Probabilistic Analysis of Masked Loads with Aggregated Photovoltaic Production

Shaohui Liu
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, TX, USA
shaohui.liu@utexas.edu

Daniel Adrian Maldonado
Emil M. Constantinescu
Mathematics and Computer Science
Argonne National Laboratory
Lemont, IL, USA
{maldonadod, emconsta}@anl.gov

***Abstract*—In this paper we present a probabilistic analysis framework to estimate behind-the-meter photovoltaic generation in real time. We develop a forward model consisting of a spatiotemporal stochastic process that represents the photovoltaic generation and a stochastic differential equation with jumps that represents the demand. We employ this model to disaggregate the behind-the-meter photovoltaic generation using net load and irradiance measurements.**

***Index Terms*—distributed PV system, SDE, spatiotemporal model, real-time prediction**

I. Introduction

The increase in penetration of user-sized distributed energy resource (DER) systems poses challenges for the planning and operation of the grid. A major issue with behind-the-meter (BTM) solar generation is the lack of direct measurements of the instantaneous power injections. Furthermore, the volatility of solar generation production caused by weather variability (e.g., cloud coverage) brings additional uncertainty to forecasts [1]. This lack of observability makes it difficult to quantify the aggregated effect of BTM photovoltaics (PV) generation on the transmission grid. Proper characterization of BTM PV generation in real time would allow utilities to prepare for and quantify the risk of situations in which sudden ramps in generation occur or in which a large amount of DER systems trip after a fault.

Given the constraints that utilities face with respect to measurements, forecasts, and parameters of inverters; recent studies have investigated the issue of "disaggregating" the PV power signal from the measurements or inferring the instantaneous PV power through irradiance measurements and other proxy regressors. One can consider the disaggregation problem as an approximate algebraic relationship:

$$P_{\text{NET}} = P_{\text{MASKED}} - P_{\text{PV}} \tag{1}$$

where P_{NET} is the net power seen by the transmission grid (also called measured power), P_{PV} is the aggregated power generated by the distributed PV inverters, and P_{MASKED} is the actual aggregated load demand that remains *masked* by the PV production. In a high loading and high PV production scenario, the utility might underestimate the actual load in the feeder, and a voltage transient that trips a large amount of PV inverters may jeopardize the dynamic stability of the system.

Several researchers are investigating how to disaggregate the masked load from the net load signals. Vrettos et al. [2] characterize the literature in three main groups: transposition model approaches, data-driven approaches, and hybrid approaches. The transposition methods involve extrapolating irradiance to a set of inverter models to compute the PV generation. For instance, Engerer and Mills [3] use proxy measurements from a PV inverter together with the clear-sky index and a PV inverter performance model to extrapolate the generation of the rest of the inverters. In [4], Killinger et al. further delve into the cases in which the parameters of the PV inverters are not uniform, and they develop a "projection method" to calculate the global horizontal irradiance (GHI) using a proxy power measurement. The GHI measures the total amount of irradiance received by a flat surface at the ground from above, and is a central indicator of the solar energy that can be produced by PV panels. With regard to the data-driven methods, Sossan et al. [5] and Patel et al. [6] analyze the impact of global horizontal irradiance fluctuations on the time series and use this information to desegregate the PV generation from the net load signal. The hybrid approaches include [7] and [8]. In the former publication, neural networks together with load forecasts and PV production models are used to forecast the net load. In the latter, Bright et al. use satellite-derived GHI estimates in 10-minute intervals, together with PV generation models, to interpolate to the aggregated PV generation.

In this paper we propose a novel algorithm for the disaggregation of instantaneous PV generation in a feeder that falls into the category of hybrid methods. Our methodology differs from previous work in that we consider high-frequency measurements of irradiance and net load. It has been shown in [9] and [10] that both the load and the PV generation, in short time intervals, can be characterized as stochastic processes with certain properties.

To extract information from the higher-resolution time series, we fit statistics such as temporal variance, autocorrelation

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research under Contract DE-AC02-06CH11357.

and variogram, which allow us to obtain the parameters of the underlying masked load process, provided that we have a model for the spatial irradiance and the installed PV panels. While modeling the instantaneous PV generation with irradiance data and inverter parameters can be error-prone, by leveraging the spatiotemporal statistics produced by the irradiance fluctuation we can tolerate higher errors in the model. To this end, we develop a model for the geographically distributed PV aggregation power injection with limited solar irradiance measurements. We then design a modified stochastic differential equation model based on the Ornstein-Uhlenbeck process with jumps to simulate the masked load yielding a jump-diffusion process [9]. The net load model then results from the combination of the two submodels. We further design a disaggregation algorithm to mitigate the error of masked load estimation caused by the estimated PV generation. Simulation studies with real recorded solar irradiance data and load data recorded by μPMU show that the spatiotemporal model with the disaggregation algorithm is a tenable method to reduce disaggregation error. Moreover, simulations indicate that we can accurately estimate the aggregated PV active power generation at a distribution feeder with limited sensor deployment.

The rest of the paper is organized as follows. In section II we model the PV power with a novel spatial Gaussian process (GP) for predicting solar irradiance under limited observations. In section III we model the load power by an Ornstein-Uhlenback (OU) process with jumps. In section IV we propose a disaggregation algorithm that separates the net power with real-time solar irradiance. In section VI we concluded our work and discussed the possible future work.

II. Solar Generation Model

In this section we consider the construction of a stochastic model of the aggregated PV power generation in a region, using sparse irradiance measurements and inverter performance equations. The irradiance GP-based forecast follows a standard kriging framework with examples that more recently include [11].

A. Spatial Gaussian Process for Clear Sky Index

Whereas direct measurement of the instantaneous power injection of each PV inverter is infeasible, one can build approximate models that simulate the spatial distribution of the irradiance and, together with data and models of the installed PV inverters, approximate the total PV generation in a feeder. Thus, we focus our efforts on developing a reliable model for the prediction of the aggregated irradiance that acknowledges the sparsity of measurements.

To build this model we employ a spatial Gaussian process to represent the variability and spatial correlations in solar irradiance. Although the normal marginals do not represent well the solar distribution, which tends to be bimodal (i.e., have two concentration peaks corresponding to cloudy and sunny conditions [12]), the correlations are useful in determining how the irradiance in the geographic region co-varies. This relation is expressed through the conditional distribution, which is the critical ingredient in our predictive framework. GPs have closed forms for the posterior and conditional distributions and this confers a distinctive advantage in achieving fast simulation and sampling, which can be critical in real-time applications. We expect this conditional distribution to depend on the weather conditions, season, and climate. Such a GP calibration process likely needs to take place with varying degrees in each region where it is deployed. In the next step, the solar irradiance is used to estimate the PV power by propagating the irradiance through a set of inverter models whose location and parameters are assumed to be known. For simplicity, we will assume the parameters of these inverters are uniform, and we neglect model errors. One approach to alleviate this restriction is discussed in [4].

We consider a realistic setting that assumes we have one or two global horizontal irradiance (GHI) observations per neighborhood. These observations are used to estimate the total amount solar generation for the entire area.

The procedure we used to estimate the forecast solar production is as follows. First we measured the global horizontal irradiance, G, for several spatial locations and calculated the clear-sky horizontal irradiance, G_c, for the same locations. With these quantities we estimated the clear-sky index, κ, for each site by

$$G = \kappa G_c \,. \tag{2}$$

The clear-sky index represents the fraction of irradiance that passes through atmosphere relative to clear-sky conditions. The advantage of using the clear-sky index is that it is a detrended quantity. We assume a joint distribution for $\kappa \sim \mathcal{N}(\mu, \Sigma)$, where the mean μ is set to zero by debiasing the data, and the covariance matrix Σ is a symmetric positive definite matrix, $\Sigma = [\epsilon_{ij}]_{1\le i,j\le n}$. Many models for the covariance function exist. Here we employ a relatively simple anisotropic kernel:

$$\epsilon_{ij} = \alpha \cdot \exp\left(-\left(\theta_x^2 (r_{ij}^x)^2 + \theta_y^2 (r_{ij}^y)^2\right)\right) + \beta \cdot \delta_{ij} \,, \tag{3}$$

with $i, j = 1, 2, \ldots, n$, where α, β, θ_x, and θ_y are parameters; r_{ij}^x and r_{ij}^y are spatial distances between site i and site j in the x and y directions, respectively; δ_{ij} is the the Kronecker delta function; and $\beta \cdot \delta_{ij}$ has the effect of a statistical nugget. We estimate the kernel parameters by a least-squares method:

$$\min_{\alpha,\beta,\theta_x,\theta_y} ||\Sigma_{model} - \Sigma_{obs}||_2 \,, \tag{4}$$

where Σ_{obs} is the empirical covariance of measured κ and Σ_{model} is given by the parametric function (3). We note here that this is a spatial model aimed at characterizing the irradiance variability in a small area. Because our final measure in this study is total solar irradiance, we argue that this model has sufficient complexity as our numerical experiments illustrate.

We have also implemented a maximum likelihood estimation procedure; however, for our setup the differences were negligible. The anisotropy, measured as the difference between latitudinal (north-south) and longitudinal (east-west) components of the GP kernel, seems to play an important role. We

recorded a difference of about 20% between these components in our experiments, which represents a point of departure from studies such as [11]. This indicates a predominant flow direction, which confers more accurate predictions in space.

We use the joint distribution to infer the clear sky index at unobserved locations. If we denote by X_1 the unobserved locations and the observation sites X_2 then their joint distribution is represented by

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right), \tag{5}$$

where Σ_{ij} are block covariance matrices as in (3). We compute the conditional distribution of X_1 provided observations X_2: $(X_1|X_2) \sim \mathcal{N}(\mu', \Sigma')$ and expressed in closed form by

$$\mu' = \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)\,, \tag{6a}$$

$$\Sigma' = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\,. \tag{6b}$$

We use a dataset that provides one-year's worth of global horizontal solar irradiance data of 17 distributed sites collected every second during daylight located on Oahu island, Hawaii [13]. The clear-sky irradiance, G_c, is calculated by pvlib [14] using the precise time and coordinates data of solar panels assumed to be collocated with the irradiance measurements. The site location and names are illustrated in Fig. 1. Indicated also are two sites used as observations and two that represent outliers in terms of proximity to the site clusters.

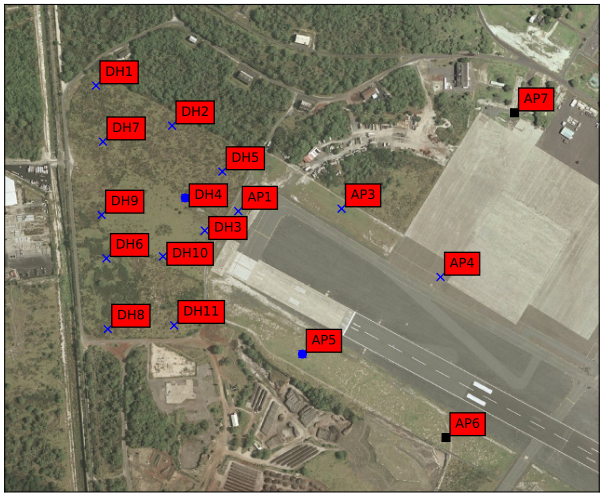

Fig. 1. Illustration of the solar PV and observation sites. The blue circles (DH4 and AP5) indicate stations that are used as observations. Black squares indicate remote sites that show smaller correlation with the rest of them.

We compute the empirical covariance of the 17 sites and plot the matrix entries and the entries of the calibrated covariance model (3) in Fig. 2. The covariance represents how the 17 stations covary at every time instance in space. We note that the model covariance approximates well the structure of the empirical covariance. Moreover, we can see the outlier locations being less correlated with the rest of them.

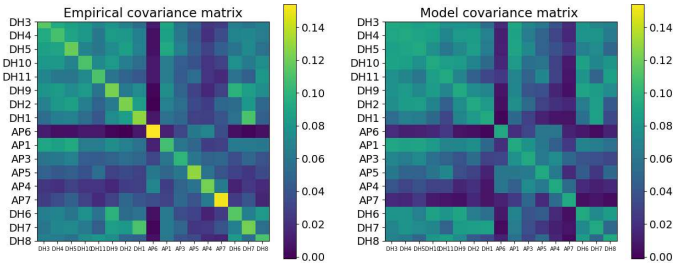

Fig. 2. The covariance matrix calculated by real κ data and the covariance matrix model calculated by using kernel (3).

Notation	Meaning
κ	clear-sky index
G_c	clear-sky irradiance
k_d	diffuse fraction: $\frac{\text{global irradiance}}{\text{extraterrestrial irradiance}}$
R_b	geometric factor: scaling factor of incidence angle
A_i	anisotropy index: $\frac{\text{beam radiation}}{\text{extraterrestrial radiation}}$
β	tilt angle of the tilted plane
ρ_g	albedo of the ground
A	total area of the PV array
η	PV module conversion efficiency
q_a	additional module/array loss
P_{ac0}	rated max AC power of inverter
P_{dc0}	DC power at which inverter reaches AC rating
P_{s0}	inverter threshold power (start to give AC power)

TABLE I
NOTATION IN PV MODEL AND DEFINITION

B. Power Model of PV Systems

The following are derivations of the power model of PV system based on solar irradiance.

GHI (global horizontal irradiance):

$$G = \kappa G_c\,. \tag{7}$$

Diffuse irradiance and beam irradiance:

$$G_d = k_d G\,, \tag{8}$$

$$G_b = G - G_d\,. \tag{9}$$

Global irradiance on tilted plane:

$$G_T = G_b R_b + G_d\left((1-A_i)\frac{1+\cos\beta}{2} + A_i R_b\right) + G\rho_g\frac{1-\cos\beta}{2}\,. \tag{10}$$

AC power output is calculated as in [1], [14]

$$P_{ac} = P_{ac0}\frac{P_{dc} - P_{s0}}{P_{dc0} - P_{s0}}\,. \tag{11}$$

We designed an experiment on synthetic data involving 17 sites with 2 observed sites. The irradiance data was sampled at 1Hz frequency, which is the same as with commonly used sensors.

Numerical results suggest that the GP model can precisely recover the covariance matrix with limited observations. Furthermore, the forecast method predicts the PV power production of the aggregated sites without complete observations, except for when sharp jumps in the irradiance are caused by

clouds moving in or our of the area. In Fig. 3 we illustrate the aggregated PV power computed by collocating uniform PV panels with deployed GHI sensors (Fig. 1), which is referred to as the observed PV power. We also use the GP procedure to estimate the clear sky index based on two observation sites for the entire region, compute the irradiance, and use the same PV model to estimate the PV power. The joint GP process

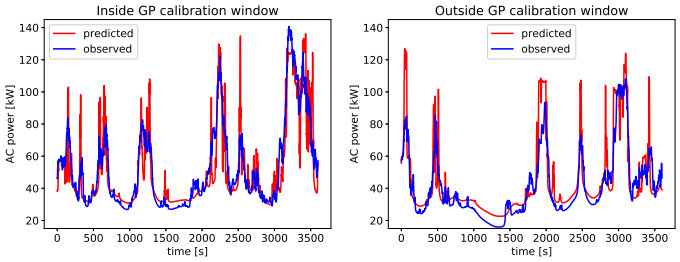

Fig. 3. Aggregate AC solar power observed and predicted by using two observation sites using eq. (11). The forecast correspond to April 8, 2010, (left) 11 am to noon local time where three sites were used for forecast, and (right) noon to 1 pm where the same GP fit was used to make predictions.

(5) is calibrated by using data that corresponds to the time frame in Fig. 3 (left). The same GP process is used to make prediction corresponding to the time frame in Fig. 3 (right). In other words, the GP is calibrated with the entire sensor network data for Fig. 3 (left), whereas in the second case the GP has access only to the designated observed sites. These results indicate a relatively good prediction capability of the spatial GP. The least accurate predictions are likely associated with sudden irradiance jumps linked to the incoming of clouds through an unobserved section of the PV generators. Extensions to temporal models might help alleviate these aspects, as illustrated for short forecast-time ahead in [11], [15].

III. Ornstein-Uhlenbeck Process for Load Modeling

We model the masked load as an Ornstein-Uhlenbeck process with spikes following [9]:

$$dx_t = \gamma(\mu - x_t)dt + \sigma d\omega_t + J_t dq_t, \tag{12}$$

where γ is the mean reversion rate and μ the long-term mean of the OU process. The variance of the OU process represents normal load changes such as small loads being switched on and off, and larger spikes represent sudden and less frequent switching of larger loads. We model $d\omega_t$ as a a standard Wiener process with diffusion σ [16]. For the jumps, we follow a Poisson process, where J_t is a random variable, $|J_t|$ follows Gamma distribution, and q_t is the Poisson random variable with intensity λ:

$$dq_t = \begin{cases} 1, \text{ w/ probability } \lambda dt \\ 0, \text{ w/ probability } 1 - \lambda dt\,. \end{cases} \tag{13}$$

A. Numerical Solution

We discretize (12) and derive the first-order numerical solution by the Euler-Maruyama method with step Δt:

$$x_{i+1} = x_i + \gamma(\mu - x_i)\Delta t + \sigma(W_i - W_{i-1}) + J_i(P_i - P_{i-1}) \tag{14}$$

where $\Delta W_i = W_i - W_{i-1}$ and $\Delta W_i \sim \sqrt{\Delta t}\mathcal{N}(0,1)$ are the independent increments. In particular, the model 12 has an explicit solution form that could be discretized as

$$x_i = \mu + (x_0 - \mu)e^{-i\Delta t\gamma} + \sigma \sum_{j=1}^{i} e^{-\gamma(i-j+1)\Delta t} \cdot (W_j - W_{j-1}) + \sum_{j=1}^{i} e^{-\gamma(i-j+1)\Delta t} \cdot J_{i-1}(q_i - q_{i-1}). \tag{15}$$

Note that our stochastic differential equation (SDE) model is for a relatively smooth system. Thus the numerical error for the Euler Maruyama scheme is relatively small, and the two methods generate nearly identical numerical solutions in our test cases. Furthermore, the computational cost of numerical scheme (14) is $O(n^2)$, while the cost of exact solution (15) is $O(n^3)$ with respect to the number of steps. So we implement the Euler Maruyama scheme in our algorithm.

B. Parameter Estimation

For parameter estimation with the discrete time series $\{X_i\}_{i=0}^N$ we first consider a simple filter. We let $y_{i+1} := f(X_i)$, where $f(X_i) = X_i + \gamma(\mu - X_i)\Delta t$. Then:

$$X_{i+1} = y_{i+1} + \xi_{i+1}, \tag{16}$$

where $\xi_i \sim \mathcal{N}(0, \sigma^2\Delta t)$ if we do not consider the rare jumps at first. With this preprocessing, we have the following algorithm.

We used an unbiased method based on a martingale estimation function to estimate the mean reversion rate γ [9]. In particular, the estimator is unbiased, consistent, and asymptotically normally distributed given the assumption that the underlying diffusion in the SDE model is ergodic [17]. We first write the martingale estimation function as

$$G_N(\gamma) = \sum_{i=1}^{N} \frac{\dot{b}(x_{i-1};\gamma)}{\sigma_{i-1}^2}\{x_i - \mu_t - (x_{t-1} - \mu_{t-1})e^{-\gamma}\}, \tag{17}$$

where

$$b(x_t;\gamma) = \frac{d\mu_t}{dt} + \gamma(\mu_t - x_t). \tag{18}$$

The estimation of the γ is the unique zero point of (17):

$$\hat{\gamma} = -\log\left(\frac{\sum_{i=1}^N Y_{i-1}\{x_i - \mu_i\}}{\sum_{i=1}^N Y_{i-1}\{x_{i-1} - \mu_{i-1}\}}\right), \tag{19}$$

where

$$Y_{i-1} = \frac{\mu_{i-1} - x_{i-1}}{\sigma_{i-1}^2}. \tag{20}$$

Algorithm 1 OU Parameter Estimation

Input: PMU data $\{x_i\}_{i=0}^N$
·Estimate the mean reversion rate γ by martingale function

·Calculate the random process set $\{\xi_i\}_{i=0}^N$ by (16)
·Calculate the mean μ_0 and variance σ_0 of $\{\xi_i\}$
for $1 \leq i \leq N$ **do**
  ·Identify jumps $\{J_j\}$ by $3\sigma_0$
end for
if $\#jumps > 0$ **then**
  ·Calculate Poisson parameter λ
  ·Estimate Gamma parameters shape k, scale θ for $\{J_j\}$
else
  ·No jumps identified
end if
· Calculate the modified increment mean μ_1 and variance σ_1 for $\{\xi_i\}_{i=0}^N \backslash \{J_j\}$
· Run Kolmogorov-Smirnov test for Gaussian and Gamma
if KS test passed **then**
  Return: $\mu_1, \sigma_1, \lambda, k, \theta$
end if

With the parameter estimation Algorithm 1, we can estimate the parameters for the OU process and then generate the predictions using the numerical solution scheme (14).

IV. BTM PV Generation Disaggregation

In real cases, after the installment of PV panels in the grid, we no longer have direct measurements of the load power. The only accessible data is net power measured by μPMU and limited measurements of solar irradiance. Thus we need to reduce the uncertainties for more precise prediction and planning. As with the separate tests in PV power and load power, the data set of GHI is in 1 Hz for 10 minutes, and we also down sampled the μPMU data to 1 Hz for 10 minutes for convenience.

A. Net Load Model

We consider the linear power model of the distributed system with three components: the net load - external power injection, the masked load - the sum of the consumer demand and the aggregated BTM PV power within the system (1). In real cases, while we have direct measurements on the net load, we have no direct information about the PV power and masked load power. PV power production could be predicted by the partial information of the solar irradiance. Moreover, we can estimate the corresponding masked load based on the estimate of BTM PV power and make further predictions. Thus the key to accurate prediction is a reliable disaggregation algorithm using limited information.

B. Disaggregation Strategy

For the Ornstein-Uhlenbeck process with jumps, we consider a full parameter vector:

$$\Theta = [\gamma, \mu, \mu_1, \sigma_1, k, \theta, \lambda] \tag{21}$$

where γ, μ are the mean reversion rate and the long-term mean of OU process; μ_1, σ_1 are the mean and standard deviation of Wiener process; k, θ are the shape and scale parameters of the Gamma distribution that describes jumps; and λ is the parameter for the Poisson process.

Then we can take the OU parameters calculated by Algorithm 1 using the μPMU data recorded before the installment of PV panels (thus without PV powers) as reference to calibrate the OU parameters of the masked load:

$$\min_{\Theta} ||S(P_{\text{NET}}^{\text{obs}}) - S(P(\Theta))||_2 + ||\Theta_{\text{prior}} - \Theta||_2, \tag{22}$$

where $S(\cdot)$ is a statistic of the net power time series of observed data and of data generated through simulations by using parameters Θ. The first term estimates the discrepancy between the statistics observed and the one generated by the simulated process. The second term represents a regularization, where Θ_{prior} can be either nominal values or zero. The statistic S is defined by considering the time series that generates a stochastic process $\hat{X}(\Theta) = \{\hat{x}_t(\Theta)\}$ by P_{NET} via (14) and P_{PV} in (1) and computes a series of statistics such as mean, standard deviation, and weighted autocorrelations' norm:

$$S(\hat{X}) := \left[\mu_{\hat{X}}, \sigma_{\hat{X}}, \frac{1}{t_1}||R_{\hat{X}\hat{X}}(\tau)||_2\right], \ \tau = 1, \cdots, t_1. \tag{23}$$

V. Numerical Results

We present two examples: a synthetic example (§V-A) and a realistic one (§V-B). In the synthetic example we generate the solar irradiance with known spatial distribution and consider one of the Oahu island measurements for the temporal correlation. We also generate a simplified OU masked load power. In the realistic case we use real μPMU measurements and irradiance to generate the net power. The total compute time in all our examples takes a few minutes on a regular laptop.

A. Synthetic Example

We start the numerical illustrations of the proposed framework by using a synthetic example. The point of this example is to test the framework in ideal situations that correspond to good parametric modeling of the irradiance, PV, and masked load. To this end, we generate the true masked load by using an OU process and the true PV power generation by using a Gaussian process, both with known parameters. This data set is used to generate the net load data. The observables in this system are the net load data and the irradiance at the two locations indicated in Fig. 1. In this setup we assume that the Gaussian process has an exact spatial structure and that the masked load is described by the correct OU process, but with unknown parameters. We aim to (i) recover the GP parameters from data and (ii) recover the OU process parameters of the masked load that together with the PV power best explain the observed net load.

1) Calibrating the GP: We assume that we have 17 PV panels in a limited area that corresponds to the Oahu irradiance sensor network (Fig. 1), out from we pick two as observations. The exact GP model has the following parameters: $\alpha = 0.0108$, $\beta = 0.0001$, $\theta_x = 61.6522$, $\theta_y = 74.081$. We first calibrate the GP model by using a least squares fit and one hour's worth of data (3600 seconds). The resulting GP parameters are $\alpha = 0.01085$, $\beta = 1.01e-05$, $\theta_x = 64.44631$, and $\theta_y = 70.899$, which is an excellent fit, as expected. Then we use the data from the two observation sites, the calibrated GP model and the conditional distribution (6), to predict the clear sky index at the remaining 15 sites.

2) Masked load and PV disaggregation: The disaggregation problem follows the steps described in §(IV-B). In particular, we solve an optimization problem that yields the maximum likelihood of the OU parameters (that define the masked load) that best explain the data (the net load). The likelihood is expressed in terms of the statistics of the observed and simulated data.

We set the true OU parameters μ_{OU}, γ_{OU}, and σ_{OU} to be $[400000, 0.01, 200]$. Note that in this case we do not use the jump process. We performed 50 solves with initial guesses initialized at $\pm 60\%$ around the true values. The results from solving these problems came to $\mu^*_{\text{OU}} \in (400463.6, 400463.8)$, $\sigma^*_{\text{OU}} \in (156.1, 156.8)$, and $\gamma^*_{\text{OU}} \in (0.0128, 0.0131)$. These results indicate that the minimizer is found closer to the true solution, and thus the estimator approximates the true values. This results in a good representation of the masked load process. In Fig. 4 we illustrate the masked load reconstruction from one such optimization (results look similar for the other ones).

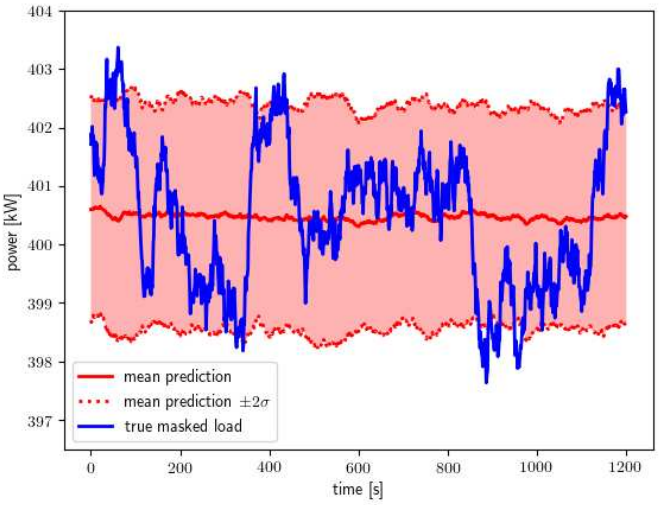

Fig. 4. Masked load (truth) and its reconstruction (prediction) through our disaggregation strategy. The predicted masked load is represented by the mean value process and $\pm 2\sigma$ deviation. The true value is covered as close to 90% by the prediction envelope, as expected.

B. Realistic Example

We also tested our framework on real data sets. We used the same solar PV sites as in the synthetic example, only now the solar data set is the one actually collected from the sensor network consisting of the 17 pyranometers, measuring GHI with a 1 Hz resolution [13]. The masked load power measurements are collected by μPMUs and PQube3 power quality meters manufactured by Power Standards Laboratory in Alameda, CA, at 120 Hz [18]. These measurements are downsampled to 1 Hz to match the solar sampling rate. The net load power data is obtained by (1).

1) Calibrating the GP: We assume that among the 17 PV panels we have the observable set $\{DH4, AP5\}$ (Fig. 1). We first calibrate the GP model by using a least squares fit and ten minutes of data (600 seconds). The resulting GP parameters are $\alpha = 0.09243$, $\beta = 1.00e{-}03$, $\theta_x = 20.14$, and $\theta_y = 17.63$, and the estimation error is $||\Sigma_{model\_opt} - \Sigma_{obs}||_2 = 0.223$, which is an excellent fit. Then we use the data from the two observation sites, the calibrated GP model and the conditional distribution (6), to predict the clear sky index at the remaining 15 sites and compute the AC PV power prediction. Comparing with the PV power computed by full observation data (all 17 locations) in Fig. 5, our prediction is close to the true value and successfully predicts the sudden injection jumps.

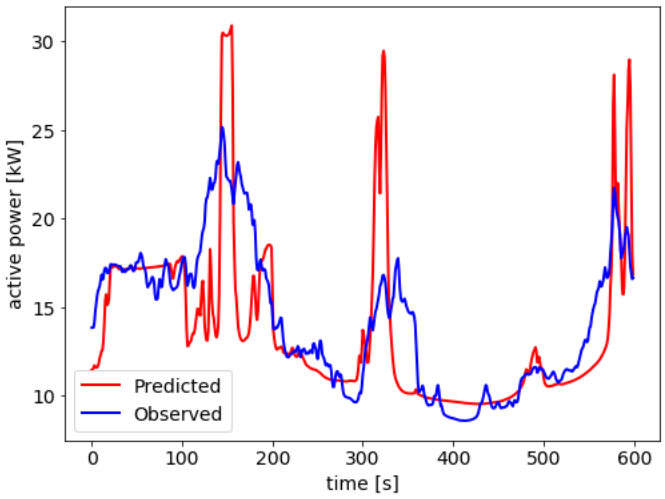

Fig. 5. Active PV power generation and its prediction through our GP model using eq. (11). The predicted PV power is inferred by GHI observations on site DH4 and AP5. The true value is calculated by full observation of 17 sites.

2) Masked load and PV disaggregation: The disaggregation problem follows the steps described in §(IV-B). We get a rough estimation of OU process parameter set Θ_1. Then we solve an optimization problem that yields the maximum likelihood of the OU parameters for Θ_{opt}, taking Θ_1 as the initial.

In the numerical experiment we use 5 minutes of irradiance data and net load data generated by recorded PMUs for disaggregation and parameter estimation. Then we make predictions of the masked load for the next 5 minutes using the estimated parameters. The absolute error of the parameter set suggests the parameter estimation of the OU process is significantly improved by introducing maximum likelihood, compared with the rough estimation in Table II.

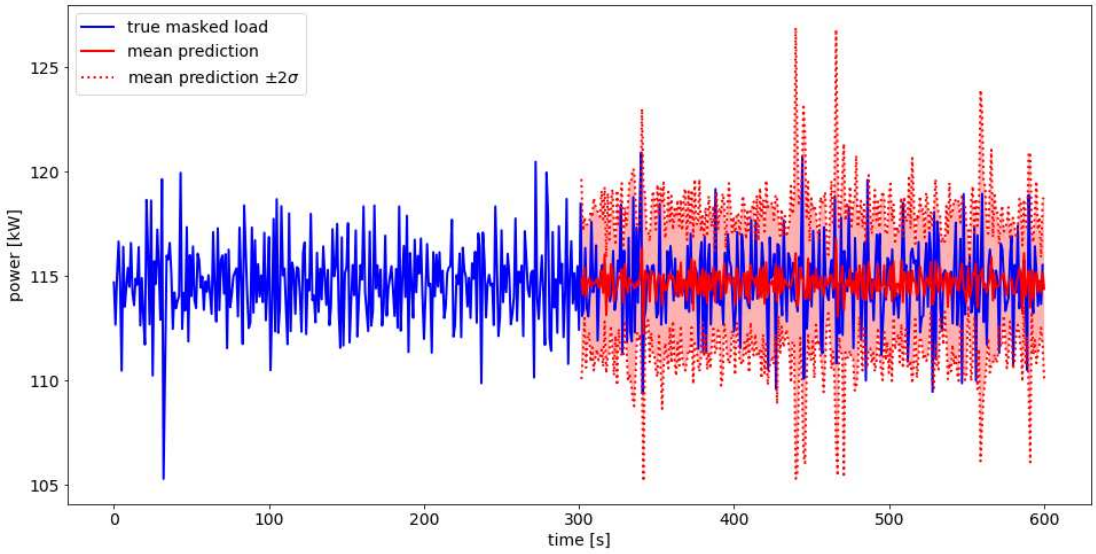

Fig. 6. Real load power data vs. BTM real power prediction by the Euler-Maruyama scheme for the OU process using parameters calibrated by maximum likelihood. We used 5 minutes of net load and irradiance data for the estimation and make predictions of the masked load for the next 5 minutes. The predicted masked load is represented by the mean value process and $\pm 2\sigma$ deviation calculated by 10 realizations. The true value is covered as close to 96.67% by the prediction envelope, as expected.

	μ	γ	μ_1	σ_1
$\lvert\Theta_{ref} - \Theta_1\rvert$	5.30e+02	1.97e-01	5.74e+01	1.53e+03
$\lvert\Theta_{ref} - \Theta_{opt}\rvert$	1.58e+01	8.39e-02	2.46e-02	2.11e-02
	k	θ	λ	
$\lvert\Theta_{ref} - \Theta_1\rvert$	1.45e-02	1.13e+03	0.00e+00	
$\lvert\Theta_{ref} - \Theta_{opt}\rvert$	5.64e-01	3.51e-02	0.00e+00	

TABLE II
OU PARAMETER ESTIMATION BY MAXIMUM LIKELIHOOD

The predicted masked load is represented by the mean value process and $\pm 2\sigma$ deviation calculated by 10 random realizations. The true value is covered as close to 96.67% by the prediction envelope. These intuitive results indicate that the minimizer is found close to the true solution, which results in a correct representation of the masked load process. In Fig. 6 we illustrate the masked load reconstruction from one such optimization (results look similar for the other ones). Numerical results indicate that our approach can disaggregate the masked load well from the net load using limited irradiance observation. It also capture its trend as well as the variability thus provide a more accurate prediction of the masked load than naive predictions by just assuming the mean and standard deviation from the historical data.

VI. CONCLUSION AND FUTURE WORK

In this paper we present a probabilistic analysis framework to estimate behind-the-meter photovoltaic generation in a single feeder network in real time. Within this framework we develop a forward model consisting of a spatial stochastic process that estimates the photovoltaic generation based on a couple of sensors and a temporal stochastic differential equation with jumps that estimates the masked user load demand. These models are used to disaggregate the behind-the-meter photovoltaic generation by using net load and partial irradiance measurements. Simulation studies with both synthetic and real recorded solar irradiance data and μPMU data indicate that the proposed framework is a tenable method to provide a reliable disaggregation procedure. Moreover, simulations indicate that we can accurately estimate the aggregated PV active power generation at a distribution feeder with limited sensor deployment. This model takes full consideration of major characteristics of masked load and PV production and thus leads naturally to predictive capability in real time. For larger areas and same density of observations, we expect this strategy to perform similarly and arguably better with more sophisticated models that can take advantage of more information. Nevertheless, the results presented in this study are limited by the availability of measurements and future studies should address larger areas if data becomes available.

Our novel framework can be naturally extended to several other directions, which we plan to investigate. For the solar generation predicted by partial irradiance measurements, improvements could be made such that we can accurately predict the irradiance jumps. On the masked load model and disaggregation side, further improvements could be made to the SDE model and computational framework. Variability on different time horizons could also be considered in the future for real applications.

REFERENCES

[1] J. Widén, M. Shepero, and J. Munkhammar, "On the properties of aggregate clear-sky index distributions and an improved model for spatially correlated instantaneous solar irradiance," *Solar Energy*, vol. 157, pp. 566–580, 2017.

[2] E. Vrettos, E. Kara, E. Stewart, and C. Roberts, "Estimating PV power from aggregate power measurements within the distribution grid," *Journal of Renewable and Sustainable Energy*, vol. 11, no. 2, p. 023707, 2019.

[3] N. Engerer and F. Mills, "KPV: A clear-sky index for photovoltaics," *Solar Energy*, vol. 105, pp. 679–693, 2014.
[4] S. Killinger, F. Braam, B. Müller, B. Wille-Haussmann, and R. McKenna, "Projection of power generation between differently-oriented PV systems," *Solar Energy*, vol. 136, pp. 153–165, 2016.
[5] F. Sossan, L. Nespoli, V. Medici, and M. Paolone, "Unsupervised disaggregation of photovoltaic production from composite power flow measurements of heterogeneous prosumers," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 9, pp. 3904–3913, 2018.
[6] V. Patel, D. A. Maldonado, and M. Anitescu, "Semiparametric estimation of solar generation," in *2018 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, Aug. 2018.
[7] T. Landelius, S. Andersson, and R. Abrahamsson, "Modelling and forecasting PV production in the absence of behind-the-meter measurements," *Progress in Photovoltaics: Research and Applications*, 2018.
[8] J. M. Bright, S. Killinger, D. Lingfors, and N. A. Engerer, "Improved satellite-derived PV power nowcasting using real-time power data from reference PV systems," *Solar Energy*, vol. 168, pp. 118–139, Jul. 2018.
[9] C. Roberts, E. M. Stewart, and F. Milano, "Validation of the Ornstein-Uhlenbeck process for load modeling based on μPMU measurements," in *Power Systems Computation Conference (PSCC)*. IEEE, 2016, pp. 1–7.
[10] G. M. Jónsdóttir and F. Milano, "Modeling solar irradiance for short-term dynamic analysis of power systems," to appear in 2019 IEEE Power & Energy Society General Meeting (PESGM). IEEE, 2019.
[11] A. W. Aryaputera, D. Yang, L. Zhao, and W. M. Walsh, "Very short-term irradiance forecasting at unobserved locations using spatio-temporal kriging," *Solar Energy*, vol. 122, pp. 1266–1278, 2015.
[12] J. Widén, M. Shepero, and J. Munkhammar, "Probabilistic load flow for power grids with high PV penetrations using copula-based modeling of spatially correlated solar irradiance," *IEEE Journal of Photovoltaics*, vol. 7, no. 6, pp. 1740–1745, Nov 2017.
[13] M. Sengupta and A. Andreas, "Oahu solar measurement grid (1-year archive): 1-second solar irradiance; Oahu, Hawaii (data)," NREL, Tech. Rep. DA-5500-56506, 2010.
[14] W. F. Holmgren, C. W. Hansen, and M. Mikofski, "pvlib python: a python package for modeling solar energy systems." *J. Open Source Software*, vol. 3, no. 29, p. 884, 2018.
[15] J. Bessac, E. Constantinescu, and M. Anitescu, "Stochastic simulation of predictive space–time scenarios of wind speed using observations and physical model outputs," *Annals of Applied Statistics*, vol. 12, no. 1, pp. 432–458, 2018.
[16] C. W. Gardiner, *Handbook of stochastic methods*. springer Berlin, 1985, vol. 3.
[17] B. M. Bibby and M. Sørensen, "Martingale estimation functions for discretely observed diffusion processes," *Bernoulli*, pp. 17–39, 1995.
[18] S. Peisert, R. Gentz, J. Boverhof, C. McParland, S. Engle, A. Elbashandy, and D. Gunter, "LBNL open power data," 2017.