Abstract—The wide variety of inverter control settings for solar photovoltaics (PV) causes the accurate knowledge of these settings to be difficult to obtain in practice. This paper addresses the problem of determining inverter reactive power control settings from net load advanced metering infrastructure (AMI) data. The estimation is first cast as fitting parameterized control curves. We argue for an intuitive and practical approach to preprocess the AMI data, which exposes the setting to be extracted. We then develop a more general approach with a data-driven reactive power disaggregation algorithm, reframing the problem as a maximum likelihood estimation for the native load reactive power. These methods form the first approach for reconstructing reactive power control settings of solar PV inverters from net load data. The constrained curve fitting algorithm is tested on 701 loads with behind-the-meter (BTM) PV systems with identical control settings. The settings are accurately reconstructed with mean absolute percentage errors between 0.425% and 2.870%. The disaggregation-based approach is then tested on 451 loads with variable BTM PV control settings. Different configurations of this algorithm reconstruct the PV inverter reactive power timeseries with root mean squared errors between 0.173 and 0.198 kVAR.

Index Terms—Disaggregation, data-driven control, feature extraction, inverters, optimization , reactive power control, system identification.

I. INTRODUCTION

W ITH the proliferation of behind-the-meter (BTM) solar photovoltaics (PV), the use of advanced inverters with reactive power control technology is expanding in tandem. Recent advances in distribution analysis tools and algorithms have culminated in efficient and accurate model-based analysis methods for determining the impact of these inverter technologies on distribution feeders [1]–[5]. However, the data used to drive these algorithms are often based on incomplete feeder models that rely on manual data entry and topology configuration. In particular, a priori knowledge of the inverter settings is typically assumed or derived from the model. Records of these settings may be out-of-date or completely unavailable, and the manual entry process for the collection and maintenance of these data is highly prone to error [6], [7].

In contrast with utility-scale PV, the impact of BTM PV inverters is typically invisible to utilities. Even if the presence of an advanced inverter is known, distribution engineers typically do not have the ability to remotely monitor BTM inverters or access the installation premises to verify the inverter control settings of grid-connected solar PV installations. Furthermore, with the advent of the IEEE 1547-2018 standard update [8], the deployment of advanced BTM solar PV inverters with reactive power control and voltage regulation capabilities will increase. Notably, there is a broad range of control settings permitted by the standard. For these reasons, it is also likely that utilities of the future may face limited information about these advanced inverters and their control settings for BTM PV. As a result, engineers may be unable to accurately predict their effects on a distribution network. It is also likely that limited information about the specific operating conditions of BTM inverter-based resources is attainable. Control parameters of the system such as the power factor, curtailment, and voltage control settings may be completely unavailable. These problems are an active area of research [4], [9]–[15].

Fortunately, data from advanced metering infrastructure (AMI) are becoming increasingly available [16], presenting a significant number of opportunities for the creation of entirely data-driven methods to reconstruct the control settings of solar PV inverters. These contribute to existing AMI-derived estimation methods, which have been developed for estimating the PV hosting capacity of distribution systems [7], [17], disaggregating the active power contribution of solar PV [11]–[15], and many other applications.

However, one of the key obstacles in implementing entirely data-driven analysis methods is the detection and estimation of inverter control settings, particularly those for BTM PV. Currently, this problem has been unaddressed in the literature. The contributions of this paper are methods to reconstruct these settings, even if they are unknown. In this paper, we build up two related mathematical models for achieving this reconstruction from heavily corrupted, aggregated load data. The proposed framework uses power system physics with data-driven estimation and optimization methods. The two models together
form the first approach to solve this problem. These estimation techniques allow engineers to acquire information about the reactive power behavior of BTM PV interconnections without the use of a distribution circuit model. The method also does not require the input of the customer or the deployment of field technicians, which allows for easier power system planning and power quality monitoring. The techniques could also be used in the form of a non-intrusive method to monitor for changes in these settings. Knowledge of these changes is critical for feeder-wide control and optimization tasks such as Volt-VAR optimization (VVO) [18]. It may also be useful for other DER management tasks such as acquiring operational impacts with other regulating devices and impacts to system protection. The methods implicitly provide information about the operational impacts of the inverter, which simultaneously reconstructs the control curve. Correspondingly, this allows for the user to infer the actuation of the control states, thereby reconstructing the PV reactive power timeseries.

In summary, the methods provide access to unknown device-level PV inverter information that is important for other distribution system planning and analysis tasks, such as the computation of PV hosting capacity and ensuring model fidelity. We achieve this with only historical voltage, active power, and reactive power AMI data, without a need for a distribution circuit model. Improved knowledge and visibility of inverter settings will increase utility confidence in their studies, which has the potential to reduce the need for BTM PV interconnection regulations and restrictions, enabling a swifter transition to a sustainable energy portfolio.

II. BACKGROUND

A. Advanced Inverter Reactive Power Control Settings

Modern inverters can operate in several different control modes. These modes are defined by characteristic curves that dictate the injection or absorption of reactive power at the point of common coupling (PCC) of the DER. An inverter with these capabilities is commonly referred to as a “smart inverter” or “advanced inverter”. For consistency, we adopt the latter terminology.

The simplest and most common is the fixed power factor mode. This mode is favored for its simplicity and interpretability. In this control mode, the inverter controller is designed to inject or absorb reactive power such that a constant power factor is always maintained at the PCC. To achieve this control objective; as the PV active power output is increased, the reactive power output will increase proportionally to the power factor [5]. This behavior is illustrated in Fig. 1. The default setting is unity power factor control [8], in which case the voltage impacts of the PV generator will not be mitigated [19].

In contrast with the default unity power factor control mode, nonunity power factor control, and another control mode, autonomous Volt-VAR, can mitigate distributed generator voltage impacts. These control modes have received significant attention from both researchers and industry due to their potential to increase PV hosting capacity. Notably, it has been shown that Volt-VAR control yields a superior reduction in voltage profile volatility in comparison with both fixed unity and nonunity power factor control, while requiring a smaller amount of reactive power injections [5], [20], [21]. When an inverter is operating with Volt-VAR control, the reactive power injection or absorption of the PV system’s inverter is controlled according to one of the following control states:

\[
S_{\text{inv}} = \begin{cases} 
R_1, & V_{pcc} \leq V_{P1}, \\
R_2, & V_{P1} \leq V_{pcc} \leq V_{P2}, \\
R_3, & V_{P2} \leq V_{pcc} \leq V_{P3}, \\
R_4, & V_{P3} \leq V_{pcc} \leq V_{P4}, \\
R_5, & V_{P4} \leq V_{pcc}, 
\end{cases}
\]  

(1)

where the PCC voltage \(V_{pcc}\) is used as the input signal to the controller to determine the state. This behavior is illustrated in Fig. 2, and an example of the state regions corresponding with the control curve is shown in Fig. 3. Research and industrial interest in Volt-Watt control have also grown, primarily serving as a “backup” to reactive power control modes. Currently, reactive
power control modes—specifically, fixed power factor and Volt-VAR—have emerged as a primary focus of industry and local regulators [19]. For this reason, and for brevity, we leave the detection and estimation of Volt-Watt and other control modes to future work.

B. Net Load Disaggregation

Net load disaggregation is a well-studied topic in the power engineering literature. Reference [22] considered one of the earliest examples of structured net load disaggregation, with the intent of identifying the presence of different household appliances from variations in voltage and current waveforms. More recently, [23] built on this idea with a non-negative sparse coding approach, integrating concepts from the blind-source separation literature in the signal processing field. Using a dictionary of example signals for the energy consumption of various device types, they find representations for aggregate energy demand in a basis of test signals. Later works such as [24] continued to use sparse coding with various enhancements. Even more recently, game-theoretic approaches, such as [13] have appeared, in addition to numerous unsupervised methods such as [12], [25], [26]. Furthermore, reference [27] developed an online learning approach for disaggregation.

A particular application of this topic that has received considerable attention is the estimation—or disaggregation—of BTM solar PV parameters from net load AMI data, i.e., the total energy summation of the customer’s demand and PV generation. There has been a considerable amount of work in both PV location, state, size, and presence detection [9], [28], [29]. Other works have explored the direct disaggregation of solar generation signals and “native” demand signals solely from net load AMI measurements [11]–[15], [30]. These are closely related to our work.

Notably, there has been a growing interest in solving disaggregation problems with a completely data-driven approach, i.e., without considering the system feeder model. Some solutions to this problem are dependent on localized feeder information, with newer works pursuing methods increasingly decoupled from localized information, [12], [13], [30]. A particularly popular method is to use “exemplars” or proxy measurements of solar PV generation to provide a basis for the estimation or to extract features of interest. In [11] a multiple linear regression approach based on proxies for localized irradiance information from a nearby metered PV system was established. A similar approach was generalized to online estimation in [15]. It has also been shown in [13] that economic data can be a useful feature input for the estimation.

Probabilistic methods and unsupervised machine learning methods have shown promise for removing the need for model-derived information. Works such as [26], [28], [31] develop data-driven methods using a variety of input data assumptions to estimate size, azimuth, or tilt information for BTM PV installations in an unsupervised fashion. An estimated probability density of nighttime and daytime active power from loads without PV on the feeder was used to disaggregate solar generation in [12]; a similar concept was developed in [32]. Reference [12] then uses a weighted combination of the exemplars to subtract away from the nighttime net AMI measurements, where the weights are chosen such that they maximize a log likelihood function derived from a learned generative distribution. In all these works, the focus is solely on disaggregating the active power production of the PV from the customer’s demand. In this paper, we take inspiration from these results and construct an approximate generative distribution for the native reactive power. We also improve upon the methodology of [12] by incorporating a metric for estimating the dimension of the model, i.e., the number of Gaussian mixtures that maximally captures the information of the load population without PV.

C. Advanced Metering Infrastructure (AMI)

The growing rate of interconnections of BTM PV systems has also come with a growth in the use of AMI data, with many applications appearing in the literature [7]; including load forecasting [10], [33] and estimating the error of the meter itself [34].

In our analysis, we consider an electric power system with $L$ loads containing BTM PV operating under reactive power control. For each load $l = 1, \ldots, L$, we assume access to a historical multivariate timeseries AMI dataset $D_l$ containing measurements of net load voltage, active power, and reactive power demands over a time horizon $M$:

$$D_l := \{(v_{l,t}^{\text{net}}, p_{l,t}^{\text{net}}, q_{l,t}^{\text{net}})\}_{t=1}^M.$$  \hspace{1cm} (2)

Hereafter, we drop the subscript $l$ for brevity. For a load with BTM PV, the net demand measured by the AMI, $p_{l,t}^{\text{net}}$ and $q_{l,t}^{\text{net}}$, will be additive combinations of the “native” active and reactive demands of the load, which we will denote as $p_l$ and $q_l$; the injection or absorption of active and reactive power by the solar PV and its inverter system, denoted as $p_l^{\text{pv}}$ and $q_l^{\text{pv}}$; and random measurement noise, denoted as $\epsilon$.

Remark (Sign Convention): Throughout the paper, we adopt the convention that injections from a node into the distribution system are positive and flows from the distribution system into the load through the AMI are negative.

Considering these conventions, native demand is a negative value. Therefore, the net AMI demands are then:

$$p_{l,t}^{\text{net}} = p_l + p_l^{\text{pv}} + \epsilon \quad t = 1, \ldots, M,$$

$$q_{l,t}^{\text{net}} = q_l + q_l^{\text{pv}} + \epsilon \quad t = 1, \ldots, M.$$  \hspace{1cm} (3)

The above convention significantly simplifies notation throughout the paper. Importantly, $p_{l,t}^{\text{net}}$ and $q_{l,t}^{\text{net}}$ are the negative of what will typically be read by AMI in practice.

The net metered demands in (2) and (3) are assumed to be sampled at 15-minute granularity. Furthermore, we assume that measurement errors due to sensor manufacturing defects are distributed according to a standard normal: $\epsilon \sim N(0, \sigma^2)$

We will use the symbol $\mathcal{K}$ to denote the set of all feasible control parameters $\Theta$ for the mode. The contribution of the PV system, $q_l^{\text{pv}}$, to the net reactive power measurement seen by the AMI, $q_{l,t}^{\text{net}}$, is modeled by a control curve $\phi_{\Theta}$, where $\Theta \in \mathcal{K}$ is a parameter vector describing the control settings. To be specific, in the case of the Volt-VAR control mode, this control curve will be a piecewise-linear function of the PCC voltage $v_{\text{pcc}}^{\text{pv}}$, whereas for the fixed power factor control, we will have a linear function
of the PV active power, i.e.: 
\[ q_{t}^{pv} = \begin{cases} \phi_\theta(v_{t}^{pcc}) & \text{if Volt-VAR,} \\ \phi_\theta(p_{t}^{pv}) & \text{if Power Factor.} \end{cases} \]  
(4)

III. CONTROL CURVE BASIS FUNCTIONS

Given the control parameter \( \Theta \), we can reconstruct the control curve and therefore the PV inverter’s reactive power response. In this section, we build up parameterized representations of common reactive power control curves to allow us to estimate this parameter.

A. Fixed Power Factor Control

In fixed power factor control mode, the PV system’s reactive power response \( q_{t}^{pv} \) can be characterized by a slope parameter \( m = \frac{\Delta q}{\Delta p} \), which is also known as the inverter’s sensitivity of reactive power injections to active power injections. If we had direct measurements of the PV system’s active power injection, we could reconstruct [18] the PV system’s reactive power injection as:

\[ q_{t}^{pv} = \phi_\theta(p_{t}^{pv}) := mp_{t}^{pv} = \frac{\Delta q}{\Delta p}p_{t}^{pv}. \]  
(5)

This sensitivity defines a simple line in the complex power plane, which maintains \( \cos(\phi_V - \phi_I) \) for any value of \( p_{t}^{pv} \), as shown in Fig. 1. Finally, the power factor setting \( pf \) can then be easily related to this slope via trigonometry and the definition of the power factor:

\[ pf := \cos(\phi_V - \phi_I) = \cos\left(\arctan\left(\frac{\Delta q}{\Delta p}\right)\right). \]  
(6)

Additionally, the setting can be recovered with a known PV apparent power measurement \( \|s_{t}^{pv}\|_2 \):

\[ pf = \frac{p_{t}^{pv}}{\|s_{t}^{pv}\|_2} = \phi_\theta\left(\frac{\Delta q}{\Delta p}\right) = \frac{q_{t}^{pv}}{\|s_{t}^{pv}\|_2}. \]  
(7)

B. Volt-VAR Control

For autonomous Volt-VAR control, the control input is the PCC voltage, assumed to be measured directly by the AMI. Inspecting the IEEE 1547-2018 standards [8] on the Volt-VAR characteristics curve, we can construct a parameterized function of PCC voltage \( v_{t}^{pcc} \), for a general\(^1\) representation of the reactive power response of a Volt-VAR controller as:

\[ \phi_\theta(v_{t}^{pcc}) := \begin{cases} Q_{1} & v_{t}^{pcc} \leq V_{1}, \\
\frac{Q_{2} - Q_{1}}{V_{2} - V_{1}} \cdot v_{t}^{pcc} + b_{1} & V_{1} < v_{t}^{pcc} < V_{2}, \\
0 & V_{2} \leq v_{t}^{pcc} \leq V_{3}, \\
\frac{Q_{4} - Q_{3}}{V_{4} - V_{3}} \cdot v_{t}^{pcc} + b_{2} & V_{3} \leq v_{t}^{pcc} \leq V_{4}, \\
Q_{4} & v_{t}^{pcc} > V_{4}, \end{cases} \]  
(8)

where \( b_{1} = Q_{1} - V_{1} \frac{Q_{2} - Q_{1}}{V_{2} - V_{1}} \) and \( b_{2} = Q_{3} - V_{3} \frac{Q_{4} - Q_{3}}{V_{4} - V_{3}}. \) By finding the best fit parameter \( \hat{\theta} \) for this basis function, we can reconstruct the control setting for load \( l \) by fitting the curve to the variables \( q_{t}^{net} \) and \( v_{t}^{pcc} \) in \( D_{l} \). Note that the parameter vector \( \Theta \) is \( N \) pairs of voltage and reactive power settings \( \theta_{i} := [P_{i}, Q_{i}]^{T} \), \( i = 1, \ldots, N \) that define the deadband, saturation, and dynamic VAR regions for the inverter:

\[ \Theta := [\theta_{1}^{T}, \ldots, \theta_{N}^{T}]^{T} = [V_{1}, Q_{1}, \ldots, Q_{N}, Q_{N}]^{T}. \]  
(9)

C. Feasible Control Curve Set

We use the symbol \( K \) to represent the set of parameter vectors \( \Theta \) that define a valid control curve for the mode of interest. For Volt-VAR control, this is defined by a set of convex constraints on the \( N \) knots under study. This can be expressed through a set of inequality constraints that are functions of the elements of \( \Theta \), as shown in (10-12), where the subscripts \( ub \), \( i \) and \( lb \), \( i \) denote the upper and lower bounds respectively for the decision variable \( i \). \( \Delta V \) represents the absolute value of the minimum dynamic VAR region width, i.e., the minimum distance between \( V_{1}, V_{2} \) and \( V_{3}, V_{4} \).

\[ V_{\text{lb},i} \leq V_{i} \leq V_{\text{ub},i} \forall i = 1, \ldots, N, \]  
(10)

\[ Q_{\text{lb},i} \leq Q_{i} \leq Q_{\text{ub},i} \forall i = 1, \ldots, N, \]  
(11)

\[ \{(V_{2} - V_{1}), (V_{4} - V_{3})\} \geq \Delta V. \]  
(12)

The above system of inequality constraints can be expressed compactly through the system \( A \Theta \leq b \). The exact bounds\(^2\) on each parameter can be found in the IEEE 1547-2018 standard [8].

For fixed power factor control, a residential load can be expected to be inductive. Therefore, \( K \) represents the reactive power sensitivities, or slopes, that meet this condition, i.e., all real numbers that are less than or equal to zero.

IV. CONTROL CURVE FITTING

The high-level goal of reconstructing inverter control settings is to find an optimal estimate \( \hat{\Theta} \) for the true control parameter \( \Theta \) in (4). For both fixed-power-factor and Volt-VAR, different methods can be used to preprocess the measurements in \( D_{l} \), such that the historical data approximates the underlying behavior of \( q_{t}^{pv} \) in the net load AMI signal (3). There are many efficient solvers for the programs described in this section, which we have used extensively throughout our work; in particular, [35], [36].

A. Regression Methods

1) Power Factor Control: Estimating a fixed power factor control parameter can be achieved by applying a simple linear regression in the complex power plane. The control setting is extracted by finding the reactive power sensitivity (slope) to active power injections that best matches the historical data relationships. If we can find a filtering function \( f(\cdot) \) that returns filtered active and reactive power timeseries vectors \( \hat{p}_{t}^{pv} \) and \( \hat{q}_{t}^{pv} \) where the contributions of the native load power to these measurements are sufficiently small, a power factor control setting can be estimated with ordinary least-squares regression, achieving the closed-form solution:

\[ \hat{\Theta} = (P^{T}P)^{-1}P^{T}q^{pv}. \]  
(13)

\(^1\)This paper assumes that the deadband region \( (V_{2}, V_{3}) \) is centered at zero kVAR, which is most common in practice, but need not be true in general.

\(^2\)Note that constructing these constraints according to the standard assumes that the user has knowledge of the nominal voltage of the load, which the bounds are constructed from. If such information is not available, the unconstrained least-squares method (16) can be used.
are the actuated control settings to be extracted from the raw dataset.

\[
\Theta := \left[ \frac{\Delta q}{\Delta \theta} \right]_T, \quad \text{and} \quad P := \begin{bmatrix} -(p_{pv})^T - 1^T \end{bmatrix}_T. \tag{14}
\]

Note that for the typical lagging power factor case, the sensitivity is less than or equal to zero, and the complex power line is fixed at the origin, so we can require that \( b = 0 \). Then, using the trigonometric relationship in (6), the power factor setting can be estimated with the regression slope:

\[
\hat{p}_f = \cos \left( \arctan \left( \frac{\Delta q}{\Delta \theta} \right) \right). \tag{15}
\]

Other typical loss functions such as the Huber loss or the \( \ell_1 \) distance can be used to achieve robust performance, as described in [37].

2) Volt-VAR Control: For residential BTM PV installations, the contribution of the native reactive power demand \( q_{n}^{\text{nat}} \) to the net signal (3) may be large at certain time points, resulting in the true control curve being significantly corrupted, and very difficult to estimate with net load AMI data.

If we can find a filtering function \( f \) such contribution of \( q_{n}^{\text{nat}} \) to \( f(q_{n}^{\text{nat}}) \) is sufficiently small, \( q_{n}^{\text{nat}} \) can effectively be lumped into the noise term \( \epsilon \), i.e., \( f(q_{n}^{\text{nat}}) = q_{n}^{\text{pv}} + \epsilon \). In this “naïve” approach, the optimal \( \Theta \in \mathbb{R}^{2N} \) is the solution to a non-linear least-squares program with the observed voltage and reactive power timeseries as data inputs:

\[
\min_{\Theta \in \mathcal{K}} \sum_{t=1}^{M} \left[ f(q_{n}^{\text{nat}})_t - \theta_{\hat{\phi}}(v_{t}^{\text{pcc}}) \right]^2. \tag{16}
\]

Furthermore, real AMI datasets with Volt-VAR—even over long time horizons—often primarily contain samples in the absorption and deadband regions. An example of such a dataset is depicted in Fig. 4. Here, \( \Theta \) can be simplified to finding the best \( \theta_3 = [V_3, Q_3]^T \) and \( \theta_4 = [V_4, Q_4]^T \):

\[
\begin{bmatrix} \hat{\theta}_3 \hat{\theta}_4 \end{bmatrix} = \arg \min_{\theta_3, \theta_4 \in \mathbb{R}^2} \sum_{t=1}^{M} \left[ f(q_{n}^{\text{nat}})_t - \hat{\phi}(v_{t}^{\text{pcc}}|\theta_3, \theta_4) \right]^2, \tag{17}
\]

where \( \hat{\phi} \) is a truncated Volt-VAR basis function and the curve parameters \( \theta_3, \theta_4 \in \mathbb{R}^2 \) are the actuated control settings to be estimated. The simplified Volt-VAR basis function can be written as:

\[
\hat{\phi}(v_{t}^{\text{pcc}}|\theta_3, \theta_4) = \begin{cases} 0 & v_{t}^{\text{pcc}} \leq V_3, \\ \frac{V_{4} - V_{3}}{Q_{4}} v_{t}^{\text{pcc}} & V_{3} \leq v_{t}^{\text{pcc}} \leq V_{4}, \\ \frac{V_{4} - V_{3}}{Q_{4}} & v_{t}^{\text{pcc}} > V_{4}. \end{cases} \tag{18}
\]

Note that the reactive power intercept \( b \) is again defined as \( b := -V_3 \frac{Q_4 - Q_3}{V_4 - V_3} \). The intercept \( b \) will go to zero in the case that the control curve does not have a deadband region. Furthermore, Volt-VAR curve definitions are often defined symmetrically about a reference voltage. In such instances, this low-dimensional method can be used, provided the user has a degree of confidence that the symmetry convention is followed at the load of interest.

B. Filtering Native Reactive Power

In order to directly fit the curve, it is necessary to form a filter \( f(q_{n}^{\text{net}}) \) that returns a preprocessed reactive power timeseries \( q_{n}^{\text{pv}} \in \mathbb{R}^{M'} \), which are the \( M' \) samples that are most favorable for the curve fitting in (13), (16), and (17). This approximation for the PV system reactive power contribution \( q_{n}^{\text{pv}} \) can then become the target for the estimation.

For Volt-VAR control, a choice for this filter becomes clear if the observations are labeled according to whether the measurement was taken during the nighttime or daytime, as shown in Fig. 5. During the nighttime, the PV inverter is generally set to turn off and to not provide voltage regulation or contribute to the reactive power measurements. Additionally, as shown in Fig. 4, observations with the most positive net load active power values more closely follow the true control curve.

Noting this trend, the user can find a “percentile threshold” for the net active power demand to filter the measurements. We denote this percentile as \( \delta \). When an observation in the AMI timeseries meets the condition \( p_{n}^{\text{net}} < p_{\delta}^{\text{net}} \), we filter out the observation, where \( p_{\delta}^{\text{net}} \) is the \( \delta \)-th percentile net active power observation. After filtering, the observations are then primarily composed of the PV contribution, i.e., \( |p_{n}^{\text{pv}}| > |p_{n}^{\text{net}}| \) and \( |q_{n}^{\text{pv}}| > |q_{n}^{\text{net}}| \). This “naïve filter” can be summarized as:

\[
f_{\delta}(q_{n}^{\text{net}}) = \{ q_{n}^{\text{net}} : p_{n}^{\text{net}} > p_{\delta}^{\text{net}} \}. \tag{19}
\]

For fixed power factor control, a similar filtering method can be applied over a different variable. As power factor control seeks to reduce overvoltages, the subset of the historical samples with the highest voltage measurements are likely to have net reactive power measurements that are primarily composed of
the inverter contribution, approximating the underlying control
curve, as discussed in [37], [38].

Experimentally, we have found that $75 \leq \delta \leq 95$ often ex-
poses both Volt-VAR and power factor curves in real AMI
datasets. The example dataset in Fig. 4 and Fig. 5 is filtered with
$\delta = 90$ in Fig. 6. In Section VI-A, we discuss the application of
the least-squares methods to the filtered datasets. These methods
are practical and convenient because they can be optionally
unconstrained, with good results.

V. REACTIVE POWER DISAGGREGATION

The least-squares methods discussed in the previous sections
have several limitations. In particular:

1) The selection of the filter threshold $\delta$ must be set by the
user. While tuning methods could be used to automate
this task, they must be repeated for every individual load,
which may be computationally costly for large systems.

2) Depending on the behavior of the individual load, differ-
ent curve regions, shown in Fig. 3, may be historically
actuated. The optimal percentile may vary significantly,
or the full curve may not be exposed at all.

In this section, we present a more general alternative to the
filtering method proposed in the first section that seeks to dis-
aggregate the native reactive power and PV-attributable reactive
power, eliminating the need for the tuning of the percentile cutoff
hyperparameter.

Similar to the methodologies applied to active power disag-
gregation in [12], [14], [32], we use a Gaussian Mixture Model
(GMM) to construct an estimate for the joint distribution of
nighttime and daytime native reactive power demands at 1-week
granularity. The data are nighttime and daytime reactive power
measurements from 678 1-year, 15-minute granularity AMI
datasets of real-world loads without BTM PV. For computational
speed and to ensure that the number of nighttime and daytime
samples are equal, the demands are summed over a 1-week period
per month to form a lower-resolution sample. All observations from
the 678 loads are shown in Fig. 7. We denote this resampling
as $x = T\{x\}$ where $T : \mathbb{R}^{M} \rightarrow \mathbb{R}^{M'}$ is a function that maps
an $M$-dimensional timeseries vector $x$ to an $M'$-dimensional
timeseries $\tilde{x}$.

Crucially, the reactive power AMI measurements for these
loads must be solely native reactive power. In practice, the user
may need to employ a PV presence detection algorithm to verify
that the training data are from loads without PV.

On loads with PV, this model can then be used to provide an
estimate for the portion of daytime net reactive power demand
that is attributable to PV by subtracting away the contribution
of a particular hypothesized control curve.

A. Gaussian Mixture Model

Throughout this section we consider pairwise nighttime and
daylight demands of the form:

$$q_{m} := [q_{m,n}, q_{d,m}]^T \quad m = 1, \ldots, M'$$

where $m$ is the augmented time index, $M'$ is the augmented
number of samples (in this case, 52 weeks), and $n, d$ denote
daylight and nighttime values respectively. Constructing a GMM
consists of finding an optimal set of $K$ multivariate Gaus-
sian densities each with mean vectors and covariance matrices
$\{\mu_k, \Sigma_k\}_{k=1}^K$ to approximate the training data’s generative
distribution. We write $\gamma_k := \{\mu_k, \Sigma_k\}$ to represent the set of
estimated parameters for each distribution $g_k$. Each density $g_k$
is then:

$$g_k(\tilde{q}_m | \gamma_k) = \frac{1}{2\pi \det(\Sigma_k)^{\frac{1}{2}}} e^{-\frac{1}{2}(\tilde{q}_m - \mu_k)^T \Sigma_k^{-1}(\tilde{q}_m - \mu_k)}$$

(21)

where $\mu_k \in \mathbb{R}^2$ is the mean vector for density $k$ and $\Sigma_k \succ 0$
is the covariance matrix for density $k$. Each density $g_k$ is then
assigned a positive scalar weight $\alpha_k > 0$.

Hereafter, let $\gamma := [\gamma_1, \ldots, \gamma_K]^T$ be the vector of parameter collections for the $K$ densities. We then model the joint
distribution of daytime and nighttime native reactive samples for the
feeder as the weighted sum of the $K$ multivariate Gaussian densities:

$$q_{m}^{nat} = [q_{m,n}^{nat}, q_{d,m}^{nat}]^T \sim \mathcal{N}(\tilde{q}_m^{nat} | \gamma) := \sum_{k=1}^K \alpha_k g_k(\tilde{q}_m^{nat} | \gamma_k)$$

(22)

where the weight argument $\alpha := [\alpha_1, \ldots, \alpha_K]^T$ is implied.

Constructing the GMM can be achieved through the optimization
problem (23), which returns the model parameters $\gamma$ and
weights $\alpha$ that maximize the data log-likelihood given a particu-
lar native reactive power GMM, taking care to ensure that the
granularities are converted:

$$\arg\max_{\alpha \succ 0, \Sigma \succ 0} \sum_{m=1}^{M'} \log \left( \sum_{k=1}^K \alpha_k g_k(q_{m}^{nat} | \gamma_k) \right)$$

(23)
Note that an advantage of this method is that we can quantify the minimum amount of training samples $M'$ that we need so that the total variation distance from the true distribution is no more than $\epsilon$. Bounds for the minimum number of samples are available in [39]. In our case, this bound is $M' \geq \frac{4K}{\epsilon^2 \log(1/\epsilon)}$.

B. Construction and Training

Using the transformation $T$, we resample the timeseries features in (2) for loads without PV up to the chosen augmented sample size $M'$. This results in the matrix $Q \in \mathbb{R}^{M' \times 2}$ where the rows are nighttime and daytime reactive power measurements $q^\text{net}_{n,m}, q^\text{net}_{d,m}$. Critically, these are composed solely of native contributions by construction.

$$Q := \begin{bmatrix} q_{1}^T \hspace{1cm} q_{M'}^T \end{bmatrix} = \begin{bmatrix} q^\text{nat}_{n,m} & q^\text{nat}_{d,m} \\ q^\text{nat}_{n,M'} & q^\text{nat}_{d,M'} \end{bmatrix}. \tag{24}$$

As before, $n$ denotes a nighttime sample and $d$ a daytime sample in (24). It is standard to use the expectation maximization (EM) algorithm first described in [40] to solve problems such as (23). The EM algorithm is not theoretically guaranteed to converge to the global optimum, but it is known to converge to a local optimum [40]–[42].

An important consideration for mixture models is the selection of the number of Gaussian components $K$, as it must be specified by the user [40], [41]. One choice for a metric to make this decision is the Bayesian Information Criterion (BIC), which we formulate based on [43] as:

$$\text{BIC}_{K_i} = (6K_i - 1) \log(M') - 2\ell(Q|\gamma_i), \tag{25}$$

where $\gamma_i$ is the locally optimal solution to (23) for a model with $K_i$ number of components, and $M'$ is the total number of samples at the chosen timestep granularity. The optimal number of components $K_i$ that minimizes the BIC (25) is then selected among a finite range $1, \ldots, K_{\text{max}}$, i.e.,

$$K_i = \arg \min_{K_i \in \{1, \ldots, K_{\text{max}}\}} \text{BIC}_{K_i}. \tag{26}$$

The corresponding parameters $\gamma_i$ and weights $\alpha_i$ are then held fixed to form the trained model. We drop the subscript $i$ hereafter. It is discussed in [43] that the BIC is a valid criterion for selecting the number of parameters of a model if the distribution behind the data is included in the set of candidate models. It was discussed in [44] that data that are distributed according to a member of the exponential family meet this condition for GMMs. Experimentally, we have found that the nighttime and daytime native reactive power samples for the feeder studied are well approximated by a joint distribution in the exponential family, and that (26) converges to 6, where the 6 components correspond to season transitions.

C. Disaggregation and Curve Reconstruction

Suppose that we are studying another load on the feeder with BTM PV. Once the GMM has been trained, we then use the model to define a likelihood function (27) for a $M' \times 2$ matrix of the form (24). Note that the rows are now net reactive power pairs $[q^\text{net}_{n,m}, q^\text{net}_{d,m}]$ for this load (at $M'$-granularity):

$$\mathcal{L}(\hat{Q}|\gamma) = \prod_{m=1}^{M'} \sum_{k=1}^{K} \alpha_k g_k(q^\text{net}_{n,m}, q^\text{net}_{d,m} | \gamma_k), \tag{27}$$

and the equivalent log-likelihood function is (28):

$$\ell(\hat{Q}|\gamma) = \sum_{m=1}^{M'} \sum_{k=1}^{K} \alpha_k g_k(q^\text{net}_{n,m}, q^\text{net}_{d,m} | \gamma_k). \tag{28}$$

Let $\Theta$ denote a hypothesized parameter vector. Note that we can construct the estimated native daytime reactive power demand for a load given the hypothesis $\Theta$:

$$q^\text{nat}_{d,m} = q^\text{net}_{d,m} - g_m = q^\text{net}_{d,m} - T[\phi(\mathbf{v})]_{[m]} \tag{29}$$

With our fitted model, our problem then becomes finding the best control parameter $\Theta$ that maximizes the likelihood of the data given the hypothesized $\Theta$, which is the solution to the following optimization problem:

$$\text{maximize } \ell(Q|\Theta, \gamma) \text{ subject to } \Theta \in \mathcal{K}, \tag{30}$$

where $\mathcal{K}$ is the set of feasible parameters for the control mode. For example, consider the Volt-VAR control mode, and recall that we can encode this set as inequality constraints such as those mentioned in Section III-C:

$$\mathcal{K} = \{ \Theta | \Theta \in \mathbb{R}^{2N} : A\Theta \leq b \}. \tag{31}$$

For notational convenience, we rewrite the objective function in the program (30) as a negative log likelihood, resulting in the minimization problem

$$\text{minimize } _{\Theta} \left[ -\sum_{m=1}^{M'} \log p(q^\text{net}_{m} | \Theta, \gamma) \right] \text{ subject to } \tag{32a}$$

$$A\Theta \leq b. \tag{32b}$$

Expanding out the problem (32) and combining with (29) yields the final problem (33).

$$\text{minimize } _{\Theta} \left[ -\sum_{m=1}^{M'} \log \left( \sum_{k=1}^{K} \Omega_{m,k} \right) \right] \text{ subject to } \tag{33a}$$

$$\Omega_{m,k} = \alpha_k g_k(q^\text{net}_{n,m}, q^\text{net}_{d,m} - T[\phi(\mathbf{v})]_{[m]} | \gamma_k), \tag{33b}$$

$$A\Theta \leq b. \tag{33c}$$

Notably, the solution $\hat{\Theta}$ to the programs (16) and (33) can be used to disaggregate the native and PV reactive power demands at the original timeseries granularity:

$$\hat{q}^\text{nat}_t = q^\text{net}_t - \hat{\phi}(\mathbf{v}) \quad \forall t \in \mathcal{M}_{\text{day}}, \tag{34}$$

where $\mathcal{M}_{\text{day}} \subset \{1, \ldots, M\}$ are the daytime points in the time horizon. See the Appendix if the inverter operates at night.

VI. EXPERIMENTS

We construct numerous semi-synthetic multivariate timeseries datasets in the form of (2) using actual native load active and reactive measurements from an electric utility. Similarly, the BTM PV active power injections are acquired from actual residential PV system measurements at 15-minute granularity for a year. We simulate the system’s reactive power response and voltage impact on the distribution network to which it is connected using a quasi-static timeseries (QSTS) simulation.

We partition the datasets into four different load categories:
we compute normalized versions of each load’s RMSE value, expressed as the percentage \( NRMSE = \frac{RMSE}{Q_{\text{max}} - Q_{\text{min}}} \times 100 \), where \( Q_{\text{max}} \) and \( Q_{\text{min}} \) are the maximum and minimum reactive power injections for the inverter of the load being evaluated. Note that we assume access to day-night transition times for simplicity. If the rough location of the distribution system is known, sunrise and sunset times are easy to obtain for each day using weather data. If the user does not have access to this information, there are many existing methods to detect time-series change points, such as [29].

### A. Least-Squares Methods

In designing our algorithm to solve the least-squares programs in (16) and (17), we use the sequential least-squares (SLSQ) programming solver introduced in [45] and implemented for Python by the SciPy project [36].

For Volt-VAR control, we test the least-squares models (16) and (17) with the first group of loads. All 701 loads have identical per-unit Volt-VAR settings and a nominal voltage of 240 V, but the differing PV sizing results in correspondingly different maximum reactive power capabilities that define the saturation parameters \( Q_1 \) and \( Q_4 \) [8].

Using \( \delta = 80 \), we estimate the rightmost curve parameters \( \theta_3 \) and \( \theta_4 \) with the unconstrained nonlinear least-squares program (17), — denoted as “UNNLSQ” — and the constrained version (17) — denoted as “CNLLSQ”. The performance of the two are shown and compared in Table I and Fig. 9.

We note that the algorithm could be used in tandem with a PV size estimation method such as [28], [29], [31], in which case we could impose near-exact constraints on \( Q_3 \) and \( Q_4 \). We make no assumptions about the size of the system, and instead simply treat the minimum observation in \( f_0(q^{\text{act}}) \) as a conservative estimate for the PV size.

Twelve customers out of the 701 have insufficient voltage volatility over the year available to allow for the use of this simple approach, i.e., the majority of samples are in the deadband for these loads. This is a key limitation of the least-squares methods and corroborates the need for the more advanced algorithms developed.

For the fourth dataset group with fixed power factor control, we use a regularized version of the least-squares approach and filter with respect to the 90th percentile voltage measurements. The results of this experiment are illustrated in Fig. 10. More details about this experiment and many others, including other loss functions, are described in [37], [38].
Fig. 9. Curve setting reconstruction results for the constrained nonlinear least-squares algorithm applied to dataset category 1. The top shows the absolute percentage errors for the estimated parameters, while the bottom shows the residuals between the estimated and true parameters for each load with BTM PV.

Fig. 10. Scatterplot of estimated power factor vs. true power factor for the power factor control setting estimation approach in [37].

B. Disaggregation Model

We implement the MLE algorithm described in Section V using interior-point methods [36], [46], and apply the algorithm in two configurations. The first assumes that $V_{ref}$ is fixed and known, and the second simultaneously estimates $V_{ref}$, an additional parameter as shown in Fig. 3. This setting is incorporated as an additional optimization variable in the parameter vector $\Theta$. This variable is essential for constructing the inequality constraints $A\Theta \leq b$ according to the IEEE 1547-2018 standard [8].

We train the GMM according to our formulation in Section V-A on the entirety of dataset group three using the popular scikit-learn [47] library. The yielded log-likelihood surface for these data, which is part of the objective function of (30), is shown in Fig. 8.

Empirical histograms of the disaggregation model performance in terms of the standard kVAR deviation of the estimated PV reactive power from the true PV reactive power (RMSE) are shown in Fig. 13. Each observation in these histograms corresponds to the execution of the disaggregation algorithm (33) over the course of an entire year. Note that the RMSE values for the 451 load datasets studied are shown in Table II along with CNLLSQ.

An example of the disaggregation-based reconstruction is shown from the curve fitting perspective, in Fig. 12, and the

Fig. 11. Distributions of control parameters $\theta_1, \theta_2, \theta_3, \theta_4$ for the 451 loads in dataset category 2, each with the corresponding curve $\phi_\theta$ shown in black. Note that the reactive power settings are given in per unit. The BTM PV of each load has a different apparent power rating, therefore the actual curve settings in kVAR will differ significantly among loads.

Fig. 12. Example control setting with a scatter plot of the historical net load reactive power vs. PCC voltage for the load. Orange: MLE control curve found via disaggregation. Black: True underlying control curve. Purple: 95% confidence interval bands for each parameter. Note that the regions that are never historically actuated remain at zero.
Fig. 13. Empirical histogram of the RMSE and normalized RMSE values for the 451 loads with varying BTM Volt-VAR control settings. Each value is calculated for the full year-long time horizon dataset.

TABLE II
SAMPLE MEAN RMSE (kVAR) OF DISAGREGATED PV REACTIVE POWER TIMESERIES (451 LOADS)

| Algorithm                        | Average RMSE |
|----------------------------------|--------------|
| CNLLSQ, $\delta = 80$           | 0.198 kVAR   |
| Disaggregation, $V_{ref}$ Known  | 0.161 kVAR   |
| Disaggregation, $V_{ref}$ Estimated | 0.173 kVAR    |

Fig. 14. Orange: Example reconstructed PV reactive power timeseries found via disaggregation. Black: Ground truth PV reactive power timeseries. Purple: 95% confidence interval bands.

disaggregation perspective, in Fig. 14. The results in Table II indicate that the disaggregation approach performs roughly equivalent to CNLLSQ with small differences in kVAR deviation, in addition to not requiring manual parameter tuning. All results are for 451 1 year, 15-minute granularity AMI datasets. The kVAR errors indicate that all of the proposed methods of this study can reconstruct the control settings within 0.1-0.2 kVAR. These values are within typical errors for AMI and are in line with recent results in the literature for active power disaggregation, such as [12], [13].

VII. DISCUSSION

A. Limitations and Future Work

The methodologies presented in this paper have both limitations and opportunities for future work. Notably, the disaggregation algorithm could use an analytical distribution for the likelihood function. This may create opportunities for theoretical guarantees and Bayesian treatments of the method. Further, intuition suggests that the methods described in this paper may be generalizable to other control curves $\phi$. If this is the case, $\phi_{\Theta}$ would not need to describe an inverter control parameter specifically.

Throughout this paper, we have assumed access to the daytime/nighttime change points for the timeseries data. It is likely that these labels can be readily approximated if the location of the BTMPV is roughly known. If the user of these algorithms has no information about the BTM PV location, the thresholds will need to be estimated through weather data or another data-driven method.

As is the case in the second dataset group, the full range of the curve is not actuated for the majority of loads. While the disaggregation MLE approach is able to predict the exposed regions, there are opportunities for future work to recover the unactuated parameters. The current algorithm is accurate only within the actuated states, as shown in Fig. 12.

Furthermore, we have developed the algorithms of this paper based on 15-minute granularity AMI data, which is the most common according to a very recent industry survey in [48]. However, some utilities may limit their AMI data collection to hourly resolutions due to bandwidth, storage, or other limitations. Therefore, there is an opportunity for future work to investigate the sensitivity of the proposed methods with respect to various measurement resolutions, collection methods, and other practical data issues.

B. Comparison to Other Disaggregation Algorithms

The work we have presented is, to the best of our knowledge, the first solution to estimate the control parameters of BTM PV systems and simultaneously disaggregate inverter reactive power from net load data. However, the effectiveness of our proposed methods can be reasonably verified by qualitatively comparing them to similar works. For instance, the least-squares methods can be compared to our past work [9] on estimating power factor control settings from net load data, where we achieved an average RMSE of 0.194 kVAR. Similarly, the disaggregation-based elements of the presented methods are comparable in performance to existing active power load disaggregation algorithms. For example, reference [11] achieved RMSEs of approximately 450 kW for a 7.5 MW PV site. In the residential setting, references [12], [13] achieved RMSEs for active power in the range of 0-2 kW and 0-10 kW respectively. In our case, we achieved an average RMSE of 0.175 kVAR for residential PV inverters with sizes ranging from 1.35 to 6.5 kVA. There is an opportunity for future work to develop more universal testing benchmarks and datasets for this new area of research to allow for new developments and extensions of this work and others.

VIII. CONCLUSION

We have presented a set of methods to reconstruct unknown reactive power control parameters of solar PV inverters from aggregate AMI measurements, without the need for a distribution circuit model. The first of these algorithms allows for efficient reconstruction via non-linear least-squares based on
parameterized control curves and requires manual tuning of a filter. This method was shown to reconstruct the parameters with MAPE values between 0.425-2.870%. To avoid tuning this filter for each individual load, we developed the more flexible net load reactive power disaggregation approach to solve the same problem with the same data inputs.

Interestingly, we demonstrated a link between the problems of reconstructing reactive power control settings and net load reactive power disaggregation. We showed that the two goals can be accomplished in a simultaneous and complementary manner with a mean prediction error of less than 0.2 kVAR. These results are comparable to the performance of existing active power disaggregation methods in the literature.

The tools developed in this paper are useful for modeling, predicting, and verifying the impact of advanced inverters on the distribution grid, without the input of the PV installer or the customer. These methods can also be used to update controller databases on BTM PV control settings regardless of whether they change or are modified in the future.

As advanced inverter and AMI deployment increases, these methods can help utilities and distribution engineers combat incomplete or out-of-date information in their feeder models, assisting the ongoing sustainable energy transition.

**APPENDIX**

**[NIGHTTIME INVERTER OPERATION]**

If reactive power control is set to continue during the night, the disaggregation problem (33) can be modified in the following manner. Let \( v_{n,m}^{(p)} \) and \( v_{d,m}^{(p)} \) denote the voltage timeseries vectors during the night and day respectively. Then let \( \hat{q}_{d,m}^{pu} := T \{ \phi_d (v_{n,m}^{(p)}) \} |_{m} \) and \( \hat{q}_{d,m}^{pv} := T \{ \phi_d (v_{d,m}^{(p)}) \} |_{m} \) be the estimated PV reactive power contributions at sample \( m \in \{1, \ldots, M'\} \). We can then modify (33) and solve

\[
\min_{\Theta} \left[ - \sum_{m=1}^{M'} \log \left( \sum_{k=1}^{K} \Omega_{m,k} \right) \right] \quad \text{subject to}
\]

\[
\Omega_{m,k} = \alpha_k q_{n,m}^{net} - \hat{q}_{d,m}^{pu} - \hat{q}_{d,m}^{pv} \gamma_k, \quad A\Theta \leq b.
\]

**ACKNOWLEDGMENT**

The authors would like to thank Logan Blakely for the helpful conversations and the reviewers for their constructive comments. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

**REFERENCES**

[1] K. Baker, A. Bernstein, E. Dall’Anese, and C. Zhao, “Network-cognizant voltage droop control for distribution grids,” *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 2098–2108, Mar. 2018.

[2] X. Zhou, M. Fairvar, Z. Liu, L. Chen, and S. H. Low, “Reverse and forward engineering of local voltage control in distribution networks,” *IEEE Trans. Autom. Control*, vol. 66, no. 3, pp. 1116–1128, Mar. 2021.

[3] S. S. Guggilam, E. Dall’Anese, Y. C. Chen, S. V. Dhople, and G. B. Gianakis, “Scalable optimization methods for distribution networks with high PV integration,” *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 2061–2070, Jul. 2016.

[4] M. U. Qureshi, S. Grijalva, and M. J. Reno, “A fast quasi-static time series simulation method for PV smart inverters with VAR control using linear sensitivity model,” in *Proc. IEEE 7th World Conf. Photovoltaic. Energy Convers.* (A Joint Conf. 45th IEEE PVSC, 28th PVPSEC 34th EU PVSEC), Waikoloa, HI, USA, IEEE, 2018, pp. 1614–1619.

[5] J. Seuss, M. J. Reno, R. J. Broderick, and S. Grijalva, “Analysis of PV advanced inverter functions and setpoints under time series simulation,” *Tech. Rep. SAND2016-48561259558*, May 2016.

[6] L. Blakely, M. J. Reno, and J. Peppanen, “Identifying common errors in distribution system models,” in *Proc. IEEE 46th Photovolt. Specialists Conf.*, May 2019, pp. 3132–3139.

[7] Y. Wang, Q. Chen, T. Hong, and C. Kang, “Review of smart meter data analytics: Applications, methodologies, and challenges,” *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3125–3148, May 2019.

[8] IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces, IEEE Std 1547-2018 (Revision of IEEE Std 1547-2003), pp. 1–138, 2018.

[9] S. Talkington, S. Grijalva, and M. J. Reno, “Power factor estimation of distributed energy resources using voltage magnitude measurements,” *J. Modern Power Syst. Clean Energy*, vol. 9, no. 4, pp. 859–869, Jul. 2021.

[10] P. Li, B. Zhang, Y. Weng, and R. Rajagopal, “A sparse linear model and significance test for individual consumption prediction,” *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4489–4500, Nov. 2017.

[11] E. C. Kara, C. M. Roberts, M. Tabone, L. Alvarez, D. S. Callaway, and E. M. Stewart, “Disaggregating solar generation from feeder-level measurements,” *Sustain. Energy, Grids Netw.*, vol. 13, pp. 112–121, Mar. 2018.

[12] F. Bu, K. Dehghanpour, Y. Yuan, Z. Wang, and Y. Guo, “Disaggregating customer-level behind-the-meter PV generation using smart meter data and solar exemplars,” *IEEE Trans. Power Syst.*, vol. 36, no. 6, pp. 5417–5427, Nov. 2021.

[13] F. Bu, K. Dehghanpour, Y. Yuan, Z. Wang, and Y. Zhang, “A data-driven game-theoretic approach for behind-the-meter PV generation disaggregation,” *IEEE Trans. Power Syst.*, vol. 35, no. 4, pp. 3134–3144, Jul. 2020.

[14] Y. Zhou, Z. Shi, Z. Shi, Q. Gao, and L. Wu, “Disaggregating solar power consumption of commercial buildings based on the finite mixture model,” *Appl. Energy*, vol. 243, pp. 35–46, Jun. 2019.

[15] M. Tabone, S. Kiliccote, and E. C. Kara, “Disaggregating solar generation behind individual meters in real time,” in *Proc. 5th Conf. Syst. Built Environ.*, Shenzhen China, 2018, pp. 43–52.

[16] A. Cooper and M. Shuster, Electric Company Smart Meter Deployments: Foundation for a Smart Grid. Washington, DC, USA: The Edison Found. Inst. Electr. Innov., Apr. 2021.

[17] X. Chen, W. Wu, B. Zhang, and C. Lin, “Data-driven DG capacity assessment method for active distribution networks,” *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3946–3957, Sep. 2017.

[18] S. Karagiannopoulos, R. Dobbie, P. Aristidou, D. Callaway, and G. Hug, “Data-driven control design schemes in active distribution grids: Capabilities and challenges,” in *Proc. IEEE Milan PowerTech, 2019*, pp. 1–6.

[19] R. O’Connell, C. Volkmann, and P. Brucke, “Regulating voltage: Recommendations for smart inverters,” GridLab, Berkeley, CA, USA, Tech. Rep., Sep. 2019. [Online]. Available: https://gridlab.org/publications

[20] F. Bell, A. Nguyen, M. McCarty, K. Atef, and T. Bialek, “Secondary volt- age and reactive power support via smart inverters on a high-penetration distributed photovoltaic circuit,” in *Proc. IEEE Power Energy Soc. Inov. Smart Grid Technol. Conf.*, 2016, pp. 1–6.

[21] A. O’Connell and A. Keane, “Volt-VAR curves for photovoltaic inverters in distribution systems,” *Transmiss. Distrib. IET Gener.*, vol. 11, no. 3, pp. 730–739, 2017.

[22] F. Sultanem, “Using appliance signatures for monitoring residential loads at meter panel level,” *IEEE Trans. Power Del.*, vol. 6, no. 4, pp. 1380–1385, Oct. 1991.

[23] J. Z. Kolter, S. Batra, and A. Y. Ng, “Energy disaggregation via discriminative sparse coding,” in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2010, vol. 1, pp. 1153–1161.

[24] B. Wang, F. Chen, H. Dong, A. P. Boedihardjo, and C.-T. Lu, “Signal disaggregation via sparse coding with featured discriminative dictionary,” in *Proc. IEEE 12th Int. Conf. Data Mining*, 2012, pp. 1134–1139.

[25] A. Harel, S. Makonin, and I. V. Bajić, “Wavenilm: A causal neural network for power disaggregation from the complex power signal,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 8335–8339.

[26] D. Chen and D. Irwin, “SunDance: Black-box behind-the-meter solar exemplars,” in *Proc. 8th Conf. Syst. Built Environ.*, 2016, pp. 1–6.

[27] G. S. Ledva, L. Balzano, and J. L. Mathieu, “Real-time energy disaggregation of a distribution Feeder’s demand using online learning,” *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 4730–4740, Sep. 2018.
[28] K. Mason, M. J. Reno, L. Blakely, S. Vejdan, and S. Grijalva, “A deep neural network approach for behind-the-meter residential PV size, tilt and azimuth estimation,” Sol. Energy, vol. 196, pp. 260–269, Jan. 2020.

[29] X. Zhang and S. Grijalva, “A data-driven approach for detection and estimation of residential PV installations,” IEEE Trans. Smart Grid, vol. 7, no. 5, pp. 2477–2485, Sep. 2016.

[30] S. Lin and H. Zhu, “Enhancing the spatio-temporal observability of grid-edge resources in distribution grids,” IEEE Trans. Smart Grid, vol. 12, no. 6, pp. 5434–5443, Nov. 2021.

[31] Y. Wang, N. Zhang, Q. Chen, D. S. Kirschen, P. Li, and Q. Xia, “Data-driven probabilistic net load forecasting with high penetration of behind-the-meter PV,” IEEE Trans. Power Syst., vol. 33, no. 3, pp. 3255–3264, May 2018.

[32] C. M. Cheung, W. Zhong, C. Xiong, A. Srivastava, R. Kannan, and V. K. Prasanna, “Behind-the-meter solar generation disaggregation using consumer mixture models,” in Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids, 2018, pp. 1–6.

[33] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, “Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities,” IEEE Trans. Smart Grid, vol. 6, no. 2, pp. 911–918, Mar. 2015.

[34] N. Zhou, J. Wang, and Q. Wang, “A novel estimation method of metering errors of electric energy based on membership cloud and dynamic time warping,” IEEE Trans. Smart Grid, vol. 8, no. 3, pp. 1318–1329, May 2017.

[35] S. Diamond and S. Boyd, “CVXPY: A Python-embedded modeling language for convex optimization,” J. Mach. Learn. Res., vol. 17, no. 83, pp. 1–5, 2016.

[36] P. Virtanen et al., “SciPy 1.0: Fundamental algorithms for scientific computing in Python,” Nature Methods, vol. 17, pp. 261–272, 2020.

[37] S. Talkington, S. Grijalva, M. J. Reno, and J. Azzolini, “Recovering power factor control settings of solar PV inverters from net load data,” in Proc. 53rd North Amer. Power Symp., College Station, TX, 2021, pp. 1–6.

[38] S. Talkington, S. Grijalva, and M. J. Reno, “Sparse time series sampling for recovery of behind-the-meter inverter control models,” in Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf., 2022.

[39] H. Ashtiani, S. Ben-David, N. Harvey, C. Liaw, A. Mehrabian, and Y. Plan, “Nearly tight sample complexity bounds for learning mixtures of Gaussians via sample compression schemes,” in Proc. 32nd Conf. Neural Inf. Process. Syst., 2018, pp. 3416–3425.

[40] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” J. Roy. Stat. Soc.: Ser. B (Methodological), vol. 39, no. 1, pp. 1–22, Sep. 1977.

[41] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York, NY, USA: Springer, Jan. 2017.

[42] C. M. Bishop, Pattern Recognition and Machine Learning. Berlin, Heidelberg: Springer, 2006.

[43] G. Schwarz, “Estimating the dimension of a model,” Ann. Statist., vol. 6, no. 2, pp. 461–464, 1978.

[44] B. G. Leroux, “Consistent estimation of a mixing distribution,” Ann. Statist., vol. 20, no. 3, pp. 1350–1360, 1992.

[45] D. Kraft, “A software package for sequential quadratic programming,” Tech. Rep. DPVLR-FB 89-28, DLR German Aerospace Center - Institute for Flight Mechanics, Köln, Germany, 1988.

[46] R. H. Byrd, M. E. Hribar, and J. Nocedal, “An interior point algorithm for large-scale nonlinear programming,” SIAM J. Optim., vol. 9, no. 4, pp. 877–900, Jan. 1999.

[47] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.

[48] J. Peppanen, M. Hernandez, J. Deboever, and M. Rylander, “Distribution load modeling - survey of the industry state, current practices and future needs,” in Proc. 53rd North Amer. Power Symp., College Station, TX, 2021, pp. 1–6.

Samuel Talkington (Student Member, IEEE) received the B.S. degree in electrical engineering from West Virginia University, Morgantown, WV, USA. He is currently working toward the Ph.D. degree with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA. From 2019 to 2020, he completed internships with the Power System Engineering and Electrical Distribution Divisions, Mitsubishi Electric Power Products.

His research interests include estimation, control, and optimization in electric power systems.

Santiago Grijalva (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA. He is the Southern Company Distinguished Professor of Electrical and Computer Engineering and Director of the Advanced Computational Electricity Systems Laboratory at the Georgia Institute of Technology, Atlanta, GA, USA. From 2002 to 2009, he was with PowerWorld Corporation. From 2013 to 2014, he was the Founding Director at the Power System Engineering Center, National Renewable Energy Laboratory. Dr. Grijalva was a Member of the NIST Federal Smart Grid Advisory Committee. His research interests include decentralized power system control, cyber-physical security, and economics.

Matthew J. Reno (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA. He is currently a Principal Member of the Technical Staff with the Department of Electric Power Systems Research, Sandia National Laboratories. His research interests include distribution system modeling and analysis with big data and high penetrations of PV.

Matthew leads several projects that are applying cutting edge machine learning algorithms to power system problems.

Joseph A. Azzolini (Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from Arizona State University, Tempe, AZ, USA. He is currently a Senior Member of Technical Staff with the Department of Electric Power Systems Research, Sandia National Laboratories. His research interests include distribution system modeling and analysis, particularly applied to study the integration of smart grid devices, advanced inverter controls, and high PV penetrations.