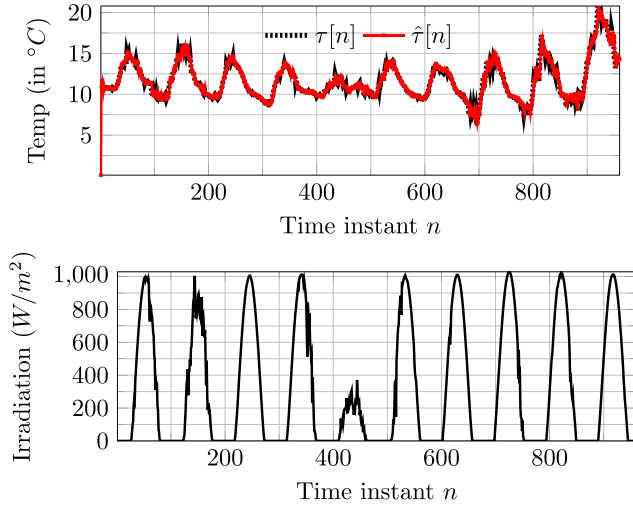


Fig. 16. Estimate of temperature, $\hat{\tau}[n]$ and corresponding irradiance $w[n]$ for May 2 – 12, 2014 at NOAA weather station Merced, CA.

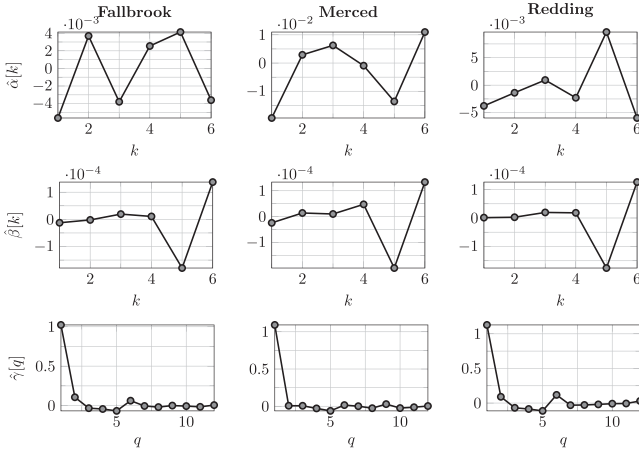


Fig. 17. Estimates, $\hat{\alpha}[k], \hat{\beta}[k], \hat{\gamma}[q]$ for 3 NOAA weather stations in California.

joint probability density takes the form of (50) with $m = 2$. This gives a 4-D probability distribution (2-D for temperature and 2-D for irradiance). The accuracy of the probabilistic forecast is evaluated in two ways: Continuous ranked probability score (CRPS) for one observation,

$$\begin{aligned} \text{CRPS}(\mathbf{w}_n^{+m}, \boldsymbol{\tau}_n^{+m}) &= \int_{\tilde{\mathbf{w}}_n^{+m}, \tilde{\boldsymbol{\tau}}_n^{+m}} \left[F(\tilde{\mathbf{w}}_n^{+m}, \tilde{\boldsymbol{\tau}}_n^{+m} | \mathbf{w}_d^{-M_1}, \boldsymbol{\tau}_n^{-M_2}) \right. \\ &\quad \left. + -u(\tilde{\mathbf{w}}_n^{+m} - \mathbf{w}_n^{+m})u(\tilde{\boldsymbol{\tau}}_n^{+m} - \boldsymbol{\tau}_n^{+m}) \right]^2 d\tilde{\mathbf{w}}_n^{+m} d\tilde{\boldsymbol{\tau}}_n^{+m} \end{aligned} \quad (66)$$

where $(\mathbf{w}_n^{+m}, \boldsymbol{\tau}_n^{+m})$ are the observed values and $F(\cdot|\cdot)$ is the joint CDF (from (50)) and $u(\cdot)$ is the Heaviside function acting on a vector. CRPS in the point forecast (single value) case reduces to mean absolute error (MAE) [58] after averaging over all observed samples $\mathbf{w}_n^{+m}, \boldsymbol{\tau}_n^{+m}$. Thus, the units of CRPS becomes $(\text{W/m}^2)^2 (\text{°C})^2$. Forecasting experiments for 30 days and for 15 minute time-intervals within each day were performed.

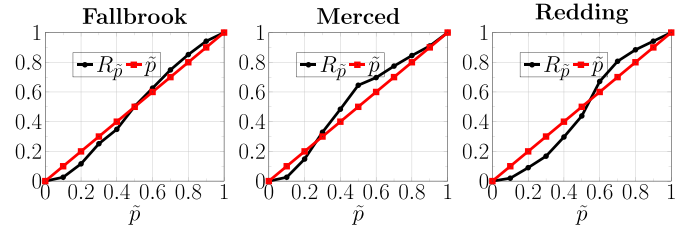


Fig. 18. Reliability $R_{\tilde{p}}$ of joint probabilistic forecasts: 2 steps ahead for Fallbrook, Merced and Redding, CA.

TABLE II
TABLE CRPS FOR JOINT FORECAST (IN $(\text{W/M}^2)^2 (\text{°C})^2$)

Fallbrook	Merced	Redding
489.5	440.02	443.80

To evaluate the metric of reliability for the joint distribution, let a probability interval (PI), $I_{\tilde{p}}$, cover the observed values $\xi = \{\mathbf{w}_n^{+m}, \boldsymbol{\tau}_n^{+m}\}, \xi \in \mathcal{P}$ with probability \tilde{p} . Then reliability $R_{\tilde{p}}$ is calculated as,

$$R_{\tilde{p}} = |\mathcal{P}|^{-1} \sum_{\xi \in \mathcal{P}} \mathbb{I}_{(\xi \in I_{\tilde{p}})} \quad (67)$$

where \mathbb{I} is an indicator function with value 1 if the observed sample belongs to the probability interval obtained using the joint CDF. Fig. 18 shows the reliability of prediction using the proposed forecasting technique for all three weather stations. Ideally, it should be close to the diagonal line \tilde{p} . The reliability of the proposed method closely matches the line as can be seen from the figure. Slight variation across different locations is observed. Broadly, it is demonstrative of the fact that the predictive distribution (CDF) provided is close to the observed CDF of real data. CRPS values are reported in the Table II.

C. Spatio-Temporal Model for Solar Power and Temperature

To illustrate the efficacy of the modeling approach for solar power and temperature from 15 installations within a ZIP code, data from SolarCity is used. Solar PV power from rooftop installations in Antioch, CA, with ZIP code 94531 is used along with hourly temperature information from the local weather station Concord Buchanan Field, CA available online.³ The fit of the proposed model to solar power data for a single day for all locations is highlighted in Fig. 19. Fig. 20 shows the average solar power and the corresponding temperature. The corresponding normalized root-mean squared error (NMSE)

$$\text{NMSE}_d^\ell = \sum_k (w_d^\ell[k] - \hat{w}_d^\ell[k])^2 / \sum_k (w_d[k])^2 \quad (68)$$

for each day across all locations is plotted in 21. For cloudy days in the beginning of the time-series, higher NMSE values are reported due to uncoordinated spikes among different locations when local effects are present. For temperature, the distribution of error is plotted in Fig. 22.

³<https://www.ncdc.noaa.gov/cdo-web/datasets/LCD/stations/WBAN:23254/detail>

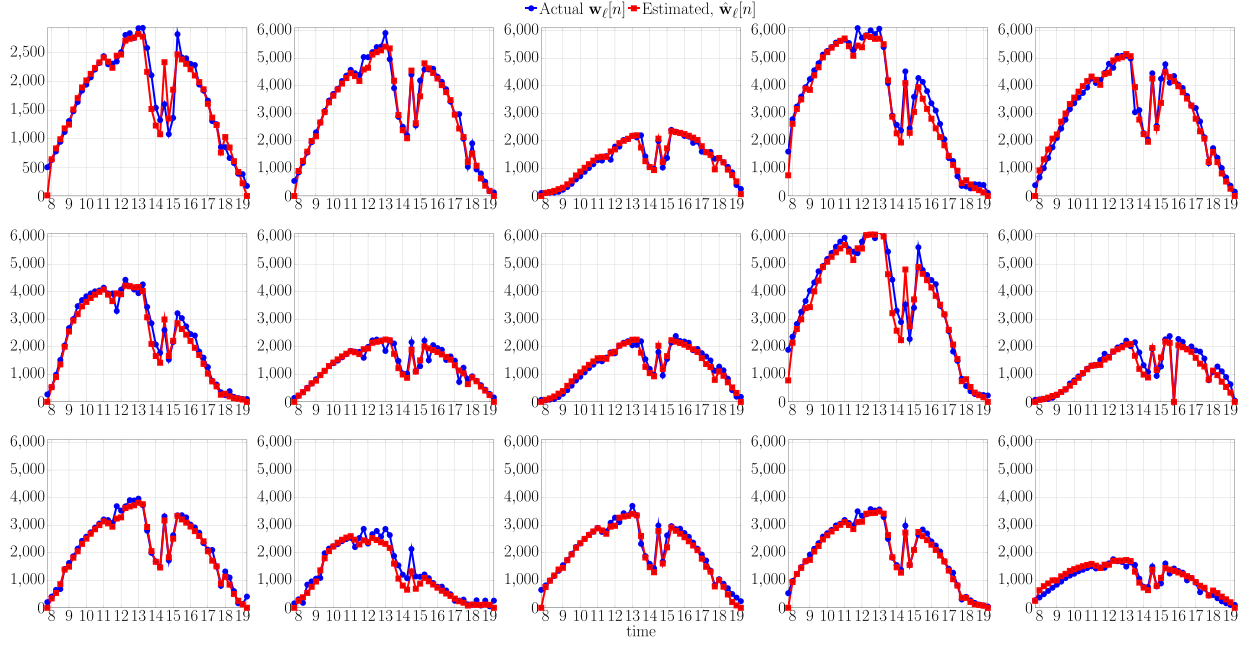


Fig. 19. Fit of the model for ZIP 94531, Antioch, CA.

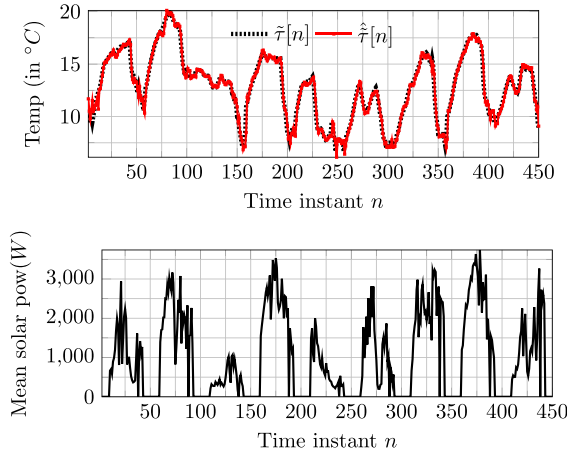
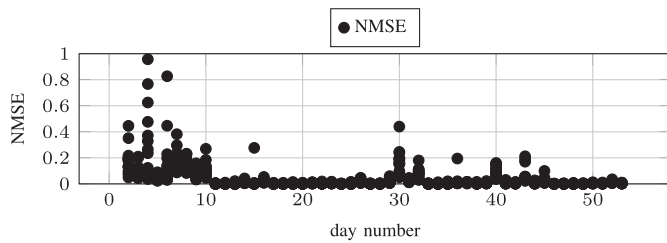
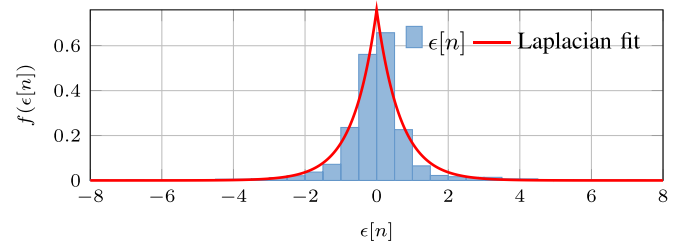

 Fig. 20. Estimate of temperature, $\hat{\tau}[n]$ and corresponding average solar power for $L = 15$ locations, $w[n] = L^{-1} \sum_{\ell=1}^L w_{\ell}^d[n]$ for March 26-April 3, 2014 in ZIP 94531, Antioch, CA.


Fig. 21. NMSE for 15 locations in ZIP 94531, Antioch, CA.

To evaluate the joint probabilistic forecasts, the CRPS measure is used with an alternative definition [62] for a multi-dimensional quantity. Let \mathbf{X} be the multi-dimensional random


 Fig. 22. Laplacian fit for $\epsilon[n]$ with $\mu = 1.5196$.

variable,

$$\mathbf{X}^T \triangleq \left[\left(\mathbf{w}_{n,1}^{+m} \right)^T \quad \left(\mathbf{w}_{n,2}^{+m} \right)^T \quad \cdots \quad \left(\mathbf{w}_{n,L}^{+m} \right)^T \quad \left(\tilde{\tau}_n^{+m} \right)^T \right]$$

with the distribution $f \triangleq f(\mathbf{X} | \{\mathbf{w}_{n,\ell}^{-M_1}\}_{\ell=1}^L, \tilde{\tau}_n^{-M_2})$ and \mathbf{X}' be also drawn from the same distribution. Let \mathbf{y} correspond to a certain realization or the observed values of power and temperature. Then,

$$\text{CRPS} \left(\left\{ \mathbf{w}_{n,\ell}^{+m} \right\}_{\ell=1}^L, \tilde{\tau}_n^{+m} \right) = \mathbb{E}_f \|\mathbf{X} - \mathbf{y}\| - \frac{1}{2} \mathbb{E}_f \|\mathbf{X} - \mathbf{X}'\|$$

This formulation is used in order to avoid computing the CDF since it is computationally intensive given the dimension of the distribution (16-D) even for a one-step ahead forecast. For the same reason of computational complexity, reliability measure is not calculated for the spatio-temporal solar power and temperature probabilistic forecast. The CRPS for one-step ahead forecast was found to be $5785 \text{ (W/m}^2\text{)}^{15} \text{ (}^\circ\text{C)}$ for the entire ZIP code.

VII. CONCLUSION

A joint stochastic model for short-term probabilistic forecast of solar PV power and temperature was presented which primarily focused on presenting the regime-switching nature exhibited in solar PV power that is dependent on cloud cover. Also, given solar power, the residual uncertainty in temperature was modeled using a second order Volterra model. The underlying statistical relationship between the quantities was captured in order to reduce the size of the sample space. The model also accounted for spatial correlation in solar PV power from multiple installations. The proposed model and methodology for forecasting were validated using real-world datasets and compared with different benchmark methods to highlight the efficacy of the proposed forecasting method. Future work involves taking into consideration spatial dependencies for a larger area where cloud motion dynamics can provide an ability to forecast and consequently connect the uncertainty in renewable power production with flexibility of electric power consumption provided by thermostatically controlled loads.

ACKNOWLEDGMENT

The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

REFERENCES

- [1] R. Ramakrishna and A. Scaglione, "A compressive sensing framework for the analysis of solar photo-voltaic power," in *Proc. 15th Asilomar Conf. Signals, Syst. Comput.*, 2016, pp. 308–312.
- [2] R. Ramakrishna, A. Bernstein, E. Dall'Anese, and A. Scaglione, "Joint probabilistic forecasts of temperature and solar irradiance," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 2018, pp. 3819–3823.
- [3] W. B. Powell and S. Meisel, "Tutorial on stochastic optimization in energy—Part I: Modeling and policies," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1459–1467, Mar. 2016.
- [4] M. Bazrafshan and N. Gatsis, "Decentralized stochastic optimal power flow in radial networks with distributed generation," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 787–801, Mar. 2017.
- [5] B. Defourny, D. Ernst, and L. Wehenkel, "Multistage stochastic programming: A scenario tree based approach to planning under uncertainty," *Decision Theory Models for Applications in Artificial Intelligence: Concepts and Solutions: Concepts and Solutions*. IGI Global, 2011, pp. 97–143.
- [6] T. J. VanderWeele and I. Shpitser, "On the definition of a confounder," *Ann. Statist.*, vol. 41, no. 1, pp. 196–220, 2013.
- [7] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, and Z. Hu, "Photovoltaic and solar power forecasting for smart grid energy management," *CSEE J. Power Energy Syst.*, vol. 1, no. 4, pp. 38–46, Dec. 2015.
- [8] H. C. Hottel, "A simple model for estimating the transmittance of direct solar radiation through clear atmospheres," *Sol. Energy*, vol. 18, pp. 129–134, 1976.
- [9] S. Pfenninger, and I. Staffell, "Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data," *Energy*, vol. 114, pp. 1251–1265, Aug. 2016.
- [10] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 2, no. 1, pp. 2–10, Mar. 2009.
- [11] D. Larson, L. Nonnenmacher, and C. F. Coimbra, "Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest," *Renewable Energy*, vol. 91, pp. 11–20, Jan. 2016.
- [12] T. E. Hoff and R. Perez, "Quantifying PV power output variability," *Sol. Energy*, vol. 84, no. 10, pp. 1782–1793, 2010.
- [13] C. W. Chow *et al.*, "Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed," *Solar Energy*, vol. 85, pp. 2881–2893, 2011.
- [14] J. Bosch and J. Kleissl, "Cloud motion vectors from a network of ground sensors in a solar power plant," *Sol. Energy*, vol. 95, pp. 13–20, May 2013.
- [15] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Solar Energy*, vol. 83, pp. 1772–1783, Mar. 2009.
- [16] E. B. Iversena, J. M. Morales, J. K. Møller, and H. Madsen, "Probabilistic forecasts of solar irradiance by stochastic differential equations," *Environmetrics*, vol. 25, no. 3, pp. 152–164, 2014.
- [17] J. Boland, M. Korolkiewicz, M. Agrawal, and J. Huang, "Forecasting solar radiation on short time scales using a coupled autoregressive and dynamical system (CARDS) model," in *Proc. 50th Annu. Conf. Australian Sol. Energy Soc.*, Dec. 2012, pp. 1–10.
- [18] C. Yang, A. A. Thatte, and L. Xie, "Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 104–112, Jan. 2015.
- [19] X. G. Agoua, R. Girard, and G. Kariniotakis, "Short-term spatio-temporal forecasting of photovoltaic power production," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 538–546, Apr. 2018.
- [20] M. Lave, J. Kleissl, and J. S. Stein, "A wavelet-based variability model (WVM) for solar PV power plants," *IEEE Trans. Sustain. Energy*, vol. 4, no. 2, pp. 501–509, Apr. 2013.
- [21] M. Fan, V. Vittal, G. T. Heydt, and R. Ayyanar, "Preprocessing uncertain photovoltaic data," *IEEE Trans. Sustain. Energy*, vol. 5, no. 1, pp. 351–352, Jan. 2014.
- [22] P. Shamsi, M. Marsousi, H. Xie, and W. Fries, "Dictionary learning for short-term prediction of solar PV production," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2015, pp. 1–5.
- [23] A. Mellit and S. Kalogirou, "Artificial intelligence techniques for photovoltaic applications: A review," *Prog. Energy Combustion Sci.*, vol. 34, no. 5, pp. 574–632, 2008.
- [24] H. Wang *et al.*, "Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network," *Energy Convers. Manage.*, vol. 153, pp. 409–422, 2017.
- [25] M. D. Tabone and D. S. Callaway, "Modeling variability and uncertainty of photovoltaic generation: A hidden state spatial statistical approach," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 2965–2973, Nov. 2015.
- [26] Y. Z. Li, R. Luan, and J. C. Niu, "Forecast of power generation for grid-connected photovoltaic system based on grey model and Markov chain," in *Proc. IEEE Conf. Ind. Electron. Appl.*, Jun. 2008, pp. 1729–1733.
- [27] Z. Ren, W. Yan, X. Zhao, W. Li, and J. Yu, "Chronological probability model of photovoltaic generation," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1077–1088, May 2014.
- [28] M. J. Sanjari and H. B. Gooi, "Probabilistic forecast of PV power generation based on higher order Markov chain," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2942–2952, Jul. 2017.
- [29] A. Bracale, G. Carpinelli, and P. D. Falco, "A probabilistic competitive ensemble method for short-term photovoltaic power forecasting," *IEEE Trans. Sustain. Energy*, vol. 8, no. 2, pp. 551–560, Apr. 2017.
- [30] S. A. Fatemi, A. Kuh, and M. Fripp, "Parametric methods for probabilistic forecasting of solar irradiance," *Renewable Energy*, vol. 129, pp. 666–676, 2018.
- [31] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," *Int. J. Forecasting*, vol. 32, no. 3, pp. 896–913, 2016.
- [32] Y. Li, B. Zhao, Z. Zhao, R. Taylor, and R. Wang, "Performance study of a grid-connected photovoltaic powered central air conditioner in the South China climate," *Renewable Energy*, vol. 126, pp. 1113–1125, 2018.
- [33] R. Perez, T. Hoff, C. Herig, and J. Shah, "Maximizing PV peak shaving with solar load control: Validation of a web-based economic evaluation tool," *Sol. Energy*, vol. 74, pp. 409–415, 2003.
- [34] E. C. Kara, M. D. Tabone, J. S. MacDonald, D. S. Callaway, and S. Kilicote, "Quantifying flexibility of residential thermostatically controlled loads for demand response: A data-driven approach," in *Proc. 1st ACM Conf. Embedded Syst. Energy-Efficient Buildings*, 2014, pp. 140–147.
- [35] A. Arabali, M. Ghofrani, M. Etezadi-Amoli, and M. S. Fadali, "Stochastic performance assessment and sizing for a hybrid power system of solar/wind/energy storage," *IEEE Trans. Sustain. Energy*, vol. 5, no. 2, pp. 363–371, Apr. 2014.
- [36] Y. Fujimoto *et al.*, "Distributed energy management for comprehensive utilization of residential photovoltaic outputs," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1216–1227, Mar. 2016.
- [37] R. Luthander, J. Widén, D. Nilsson, and J. Palm, "Photovoltaic self-consumption in buildings: A review," *Appl. Energy*, vol. 142, pp. 80–94, 2015.
- [38] J. D. Hamilton, "Regime-switching models," *The New Palgrave Dictionary of Economics*. Springer, 2016, pp. 1–7.

- [39] G. M. Masters, *Renewable and Efficient Electric Power Systems*. New York, NY, USA: Wiley, 2004.
- [40] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*. New York, NY, USA: Dover, 1965.
- [41] A. Kankiewicz, M. Sengupta, and D. Moon, "Observed impacts of transient clouds on utility-scale PV fields," in *Proc. ASES Nat Sol. Conf.*, 2010.
- [42] A. E. Curtright and J. Apt, "The character of power output from utility-scale photovoltaic systems," *Prog. Photovolt., Res. Appl.*, vol. 16, pp. 241–247, Sep. 2007.
- [43] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*. London, U.K.: Pearson, 2010.
- [44] I. Tosic and P. Frossard, "Dictionary learning," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27–38, Mar. 2011.
- [45] B. Mailhé, S. Lesage, R. Gribonval, F. Bimbot, and P. Vandergheynst, "Shift-invariant dictionary learning for sparse representations: Extending K-SVD," in *Proc. Eur. Signal Process. Conf.*, 2008, vol. 4, pp. 1–5.
- [46] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, Jun. 2010.
- [47] S. J. Wright, "Coordinate descent algorithms," *Math. Program.*, vol. 151, no. 1, pp. 3–34, Jun. 2015.
- [48] W. G. Kannuluik and E. H. Carman, "The temperature dependence of the thermal conductivity of air," *Australian J. Scientific Res.*, vol. 4, no. 3, pp. 305–314, 1951.
- [49] D. Halliday, R. Resnick, and J. Walker, *Fundamentals of Physics*. New York, NY, USA: Wiley, 2010.
- [50] G.-O. A. Glentis, P. Koukoulas, and N. Kalouptsidis, "Efficient algorithms for volterra system identification," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 3042–3057, Nov. 1999.
- [51] T. Koh and E. J. Powers, "Second-order volterra filtering and its application to nonlinear system identification," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 6, pp. 1445–1455, Dec. 1985.
- [52] R. Ramakrishna, A. Scaglione, A. Spanias, and C. Tepedelenlioglu, "Distributed Bayesian estimation with low-rank data: Application to solar array processing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2019, pp. 4440–4444.
- [53] R. H. Inman, H. T. Pedro, and C. F. Coimbra, "Solar forecasting methods for renewable energy integration," *Prog. Energy Combustion Sci.*, vol. 39, pp. 535–576, 2013.
- [54] F. Jelinek, "Continuous speech recognition by statistical methods," *Proc. IEEE*, vol. 64, no. 4, pp. 532–556, Apr. 1976.
- [55] B. H. Juang and L. Rabiner, "The segmental k-means algorithm for estimating parameters of hidden Markov models," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 9, pp. 1639–1641, Sep. 1990.
- [56] G. D. Forney, "The viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.
- [57] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
- [58] T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," *J. Roy. Statistical Soc.*, vol. 69, no. 243–268, 2007.
- [59] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Optimal prediction intervals of wind power generation," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1166–1174, May 2014.
- [60] R. L. Winkler, "A decision-theoretic approach to interval estimation," *J. Amer. Statistical Assoc.*, vol. 67, no. 33, 1972, pp. 187–191.
- [61] R. Koenker and K. F. Hallock, "Quantile regression," *J. Econ. Perspectives*, vol. 15, no. 4, pp. 143–156, 2001.
- [62] E. P. Gneiting, T. Gneiting, V. Berrocal, and N. A. Johnson, "The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification," *Quart. J. Roy. Meteorological Soc.*, vol. 132, no. 621C, pp. 2925–2942, 2006.



Raksha Ramakrishna (S'15) received the B.E. degree in electronics and communications engineering from the Rashtriya Vidyalaya College of Engineering, Bangalore, India, in 2014 and the M.S. degree in electrical engineering in 2017 from Arizona State University, Tempe, AZ, USA, where she is currently working toward the Ph.D. degree. Her research interests are in the domain of statistical signal processing and data analytics for power systems.



Anna Scaglione (F'11) received the M.Sc. degree in 1995 and the Ph.D. degree in 1999. She is currently a Professor in electrical and computer engineering with Arizona State University, Tempe, AZ, USA. Her expertise is in the broad area of statistical signal processing for information networks and intelligent cyber-physical infrastructures for secure and resilient energy delivery systems. She served IEEE in many capacities: She was an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS AND SIGNAL PROCESSING, EiC of the IEEE SIGNAL PROCESSING LETTERS, and is currently a Senior Editor for the *Transactions on Control over Networked Systems*. She was a member of the Signal Processing Society Board of Governors from 2011 to 2014. She received the 2000 IEEE Signal Processing Transactions Best Paper Award, 2013 IEEE Donald G. Fink Prize Paper Award for the best review paper in the IEEE publications and her work with her student earned 2013 IEEE Signal Processing Society Young Author Best Paper Award (Lin Li).



Vijay Vittal (S'78–F'97) received the B.E. degree in electrical engineering from the B.M.S. College of Engineering, Bangalore, India, in 1977, the M.Tech. degree from the Indian Institute of Technology, Kanpur, India, in 1979, and the Ph.D. degree from Iowa State University, Ames, IA, USA, in 1982. He is the Ira A. Fulton Chair Professor with the Department of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA. He is currently the Director of the Power System Engineering Research Center (PSERC) Headquartered at Arizona State University. Dr. Vittal is a member of the National Academy of Engineering.



Emiliano Dall'Anese (M'11) received the Ph.D. degree in information engineering from the Department of Information Engineering, University of Padova, Padua, Italy, in 2011. From January 2009 to September 2010, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of Minnesota, where he was also a Postdoctoral Associate from January 2011 to November 2014. From December 2014 to August 2018, he was a Senior Researcher with the National Renewable Energy Laboratory. Since August 2018, he has been an Assistant Professor with the Department of Electrical, Computer, and Energy Engineering, University of Colorado Boulder, Boulder, CO, USA. His research interests lie in the broad areas of optimization, control, and signal processing, with application to power grids and networked systems.



Andrey Bernstein (M'15) received the Ph.D. degree in electrical engineering from the Technion - Israel Institute of Technology, Haifa, Israel, in 2013. From 2010 to 2011, he was a Visiting Researcher with Columbia University, New York, NY, USA. From 2011 to 2012, he was a Visiting Assistant Professor with Stony Brook University, Stony Brook, NY, USA. From 2013 to 2016, he was a Postdoctoral Researcher with Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland. Since 2016, he has been a Senior Scientist with National Renewable Energy Laboratory, Golden, CO, USA. His research interests are in the optimization and control problems of complex systems, with application to intelligent power and energy systems.