Optimal Matching of Stochastic Solar Generators to Stochastic Loads

Sun Sun
University of Waterloo
200 University Avenue West
Waterloo, Ontario, Canada
sun.sun@uwaterloo.ca

Srinivasan Keshav
University of Waterloo
200 University Avenue West
Waterloo, Ontario, Canada
keshav@uwaterloo.ca

Catherine Rosenberg
University of Waterloo
200 University Avenue West
Waterloo, Ontario, Canada
cath@uwaterloo.ca

Matthew Peloso
Sun Electric Pte Ltd
3 Church Street
Singapore
mpeloso@sunelectric.com.sg

ABSTRACT

To meet the demand for locally-produced and sustainable power, community microgrids distribute power generated by roof-mounted solar PV systems to 'green' consumers. In this context, we consider the problem of matching one or more inherently intermittent solar energy producers with each green consumer so that, with a high probability, a certain component of their load is met from solar generation. We formulate this optimal matching as a stochastic optimization problem which incorporates the uncertainty of both solar and loads. To solve the problem, we propose four algorithms which make different assumptions on the distributions of solar generation and loads. We compare the performance of these algorithms using real data, and find that, for our dataset, an approach that assumes Gaussian mixture models for solar and loads best fits our design requirements. We also investigate admission control algorithms to admit customers based on our matching algorithms so that the solar allocation is feasible.

KEYWORDS

Virtual power plant; matching; solar; loads; stochastic optimization

1 INTRODUCTION

With the rapid decline in the price of solar photovoltaic systems, it has become increasingly cost-effective for both residential and commercial building owners to generate electricity from roof-mounted systems. They can either use this energy themselves or sell it to geographically-close 'green' consumers who wish to purchase renewable energy for many reasons including: reducing their carbon footprint; reducing their electricity cost, if the sale price is lower than the cost of grid electricity; reducing their price volatility due to exposure to fossil-powered grid electricity; and, in some jurisdictions, obtaining tax or 'green' building credits [9, 27]. However, most building owners do not have the skill set necessary to finance, install, operate, and manage such systems. To alleviate this problem, third party virtual power plant (VPP) operators have taken on these tasks, aggregating and operating distributed generating facilities [19, 20, 25, 33].

Most existing VPPs act as both purchasers and suppliers of green electricity, executing supply-side contracts with generators to create renewable energy certificates (RECs), and demand-side contracts with consumers to sell them RECs to match their demand, thus supplying them 'green' energy. A recent trend in such systems is to allow generators and consumers to directly enter into peer-to-peer contracts [19, 23, 25, 26]. In such a system, the VPP acts essentially as a market maker (akin to eBay or Kijiji), and supply contracts for the next billing period as peer-to-peer, i.e., between generators and consumers. Nevertheless, it is still necessary for the VPP operator to control the admission of generators and consumers into the system, and to match generators to consumers, so that, with a high probability, the supply contracts will be met.

This motivates the following matching problem: given historical data of solar generation and energy consumption of consumers, match a certain component of generation from each producer with each consumer so that, with a high probability, their anticipated load in the next billing cycle[1] is met (see Section 3 for details). This matching problem is solved at the start of each billing cycle, which is when load and solar profiles can be updated with recent measurement data, and new consumers or new generators are admitted to the system (if possible). The main challenge is that the actual future values of both solar generation and energy consumption are highly variable and unknown.

More specifically, in this paper, we consider two problems. First, we propose and compare several approaches to allocating a fixed portion of energy generated by each solar producer to each consumer for the next billing cycle, e.g., a month. Second, we present admission control algorithms that determine whether a consumer can be admitted to the system, i.e., their anticipated load can be met from the existing set of producers. Solutions to both problems use a probabilistic stochastic optimization approach to model intermittency in solar power generation and load variability.

Our key contributions are as follows.

- We formulate the matching of solar producers to loads as a stochastic optimization problem in which the uncertainty of both solar and loads are considered.

[1] Typically one month.

- To solve the stochastic optimization problem, we propose three approaches (and four algorithms) based on different assumptions on the distributions of solar and loads. We compare the effectiveness of these algorithms in a realistic setting.
- We develop admission control algorithms to guarantee the feasibility of the matching.

The remainder of the paper is organized as follows. Section 2 presents related work. In Section 3, we present our system model. Section 4 considers several matching approaches based on chance-constrained stochastic programming. We present algorithms for admission control in Section 5. Numerical results are presented in Section 6 and we discuss and conclude our work in Section 7.

2 RELATED WORK

This section reviews related work, providing a broad overview of virtual power plants, stochastic optimization, solar and load modeling, and a detailed summary of other work that has used a stochastic modeling approach in the context of renewable energy systems.

2.1 Virtual Power Plants

A virtual power plant (VPP) aggregates generators to either supply consumers or to sell their power in a market [30]. They arise naturally in the context of distributed generation, where each generator individually has neither the management capability nor the market power to interact with existing large-scale grid operators [31]. As distributed solar generation becomes more popular, numerous VPPs have entered markets around the world [30]. Our work focuses on VPPs that operate in microgrids, allowing peer-to-peer energy trading [19, 23, 25, 26].

2.2 Stochastic Optimization

Stochastic optimization problems can be solved in several ways, including:

(1) *Scenario-based approaches* that make no assumption on uncertainty distribution. However, they generally require a large number of scenarios (i.e., samples of historical data) for good performance, inevitably leading to high computational complexity [10, 24].

(2) *Probabilistic approaches* where chance constraints incorporate uncertainty [7, 12]. To handle such constraints, we can either assume uncertain parameters come from specific distributions, or adopt a distributionally-robust approach [11, 13] that relies only on statistics of a distribution (e.g., mean and variance), hence is applicable to random variables drawn from a family of distributions.

(3) *Worst-case robust approaches* which provide strategies that are guaranteed to meet constraints even in the worst-case scenario [5, 6]. However, this generally results in an overly-conservative allocation of resources.

Since the matching between solar producers and consumers can be naturally formulated as chance constraints, we adopt the second approach in our work. We first investigate methods that assume specific distribution information (i.e., Gaussian in Section 4.1 and Gaussian Mixture Models (GMMs) in Section 4.2) and then study a distributionally-robust strategy in Section 4.3.

2.3 Solar Radiation and Load Modeling

There are numerous mathematical models for solar radiation incident on the Earth's surface [2, 35]. Broadly speaking, these fall into two categories: physical and statistical. Physical models represent site attributes such as atmospheric turbidity, shadowing, and the level of diffuse radiation. In contrast, statistical models are entirely data-driven. Accurate physical models use sophisticated forms and are highly parameterized, thus are not amenable for use in optimization. Hence, we focus on statistical models.

Statistical models are used to either forecast solar radiation for the next time period, given past history, or to model the variability in solar irradiation for a specific hour of the day in a particular season [2]. Since we wish to model solar radiation for a specific time period as an i.i.d random variable, we focus on the latter approach.

Specifically, we use Gaussian Mixture Models (GMMs) and their simplest variant, a single-Gaussian model, in our work. In contrast to single models, in mixture models, each component of the mixture corresponds to a certain sky condition (such as cloudy, partly cloudy, or clear). Such use of GMMs to model solar radiation for a specific time period is well-known [1, 2, 21].

Load modeling has also been extensively studied in prior work (e.g., [3, 4, 18]). As with solar modeling, we seek to statistically model load during a specific time period. Based on prior work, we use a GMM (or a Gaussian) for load modeling [29, 32]. We also investigate a distributionally-robust approach that considers a family of distributions.

2.4 Stochastic Optimization in the Context of Renewable Energy

Several researchers have used stochastic optimization to solve problems that arise in the context of renewable energy. For example, [15] aims to reduce the variance of aggregate load (load minus renewable generation) by managing the power consumption of deferrable loads under the consideration of prediction uncertainty. To balance supply and demand, [34] proposes a real-time algorithm where the uncertainty of renewable and loads is modeled by a range.

Instead of general renewable energy, prior work has also specifically investigated systems with solar (e. g., [28, 36]) or wind generation (e.g., [17, 22]). In particular, [36] studies joint-operation in building microgrids. A finite-stage event-based optimization model is used to formulate the problem with the decision made based on the observed event. Similarly, [28] adopts a scenario-based approach to minimize electricity cost while maintaining the comfort requirements of each user.

Both [17, 22] assume specific distributions in the problem formulation involving wind generation. For example, [17] investigates controlling the charging load of electric vehicles to wind power by formulating the problem as a Markov decision process. To model system uncertainty, the Weibull distribution is used for wind speed, a truncated Gaussian distribution is assumed for the parking duration of electric vehicles, and the χ^2-distribution is considered for the driving distance. An improved charging policy is provided based on simulation.

In contrast to prior work, we study stochastic optimization based both on specific distributions as well as distributionally-robust approaches. Moreover, the problem context, i.e., optimal matching in a microgrid, to our knowledge, has not been studied in the literature.

3 SYSTEM MODEL

Consider a time slotted system with each day indexed by d, and within each day there are several disjoint time slots indexed by k (e.g., 10:30am-11am and 12:00pm-12:30pm). Denote the set of the days considered as $\mathcal{D} \triangleq \{1, \cdots, D\}$ and the set of the time slots within each day as $\mathcal{K} \triangleq \{1, \cdots, K\}$. We are given a set of $\mathcal{P}$ solar producers numbering $|\mathcal{P}|$. Let $p_i(d,k)$ be the amount of solar energy that producer i generates during time slot k on day d. If a producer has more than one solar panel, $p_i(d,k)$ represents the aggregate amount of energy it generates. Symmetrically, we are given a set of C consumers numbering $|C|$. Let $c_j(d,k)$ be the amount of energy used by consumer j during time slot k on day d. Note that due to the uncertainty of solar generation and load, information about $p_i(d,k)$ and $c_j(d,k)$ is not known in advance.

In Fig.1, we show a bipartite graph where the solar producers are in one set and the consumers are in the other set. The direction of each edge indicates the direction of energy flow. A consumer can be served by multiple generators and a generator can serve multiple customers, i.e., a generator can split its generation without any constraint.

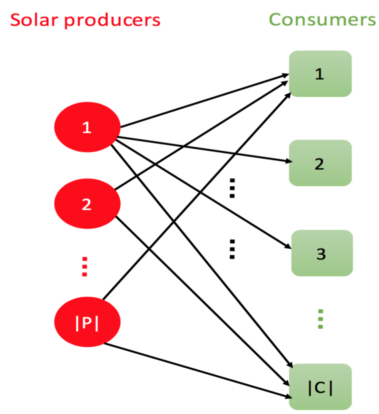


Figure 1: System model.

At the beginning of each billing cycle, the VPP recruits new generators and, if resources permit, new consumers. Moreover, each consumer negotiates contracts with a set of generators based on the result of the matching algorithm. Based on the services currently provided by SunElectric, a typical solar VPP, we model two types of consumers [20].

- Priority I: Each consumer is ensured that a predetermined fraction of their next month's load is met for a certain part of the day. For example, with the Solar100 service from SunElectric, 100% of the energy consumption during the chosen time slots is satisfied by solar generation with a probability higher than some threshold (e.g., 0.9). Once admitted into the system, the consumer is matched to one or more solar producers that will meet its demand for this month. The matching algorithm assigns producers and a corresponding percentage of their solar generation to each consumer.
- Priority II : These consumers are not guaranteed green energy for the next billing cycle, but are retroactively assigned any unallocated renewable energy production from the prior billing cycle. In other words, for such consumers, the VPP operator guarantees that a certain fraction of their actual energy consumption during the whole month will be satisfied by solar generation. This models the SolarLite service from SunElectric.

The objective of this paper is as follows: first, for Priority I consumers, design admission control strategies and also obtain the matching parameters for admitted consumers; second, for Priority II consumers, design admission control strategies[2].

4 MATCHING DESIGN FOR PRIORITY I CUSTOMERS

We consider the design of a matching algorithm that takes into account the fact that values of the solar generation and loads are uncertain in that they are unavailable until they are realized. Thus, the system operator has to rely on historical data when designing the matching parameters, knowing that this may not exactly reflect the future generation or the future load. In particular, we assume that the following information is available:

- for each solar producer, we have historical half-hourly solar generation of previous years/months; and
- for each existing/new customer, we have historical half-hourly energy consumption of previous years/months.

The matching algorithm assumes that there are enough solar resources in the system (i.e., the optimization problem is always feasible): Section 5 discusses how to admit customers so that this assumption is met. The goal of the matching algorithm is to construct a matching matrix $\mathbf{M} \in \mathbb{R}^{|\mathcal{P}|\times|C|}$ with its (i,j)-th element $m_{i,j}$ denoting the percentage of producer i's generation assigned to consumer j for the whole month.

Then, we can naturally formulate the service requirement of consumer j as a chance constraint: $\Pr\left(c_j(d,k) \le \sum_{i\in\mathcal{P}} m_{i,j} p_i(d,k)\right) \ge \alpha$. That is, the load of consumer j during time slot k on day d should be met by solar generation with a probability at least α, $\alpha \in (0.5, 1)$, e.g., 0.9. To determine the matching parameters, we formulate the following stochastic optimization problem, given the distributions of $p_i(d,k)$ and $c_j(d,k)$ and the target probability α:

$$\textbf{SP:} \min_{\mathbf{M}} \quad \mathbb{E}\left[\sum_{d\in\mathcal{D},k\in\mathcal{K}} \sum_{i\in\mathcal{P},j\in C} m_{i,j} p_i(d,k)\right]$$

$$\text{s.t. } m_{i,j} \ge 0, \quad \forall i \in \mathcal{P}, \forall j \in C, \tag{1}$$

$$\sum_j m_{i,j} \le 1, \quad \forall i \in \mathcal{P}, \tag{2}$$

$$\Pr\left(c_j(d,k) \le \sum_{i\in\mathcal{P}} m_{i,j} p_i(d,k)\right) \ge \alpha, \quad \forall d \in \mathcal{D}, \forall k \in \mathcal{K}, \forall j \in C, \tag{3}$$

[2]No matching is necessary for Priority II consumers since the allocation is after the fact, and for aggregate consumption.

where the objective is to minimize the expected solar production needed to satisfy the consumer demand, and the expectation is taken over the solar generation of all time slots. This objective allows us to subsequently allocate the most possible solar generation to Priority II customers (see Section 5.2 for details). Note that (3) is a stochastic constraint, and cannot be solved using a standard solver. Our work, therefore, lies in translating this constraint into a form that can be solved using standard solvers.

Constraint (1) requires that all matching parameters should be greater than zero. Constraint (2) ensures that the allocated solar generation of producer i cannot exceed its total resource. Note that the chance constraint (3) is more binding for higher values of α.

Next, we propose three approaches to solve **SP**, making different assumptions to model the uncertainty of future generation and load, thus resulting in different algorithms where Constraint (3) are replaced with deterministic ones.

4.1 Stochastic Optimization Using a Gaussian Approximation

To transform the chance constraint to a deterministic form we need the distributions of the uncertain parameters. The Gaussian distribution is widely used in many applications and generally results in a simple analysis. In this section, we use the Gaussian distribution to model the uncertainty of the solar and energy consumption, and propose an algorithm under this approximation. In particular, we make the following assumptions:

(1) the distribution of the solar generation and energy consumption is Gaussian;
(2) the solar generation of all producers are linearly correlated;
(3) the solar generation and energy consumption are independently distributed; and
(4) the distribution of solar and energy consumption is independent and identically distributed (i.i.d.) over days.

Assumption (2) is not essential, but will simplify the analysis, since we can focus on the profile of one solar producer instead of $|\mathcal{P}|$ producers. We expect this assumption to hold for geographically-close solar producers, and we further validate it in Section 6 using real data. Assumptions (3) and (4) are technical assumptions that are required for either the problem formulation or later the derivation of empirical statistics. Note that, since our analysis relies on historical data, we implicitly assume that the future can be characterized by historical data.

For simplicity of notation, we assume there is only one time slot every day (e.g., 12pm-12:30pm) for which there is a commitment from the generators to the consumers and omit the index k below[3]. From Assumption (4), it suffices to only consider the chance constraint of each consumer on a typical day, thus we omit the index d also. Based on Assumption (2), we define the solar generation of producer i as $\beta_i p$. Solar producer 1 is treated as the reference with $\beta_1 = 1$, and its solar generation on day d is denoted by $p(d)$ which is assumed to follow the Gaussian distribution with the mean $\overline{p}$ and variance σ_p^2. The value of other β_i can be evaluated by comparing the solar generation of producer i and producer 1 based on historical data. The energy consumption of consumer j is also assumed to follow the Gaussian distribution with mean $\overline{c}_j$ and variance σ_j^2.

With the Gaussian approximation and independence assumption, the chance constraint associated with consumer j, i.e., $\Pr\left(c_j \le \sum_{i\in\mathcal{P}} m_{i,j}\beta_i p\right) \ge \alpha$ can be equivalently written as

$$\frac{-\overline{c}_j + \overline{p}(\sum_{i\in\mathcal{P}} m_{i,j}\beta_i)}{\sqrt{\sigma_j^2 + \sigma_p^2(\sum_{i\in\mathcal{P}} m_{i,j}\beta_i)^2}} \ge \Phi^{-1}(\alpha), \tag{4}$$

where $\Phi^{-1}(\alpha)$ is the α-quantile of the standard Gaussian distribution. Since the threshold $\alpha > 0.5$, we have $\Phi^{-1}(\alpha) > 0$. We next show that (4) is a second-order cone constraint.

To facilitate the transformation to the second-order cone constraint, we introduce new optimization variables in the vector form. First we define the optimization variable $\mathbf{x}_j \in \mathbb{R}^{|\mathcal{P}|+1}$ associated with consumers j as

$$\mathbf{x}_j \triangleq [1, -m_{1,j}, -m_{2,j}, \cdots, -m_{|\mathcal{P}|,j}]^T, \tag{5}$$

where T means transpose. Then we define the optimization variable $\mathbf{y} \in \mathbb{R}^{|C|(|\mathcal{P}|+1)}$ as

$$\mathbf{y} \triangleq [\mathbf{x}_1^T, \mathbf{x}_2^T, \cdots, \mathbf{x}_{|C|}^T]^T. \tag{6}$$

Note that we define the optimization variable $\mathbf{y}$ in such a way to make the optimization variable consistent in all our proposed algorithms, which will be clear later.

Based on the definition of $\mathbf{x}_j$ and $\mathbf{y}$, we have $\mathbf{x}_j = \mathbf{E}_j\mathbf{y}$, where $\mathbf{E}_j \in \mathbb{R}^{(|\mathcal{P}|+1)\times|C|(|\mathcal{P}|+1)}$ is a constant matrix with all elements equal to zero except the j-th $(|\mathcal{P}|+1)$ columns which is the identity matrix. Define the constant vector $\mathbf{c}_i \in \mathbb{R}^{|\mathcal{P}|+1}, \forall i$, as $\mathbf{c}_i \triangleq [0, \cdots, -1, \cdots, 0]^T$ with all elements zero except the $(i+1)$-th element which is equal to -1. Then we can represent the matching parameter $m_{i,j}$ as $m_{i,j} = \mathbf{c}_i^T\mathbf{x}_j = \mathbf{c}_i^T\mathbf{E}_j\mathbf{y}$. To transform the chance constraint, we also define the constants $\mathbf{b}_j \in \mathbb{R}^{|\mathcal{P}|+1}, \forall j$, as

$$\mathbf{b}_j \triangleq [\overline{c}_j, \overline{p}\beta_1, \overline{p}\beta_2, \cdots, \overline{p}\beta_{|\mathcal{P}|}]^T$$

and $\mathbf{A}_j \in \mathbb{R}^{2\times(|\mathcal{P}|+1)}, \forall j$, as

$$\mathbf{A}_j \triangleq \begin{bmatrix} \sigma_j & 0 & 0 & \cdots & 0 \\ 0 & \sigma_p\beta_1 & \sigma_p\beta_2 & \cdots & \sigma_p\beta_{|\mathcal{P}|} \end{bmatrix}$$

Then constraint (4) is equivalent to $\|\mathbf{A}_j\mathbf{E}_j\mathbf{y}\| \le \frac{-1}{\Phi^{-1}(\alpha)}\mathbf{b}_j^T\mathbf{E}_j\mathbf{y}$.

Finally, under the Gaussian approximation we can rewrite the stochastic optimization problem **SP** in the following form, given the mean and variance of the Gaussian distribution, the linear coefficients β_i, and the target probability α:

$$\textbf{GauSP:} \quad \min_{\mathbf{y}} \quad \sum_{i\in\mathcal{P}}\sum_{j\in C} |\mathcal{D}|\overline{p}\beta_i\mathbf{c}_i^T\mathbf{E}_j\mathbf{y}$$

$$\text{s.t.} \quad \mathbf{c}_i^T\mathbf{E}_j\mathbf{y} \ge 0, \forall i \in \mathcal{P}, \forall j \in C, \tag{7}$$

$$\sum_{j\in C}\mathbf{c}_i^T\mathbf{E}_j\mathbf{y} \le 1, \forall i \in \mathcal{P}, \tag{8}$$

$$\|\mathbf{A}_j\mathbf{E}_j\mathbf{y}\| \le \frac{-1}{\Phi^{-1}(\alpha)}\mathbf{b}_j^T\mathbf{E}_j\mathbf{y}, \forall j \in C. \tag{9}$$

We will discuss how to derive the estimates of statistics of the Gaussian distribution in Section 4.4.

[3]It is straightforward to generalize our approach to multiple time slots by requiring that the constraint α be met for each of the selected time slots.

We can see that **GauSP** is a second-order cone program, which can be efficiently solved by many optimization software packages such as CVX [16]. Thus, for the matching design we first solve **GauSP**. Once the optimization variable y is obtained, it is straightforward to obtain the matching parameters $m_{i,j}$.

4.2 Stochastic Optimization Using Gaussian Mixture Models

In Section 4.1 we use the Gaussian distribution to approximate the real distributions of solar and energy consumption. In this section, we provide an algorithm obtained by approximating the distribution of solar and energy consumption as a GMM. For the derivation of the algorithm, we make the same assumptions as in Section 4.1 except that in Assumption (1), we replace the Gaussian distribution with a GMM. Indeed, based on measured generation data provided by a solar energy company[4], in Fig. 2, we show the normalized empirical distributions of solar generation during the same half-hour of the day in three representative months. We find that the empirical distribution has multiple peaks, consistent with a mixture model such as a Gaussian mixture model (GMM).

Under the approximation of a GMM, we denote the distribution of the reference solar generation as

$$p \sim \sum_{r}^{R} \pi_r \mathcal{N}(\bar{p}_r, \sigma_r^2), \tag{10}$$

where R is the number of Gaussian components, π_r is the probability associated with the r-th Gaussian component, $\bar{p}_r$ is the mean and σ_r^2 is the variance of the r-th Gaussian component. Similarly, we can denote the distribution of the energy consumption of consumer j as

$$c_j \sim \sum_{l}^{L_j} \pi_{j,l} \mathcal{N}(\bar{c}_{j,l}, \sigma_{j,l}^2). \tag{11}$$

where $\pi_{j,l}$ is the probability associated with the l-th Gaussian component, and L_j is the number of Gaussian components.

The following proposition holds.

Proposition 1. *Assume that the distribution of the solar generation and the energy consumption of each consumer follows a GMM, represented by* (10) *and* (11)*, respectively. Also assume that the solar generation and energy consumption are independent. Then c_j and p are jointly Gaussian mixture distributed. In particular, we have,*

$$\boldsymbol{\omega}_j \triangleq \begin{bmatrix} c_j \\ p \end{bmatrix} \sim \sum_{l,r} \pi_{j,l}\pi_r \mathcal{N}\left(\begin{bmatrix} \bar{c}_{j,l} \\ \bar{p}_r \end{bmatrix}, \begin{bmatrix} \sigma_{j,l}^2 & 0 \\ 0 & \sigma_r^2 \end{bmatrix}\right). \tag{12}$$

For simplicity of notation, we denote $\boldsymbol{\omega}_j \sim \sum_{h=1}^{H_j} \pi_{j,h}\mathcal{N}(\boldsymbol{\mu}_{j,h}, \Sigma_{j,h})$, where $H_j \triangleq RL_j$, $\pi_{j,h} \triangleq \pi_{j,l}\pi_r$, $\boldsymbol{\mu}_{j,h} \triangleq \begin{bmatrix} \bar{c}_{j,l} \\ \bar{p}_r \end{bmatrix}$, and $\Sigma_{j,h} \triangleq \begin{bmatrix} \sigma_{j,l}^2 & 0 \\ 0 & \sigma_r^2 \end{bmatrix}$.

Proof. The proof can be seen in Prop. 4 in [14]. □

With the definition of $\boldsymbol{\omega}_j$ in (12), we can rewrite the chance constraint associated with consumer j into a compact form as $\Pr(\boldsymbol{\omega}_j^T \mathbf{z}_j \le 0) \ge \alpha$, where $\mathbf{z}_j$ contains the optimization variables and is defined as $\mathbf{z}_j \triangleq \begin{bmatrix} 1 \\ -\sum_{i\in\mathcal{P}} m_{i,j}\beta_i \end{bmatrix}, \forall j$. Note that the introduction of $\boldsymbol{\omega}_j$ is crucial, which will be used in the proof of Prop. 2.

We now discuss how to handle the chance constraint under the GMM approximation. Denote the original matching optimization problem with the constraints (7), (8) and the chance constraints $\Pr(\boldsymbol{\omega}_j^T \mathbf{z}_j \le 0) \ge \alpha, \forall j$, as **GmmSP-P1**, and denote an optimal solution of **GmmSP-P1** as $\mathbf{y}^*$. Consider a new optimization problem **GmmSP-P2**, which is the same as **GmmSP-P1** except that the chance constraints are replaced with two deterministic constraints. Given the statistics of GMMs in (10) and (11), the linear coefficients β_i, and the target probability α, we have the following formulation:

$$\textbf{GmmSP-P2:} \min_{\mathbf{y}, \epsilon_{j,h}} \left(|\mathcal{D}| \sum_r \pi_r \bar{p}_r\right) \sum_{i\in\mathcal{P}} \sum_{j\in C} \beta_i \mathbf{c}_i^T \mathbf{E}_j \mathbf{y}$$

$$\text{s.t.} \quad \mathbf{c}_i^T \mathbf{E}_j \mathbf{y} \ge 0, \forall i \in \mathcal{P}, \forall j \in C,$$

$$\sum_{j\in C} \mathbf{c}_i^T \mathbf{E}_j \mathbf{y} \le 1, \forall i \in \mathcal{P},$$

$$\sum_{h=1}^{H_j} \pi_{j,h}\epsilon_{j,h} \ge \alpha, \forall j, \tag{13}$$

$$\epsilon_{j,h} \le \Phi\left(\frac{-\boldsymbol{\mu}_{j,h}^T \mathbf{B}\mathbf{E}_j \mathbf{y}}{\|\Sigma_{j,h}^{\frac{1}{2}} \mathbf{B}\mathbf{E}_j \mathbf{y}\|}\right), \forall j, \forall h, \tag{14}$$

where $\epsilon_{j,h}$ are new optimization variables, $\Phi(\cdot)$ is the CDF of the standard Gaussian distribution, and (13) and (14) are the new introduced constraints. We will discuss how to derive the estimates of statistics of GMMs in Section 4.4. Denote an optimal solution of **GmmSP-P2** as $(\tilde{\mathbf{y}}, \tilde{\epsilon}_{j,h})$. Below, we show the equivalence between these two optimization problems.

Proposition 2. *Under Assumptions (1)-(3), the stochastic optimization problems **GmmSP-P1** and **GmmSP-P2** are equivalent in the sense that, $\left(\mathbf{y}^*, \Phi\left(\frac{-\boldsymbol{\mu}_{j,h}^T \mathbf{B}\mathbf{E}_j \mathbf{y}^*}{\|\Sigma_{j,h}^{\frac{1}{2}} \mathbf{B}\mathbf{E}_j \mathbf{y}^*\|}\right)\right)$ is an optimal solution to **GmmSP-P2** and $\tilde{\mathbf{y}}$ is an optimal solution to **GmmSP-P1**, where $\mathbf{B} \in \mathbb{R}^{2\times(|\mathcal{P}|+1)}$ is a constraint matrix defined as $\mathbf{B} \triangleq \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \beta_1 & \cdots & \beta_{|\mathcal{P}|} \end{bmatrix}$, and $\mathbf{E}_j$ is defined in Section 4.1.*

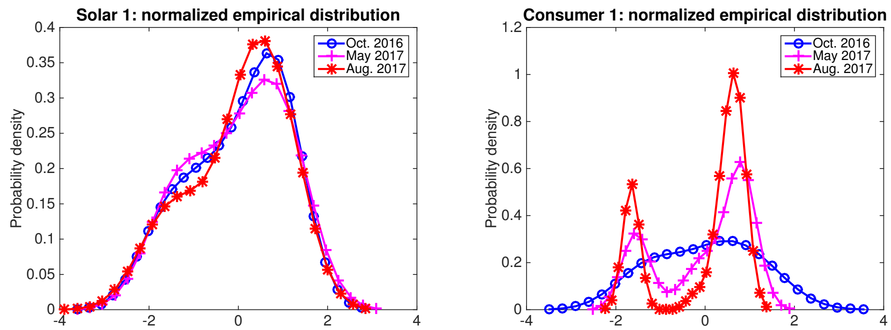
Proof. See Appendix A. □

Based on Proposition 2, we have an equivalent deterministic optimization problem **GmmSP-P2** for the matching design, which is desirable. However, **GmmSP-P2** is a non-convex optimization problem due to constraint (14). Below we propose a heuristic algorithm to solve **GmmSP-P2**. The idea is that, instead of solving a difficult optimization problem with joint variables $\mathbf{y}$ and $\epsilon_{j,h}$, we solve a number of easier problems where the variables $\epsilon_{j,h}$ are fixed. The final solution is then chosen as the one that results in the least objective. We list the detailed steps of the algorithm below.

First define the maximum number of iterations as $I_{\max}$.

(1) Initialize $\epsilon_{j,h}, \forall j, \forall h$, satisfying constraints (13) (e.g., α).

[4] Name omitted for anonymity



(a) (b)

Figure 2: Normalized empirical distributions of solar and energy consumption.

(2) Given $\epsilon_{j,h}$, solve the following second-order cone optimization problem with the optimal solution $\hat{\mathbf{y}}$, and record $\hat{\mathbf{y}}$:

$$\textbf{GmmSP-P3:} \min_{\mathbf{y}} \left(|\mathcal{D}| \sum_{r} \pi_r \bar{p}_r\right) \sum_{i\in\mathcal{P}} \sum_{j\in C} \beta_i \mathbf{c}_i^T \mathbf{E}_j \mathbf{y}$$
$$\text{s.t.} \quad \mathbf{c}_i^T \mathbf{E}_j \mathbf{y} \geq 0, \forall i \in \mathcal{P}, \forall j \in C,$$
$$\sum_{j\in C} \mathbf{c}_i^T \mathbf{E}_j \mathbf{y} \leq 1, \forall i \in \mathcal{P},$$
$$\|\Sigma_{j,h}^{\frac{1}{2}} \mathbf{B}\mathbf{E}_j \mathbf{y}\| \leq \frac{-\boldsymbol{\mu}_{j,h}^T \mathbf{B}\mathbf{E}_j \mathbf{y}}{\Phi^{-1}(\epsilon_{j,h})}, \forall j, \forall h.$$

(3) Repeat Steps (1),(2) until the number of iterations achieves $I_{\max}$.

(4) Choose the best solution $\hat{\mathbf{y}}$ that results in the least objective in **GmmSP-P3**. Obtain the matching parameters based on the definition of $\mathbf{y}$ in (6).

Since constraints (13) always hold if we initialize $\epsilon_{j,h}$ as α, we can start with α. Other initialization values can be obtained by, for example, grid search. The maximum number of iterations $I_{\max}$ is pre-selected. Obviously, the larger the value of $I_{\max}$, the lower the objective of **GmmSP-P3**, but the computational complexity accordingly increases.

4.3 Distributionally-Robust Stochastic Optimization

In Sections 4.1 and 4.2, we use specific distributions to approximate the real distributions of solar and energy consumption in the future. Such approximations, however, may result in poor performance if the future data do not fit the assumed distribution. In this section, we instead adopt a distributionally-robust approach by considering a family of distributions with some common statistical information (e.g., mean and variance).

A chance constraint is generally hard to deal with except in some very specific cases. For example, when the distribution of the uncertain parameter is Gaussian, as seen in Section 4.1, we can equivalently transform the chance constraint to a second-order cone constraint which is tractable. For other distributions, such as GMM in Section 4.2, we can only resort to heuristic algorithms. As mentioned before, by a distributionally-robust approach, instead of focusing on a specific distribution we consider a family of distributions. This would lead to a conservative approximation of the chance constraint. With appropriate definition of this family, we can make this approximation tractable.

For simplicity of notation, we define a slightly different random variable $\boldsymbol{\omega}_j \in \mathbb{R}^{|\mathcal{P}|+1}$ associated with consumer j:

$$\boldsymbol{\omega}_j \triangleq [c_j, p, \beta_2 p, \cdots, \beta_{|\mathcal{P}|} p]^T,$$

which denotes the consumption of consumer j and solar generation of all producers. Then for consumer j the chance constraint (3) can be rewritten in a compact form as

$$\Pr(\boldsymbol{\omega}_j^T \mathbf{x}_j \leq 0) \geq \alpha, \tag{15}$$

where $\mathbf{x}_j$ is defined in (5).

Define the mean and covariance matrix of $\boldsymbol{\omega}_j$ as $\boldsymbol{\mu}_j$ and Σ_j, respectively. Next we transform the chance constraint into a distributionally-robust chance constraint as

$$\inf_{\boldsymbol{\omega}_j \sim \mathcal{F}(\boldsymbol{\mu}_j, \Sigma_j)} \Pr(\boldsymbol{\omega}_j^T \mathbf{x}_j \leq 0) \geq \alpha, \tag{16}$$

where $\mathcal{F}(\boldsymbol{\mu}_j, \Sigma_j)$ denotes the distribution family that contains all distributions with mean $\boldsymbol{\mu}_j$ and covariance matrix Σ_j. It is obvious that if (16) holds then (15) holds. In the following theorem we show that (16) is equivalent to a second-order cone constraint.

THEOREM 1. *For any given $\alpha \in (0.5, 1)$, the distributionally-robust chance constraint $\inf_{\boldsymbol{\omega}_j \sim \mathcal{F}(\boldsymbol{\mu}_j, \Sigma_j)} \Pr(\boldsymbol{\omega}_j^T \mathbf{x}_j \leq 0) \geq \alpha$ is equivalent to the second-order cone constraint*

$$\|\Sigma_j^{\frac{1}{2}} \mathbf{x}_j\| \leq -\frac{1}{\sqrt{\alpha/(1-\alpha)}} \boldsymbol{\mu}_j^T \mathbf{x}_j,$$

where $\|\cdot\|$ denotes L_2 norm.

PROOF. The proof follows that of Theorem 3.1 in [11]. □

With Theorem 1, we can formulate a second-order cone program, which is a conservative approximation of **SP**. Given the mean and covariance matrix of $\boldsymbol{\omega}_j$, the linear coefficients β_i, and the target probability α, we have the following formulation:

$$\textbf{DRSP:} \min_{\mathbf{y}} \sum_{i\in\mathcal{P}} \sum_{j\in C} |\mathcal{D}| \beta_i \bar{p} \mathbf{c}_i^T \mathbf{E}_j \mathbf{y}$$
$$\text{s.t.} \quad \mathbf{c}_i^T \mathbf{E}_j \mathbf{y} \geq 0, \forall i \in \mathcal{P}, \forall j \in C,$$

$$\sum_{j \in C} \mathbf{c}_i^T \mathbf{E}_j \mathbf{y} \leq 1, \forall i \in \mathcal{P},$$

$$\|\Sigma_j^{\frac{1}{2}} \mathbf{E}_j \mathbf{y}\| \leq -\frac{1}{\sqrt{\alpha/(1-\alpha)}} \boldsymbol{\mu}_j^T \mathbf{E}_j \mathbf{y}, \forall j \in C. \quad (17)$$

We will discuss the estimation of statistics of $\boldsymbol{\omega}_j$ in Section 4.4. Comparing constraint (17) with constraint (9) in **GauSP**, we can see that the coefficients $\mathbf{b}_j$ and $\mathbf{A}_j$ in (9) play similar roles as $\boldsymbol{\mu}_j$ and $\Sigma_j^{\frac{1}{2}}$ in (17), respectively. Since $\sqrt{\alpha/(1-\alpha)}$ is greater than $\Phi^{-1}(\alpha)$ for a particular α, the upper bound in (17) is tighter than that in (9). In other words, (17) is more conservative than (9), which is easy to understand since **DRSP** is subject to a family of distributions which includes the Gaussian distribution.

4.4 Derivation of Statistics

Recall that matching parameters are recomputed at the beginning of each billing cycle. To do so, our proposed algorithms require statistics derived from past history as input. For example, in Section 4.1, we assume availability of the mean and variance of the Gaussian distribution, in Section 4.2 the statistics of GMM, and in Section 4.3 the mean and covariance matrix of $\boldsymbol{\omega}_j$. With sufficient historical data, the empirical estimates of these statistics can be derived with a reasonably high accuracy; the greater the availability of historical data, the more accurate the statistical information, and hence the greater the accuracy of the matching parameters.

We now describe how to extract the desired statistics. Assume that the data of solar and energy consumption are i.i.d., and denote $\{\boldsymbol{\omega}_{(i)}\}_{i=1}^N$ as a set of N independent samples. Then the empirical mean of the random variable $\boldsymbol{\omega}$ based on these N samples is $\hat{\boldsymbol{\mu}}_{(N)} = \frac{1}{N}\sum_{i=1}^N \boldsymbol{\omega}_{(i)}$ and the empirical covariance matrix of $\boldsymbol{\omega}$ is

$$\hat{\Sigma}_{(N)} = \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{\omega}_{(i)} - \hat{\boldsymbol{\mu}}_{(N)}\right)\left(\boldsymbol{\omega}_{(i)} - \hat{\boldsymbol{\mu}}_{(N)}\right)^T.$$

It has been shown that, when the number of the samples N is large enough the distributionally-robust chance constraint (16) will hold with a high probability (Theorem 4.1 in [11]). In addition, using historical data, we can estimate the statistics of a GMM using an approach such as the Expectation-Maximization (EM) algorithm[5].

5 ADMISSION CONTROL

In the discussion so far, we have assumed that there are enough solar resources to accommodate all consumers. In practice, the VPP may only be able to support a subset of all consumers who want to enter the system because of insufficient solar generation. In this section, we propose a greedy sequential allocation strategy for the admission of Priority I and II consumers.

5.1 Priority I Consumers

Assume that the consumers are ordered in descending order of priority (e.g. first-come-first-served). Then, consumers are added in order of decreasing priority until the VPP runs out of solar resources. Usually, we can start with a conservative estimate of j' consumers based on the operation of previous months. Specifically we begin with the consumer set $\{1, 2, \cdots, j'\}$ and design the matching parameters by some algorithm, e.g., **DRSP**. If the problem is feasible we add the next consumer and run the algorithm again. The process continues until the problem becomes infeasible (i.e., the optimization software cannot generate a solution.)

[5]In Matlab, this is implemented by the *fitgmdist()* function.

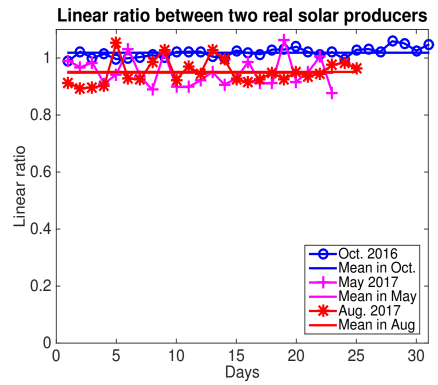


Figure 3: Linear ratio between real solar generation

5.2 Priority II Consumers

We now consider the admission of Type II consumers. For each Type II consumer, we require that its monthly energy consumption be completely satisfied by unallocated solar generation from the prior billing cycle. Denote the set of Type II consumers that request to join the program as $\mathcal{A}$ each indexed by q. Denote the prior billing cycle's energy consumption of consumer q by a_q. We assume that these consumers are ordered in descending order of priority. With historical data, we know the value of $a_q, \forall q \in \mathcal{A}$.

At the end of each billing cycle, using the matching parameters for Type I consumers, the VPP can derive the total amount of solar energy that has already been contracted for. Then the amount of the un-assigned solar generation for the whole month is U given by:

$$U = \sum_{i \in \mathcal{P}, d \in \mathcal{D}, k \in \mathcal{K}} p_i(d,k) - \sum_{i \in \mathcal{P}, d \in \mathcal{D}} p_i(d,k') \sum_{j \in C} m_{i,j}^*, \quad (18)$$

where $m_{i,j}^*$ are the matching parameters for Type I consumers, and k' is the particular time slot requested by Type I consumers. In other words, (18) gives the maximum amount of solar generation that can be used for Type II consumers.

Denote the set of the admitted Type II consumers as $\{1, 2, \cdots, q^*\}$. Then the total amount of solar energy for Type II consumers cannot exceed the available resources, i.e., there should be inequality $\sum_{q=1}^{q^*} a_q \leq U$. The method we propose is to add Type II consumers sequentially until this inequality becomes infeasible.

6 EXPERIMENTAL RESULTS

In this section, we evaluate and compare the performance of the algorithms proposed in Section 4 using real-world data.

6.1 Data Description

We were given access to one year (Sept. 2016 to Aug. 2017) of anonymized half-hourly solar generation and load data from 2 geographically close solar panels and 15 consumers by a Singapore-based VPP[6]. Note that the weather in Singapore is all-year similar with no obvious seasonal characteristics. Fortuitously, this is aligned with our i.i.d. assumption on solar data.

We compared the normalized generation traces of several solar panels for the whole year, and found that the ratio between the solar generators was close to 1. See Fig. 3 for the comparison of two representative generators for three representative months. This is in line with our assumption that solar generation from different geographically-close generators is linearly correlated.

We set one of the solar producers as the reference, and, based on this reference solar producer, we generate 8 synthetic solar producers with the linear coefficients 4, 5, 3, 2, 4, 4, 3, and 1, respectively. In principle, these linear coefficients can be arbitrary, and should reflect the actual situation on the ground. In this experiment we choose a representative set of values that generally results in feasible matchings under both our proposed algorithms and a benchmark **Oracle** algorithm described in Section 6.3.1.

In our experiment, we generate a matching for generation and load considering only one time slot each day, i.e., 12pm – 12:30pm. The solar generation of the i-th ($i \geq 3$) producer during 12pm –12:30pm on the d-th day is thus denoted by $p_i(d) = \beta_i p_1(d)$.

6.2 Experiment Description

In the experiment we focus on Priority I consumers and answer the question: if we are given a target α and historical data, what algorithm should we use to design the matching parameters?

We answer this questions using a 12-fold cross validation based on real solar and load data. That is, we divide the data into months and run 12 experiments. For each experiment, we set one month as the test month and the rest as the training months. The purpose of training is to obtain the matching matrix $\mathbf{M}$ that corresponds to the training data. In the testing phase, the derived matching matrix is used in the test month. We calculate the actual test α for each consumer, which may be different from the target α in the training phase. In the training phase, we try different values of the target α, from the set $[0.75\ 0.8\ 0.85\ 0.9\ 0.95\ 0.99]$. A high value of α indicates a more restricted chance constraint, in other words, the energy consumption would be satisfied with a high probability. The target α is the same for all consumers.

6.3 Algorithm Comparison

In Section 4, we proposed three approaches to solve the matching problem, where the chance constraints in (3) are replaced by deterministic constraints, with stochastic parameters modeled either as Gaussian random variables, Gaussian Mixture Model random variables, or based on a family of distributions. In this section, we compare these approaches. Note that under the GMM assumption, we consider two additional cases: 1) both solar generation and energy consumption are fit by GMM with two Gaussian components; we call this **GmmSP2**; and 2) both solar generation and energy consumption are fit by the 'best fit' GMM; we call this **GmmSPb**. With **GmmSPb**, the number of Gaussian components can be up to a pre-determined threshold (e.g., 5), hence leading to a better fitted model of the training data. These two algorithms are solved in the same way, and the only difference is their distributional models.

To compare the algorithms, we observe that good algorithm should be feasible both in the training and testing phases (we discuss this in more detail in Section 6.3.2), and minimize the objective function. These, hence, are our metrics. Before comparing the algorithms, we first present our benchmark oracle algorithm.

6.3.1 Oracle. As a benchmark, we describe an **Oracle** matching algorithm that knows the future perfectly, and can therefore allocate the least generation resources to guarantee the desired load. In other words, the target α under **Oracle** for all consumers is 1.

More precisely, given $p_i(d,k)$ and $c_j(d,k), \forall i, \forall j, \forall d, \forall k$, the oracle solves the following problem:

$$\begin{aligned} \textbf{Oracle:}\ \min_{\mathbf{M}} \sum_{d\in\mathcal{D},k\in\mathcal{K}} \sum_{i\in\mathcal{P},j\in C} & m_{i,j} p_i(d,k) \\ \text{s.t. } m_{i,j} \geq 0,\ \ & \forall i \in \mathcal{P}, \forall j \in C, \\ \sum_j m_{i,j} \leq 1,\ \ & \forall i \in \mathcal{P}, \\ c_j(d,k) \leq \sum_{i\in\mathcal{P}} m_{i,j} p_i(d,k),\ \ & \forall d \in \mathcal{D}, \forall k \in \mathcal{K}, \forall j \in C. \end{aligned}$$

Since there are no uncertain parameters, **Oracle** is a deterministic (linear) optimization problem, which is easy to solve. **Oracle** provides a lower bound of the objective when the energy consumption is completely met by solar.

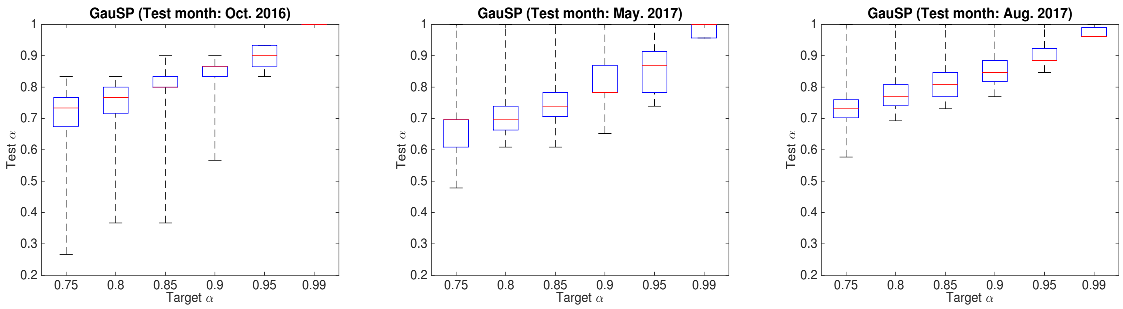
6.3.2 Feasibility. As observed earlier, there are two aspects of feasibility. First, in the training phase, the optimization problem in our proposed algorithm may become infeasible (and thus cannot generate the matching parameters) if the target α is greater than some threshold. Second, when used in the test month, the matching parameters derived in the training phase may result in a test α that is lower than the target α. We evaluate both types of feasibility in our work.

Table 1: Number of feasible training

	0.75	0.8	0.85	0.9	0.95	0.99
GauSP	12	12	12	12	12	12
GmmSP2	12	12	12	12	12	7
GmmSPb	12	12	12	12	12	12
DRSP	12	12	12	0	0	0

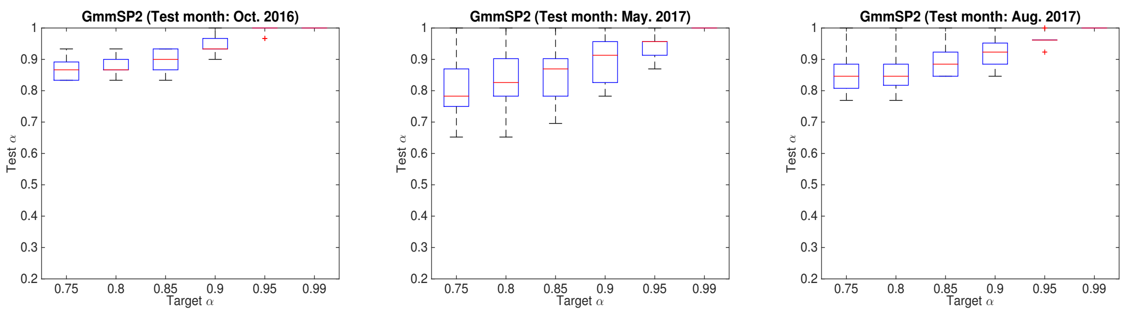
In Table 1, we investigate feasibility in the training phase. For each proposed algorithm, we list the number of times, out of the total 12 experiments, and for all values of the target α that the stochastic program was able to generate the matching parameters. We see that first, **GauSP** and **GmmSPb** are feasible in the training phase for all values of the target α; second, **GmmSP2** is generally feasible except when the target α is close to 1; and **DRSP** is only feasible for relatively small values of α. To see if this behaviour of **DRSP** was due to inadequate supply of solar energy, we further increase the linear coefficients of the 8 synthetic solar producers to 1000, 2000, 3000, 3000, 4000, 4000, 5000, and 5000. Even with this very high values for solar genearation, **DRSP** was still found to be

[6] Name omitted for anonymity



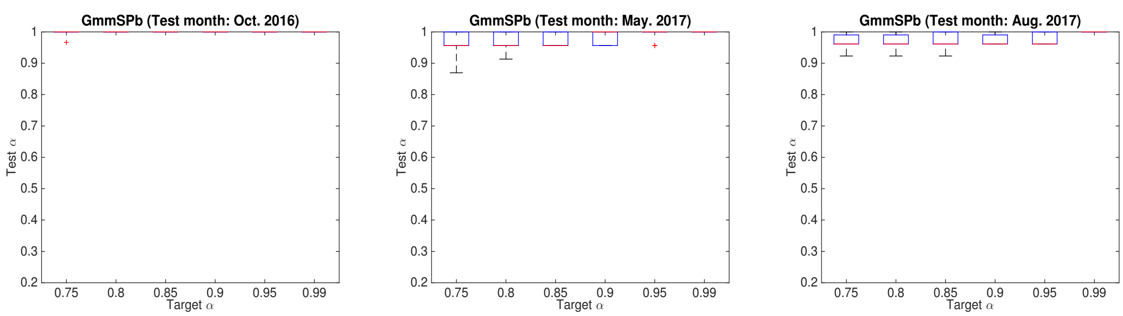
(a) (b) (c)

Figure 4: GauSP: feasibility in the test phase.



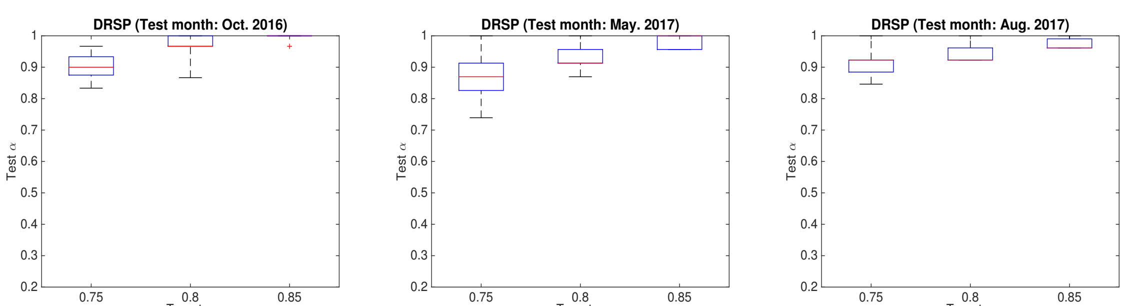
(a) (b) (c)

Figure 5: GmmSP2: feasibility in the test phase.



(a) (b) (c)

Figure 6: GmmSPb: feasibility in the test phase.



(a) (b) (c)

Figure 7: DRSP: feasibility in the test phase.

infeasible for $\alpha \geq 0.9$. Hence, we attribute the infeasibility of **DRSP** in the high range of α to its high level of conservatism, based on its need to be insensitive to the distribution of the underlying random variables.

Next we consider feasibility in the test phase. In Figs. 4-7, we use box plots to show results on the target α as a function of the test α for all 15 consumers for three different months. For each consumer, the test α is calculated as the number of days in the test month where the energy consumption is fully met by solar for the time slot under consideration, over the total number of days in the month. In each box, the central red line denotes the median, the lower and upper edge of the box are the 25th and 75th percentiles, respectively, and the whiskers extend to the minimum and maximum data points not considered outliers. Outliers, if any, are shown as red plus signs. To check feasibility, it suffices to see whether the minimum test α among all 15 consumers exceeds the target α. Due to space limitation, we only show three representative months.

From Fig. 4, it is clear that **GauSP** cannot guarantee feasibility for all consumers because, for each target α, there is at least one consumer whose constraint is unmet (except when α equal to 0.99 in Oct. 2016). For example, in the test month Oct. 16, when the target α is set as 0.75 there is one consumer whose test α is as low as 0.27. In addition, the median test α is, in general, below the target α. Thus, **GauSP** is not a desirable approach for solving the matching problem.

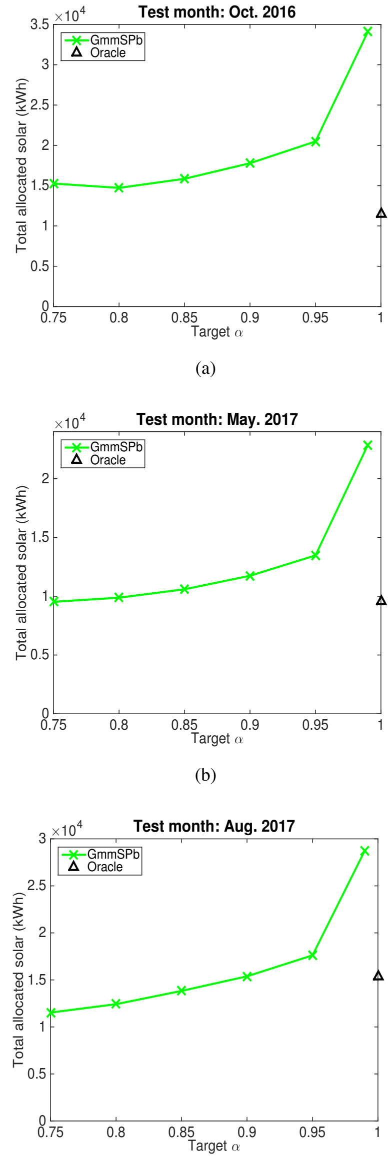
Unfortunately, similar to **GauSP**, **GmmSP2** does not provide feasibility for all consumers. Nevertheless, its feasibility performance is better than that of GauSP, since from Fig. 5 the median test α is generally greater than that of the target α.

Unlike the two prior algorithms, **GmmSPb** ensures feasibility in all cases. However, **GmmSP2** is quite conservative especially in the lower range of α, in that the test α is much higher than the target α in general. For example, in as many as seven test months all test α are greater than 0.95 regardless of the values of the target α.

Finally, in Fig. 7, we find that, **DRSP**, when feasible in the training phase, is also feasible in the test phase. Despite this, given that **DRSP** cannot find a feasible solution for typical values of $\alpha \geq 0.9$, we do not recommend it as a solution.

To sum up, based on feasibility in both the training phase and test phase, the only acceptable solution is **GmmSPb**.

6.3.3 Optimality. In this section, we compare the performance of **GmmSPb** with that of **Oracle**. Since **GauSP**, **GmmSP2**, and **DRSP** cannot ensure feasibility, we do not include these algorithms since the comparison would be meaningless without feasibility. From Fig. 8, with the target α equal to 0.99, the total amount of the allocated solar generation to Type I consumers under **GmmSPb** is 3 (Oct. 2016), 2.4 (May 2017), and 1.9 (Aug. 2017) times, respectively, that under **Oracle**. On the other hand, with the same amount of the allocated solar as that under **Oracle**, it is possible for **GmmSPb** to achieve the target α equal to 0.9 (Aug. 2017), but can also be lower than 0.75 (Oct. 2016).



(a)

(b)

(c)

Figure 8: Objective comparison between GmmSPb and Oracle.

7 DISCUSSION AND CONCLUSIONS

7.1 Discussion

We now discuss some limitations of our work and propose future directions. To begin with, our modeling of both soalr generation and load is fairly simplistic, based on the Gaussian distribution and Gaussian Mixture Models. In future work, we plan to study more sophisticated models (e.g., based on higher order Markov chain), which can better reflect the characteristics of solar and loads, and can be updated dynamically on changes. Of course, this makes the matching algorithm far more challenging.

Second, in Section 4.2, using the GMM approach, we used a heuristic to solve the the non-convex optimization problem **GmmSP-P2**. This could potentially be made more efficient, a topic for future work.

Finally, we design the matching scheme for priority I consumers based on some technical assumptions which significantly simplify our analysis. In particular, the i.i.d. assumption on the distributions of the solar and energy consumption may be far from reality. For example, solar output is highly affected by ambient temperature and solar irradiance and is in general non-stationary. Nevertheless, the matching that we produce meets the target when using the recommended **GMMSPb** approach.

7.2 Conclusions

We consider a solar-based microgrid where a VPP acts as a market maker and contracts are directly made between solar producers and consumers. For the matching design of Priority I consumers, we propose three approaches and four algorithms, i.e., **GauSP**, **GmmSP2**, **GmmSPb**, and **DRSP**, which are based on different assumptions on the distributions of the solar and energy consumption. Based on experiments in a realistic setting, we find that **GmmSPb** can always meet our design requirement and is thus recommended. We also provided sequential control strategy for the admission of Priority I and II consumers so that the solar allocation is feasible.

REFERENCES

[1] Shahrouz Alimohammadi and Dawei He. 2016. Multi-stage algorithm for uncertainty analysis of solar power forecasting. In *Power and Energy Society General Meeting (PESGM), 2016*. IEEE, 1–5.

[2] Viorel Badescu. 2014. *Modeling solar radiation at the earth's surface*. Springer.

[3] Josef Bajada, Maria Fox, and Derek Long. 2013. Load modelling and simulation of household electricity consumption for the evaluation of demand-side management strategies. In *Innovative Smart Grid Technologies Europe (ISGT EUROPE), 2013 4th IEEE/PES*. IEEE, 1–5.

[4] Sambaran Bandyopadhyay, Tanuja Ganu, Harshad Khadilkar, and Vijay Arya. 2015. Individual and aggregate electrical load forecasting: One for all and all for one. In *Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems*. ACM, 121–130.

[5] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. 2009. *Robust optimization*. Princeton University Press.

[6] Aharon Ben-Tal and Arkadi Nemirovski. 1999. Robust solutions of uncertain linear programs. *Operations research letters* 25, 1 (1999), 1–13.

[7] John R Birge and Francois Louveaux. 2011. *Introduction to stochastic programming*. Springer Science & Business Media.

[8] Stephen Boyd and Lieven Vandenberghe. 2004. *Convex optimization*. Cambridge university press.

[9] Singapore Building and Construction Authority. 2018. Singapore BCA Green Mark Assessment Criteria. (2018). https://www.bca.gov.sg/GreenMark/green_mark_criteria.html

[10] Giuseppe C Calafiore and Marco C Campi. 2006. The scenario approach to robust control design. *IEEE Trans. Automat. Control* 51, 5 (2006), 742–753.

[11] Giuseppe Carlo Calafiore and Laurent El Ghaoui. 2006. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications* 130, 1 (Jul. 2006), 1–22.

[12] Abraham Charnes and William W Cooper. 1959. Chance-constrained programming. *Management science* 6, 1 (1959), 73–79.

[13] Erick Delage and Yinyu Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research* 58, 3 (2010), 595–612.

[14] JT Flam. 2013. *The Linear Model Under Gaussian Mixture Inputs*. Ph.D. Dissertation. Ph. D. dissertation, Dept. of Elect. & Telecom. Norwegian Univ. of Science and Technology, Trondheim, Norway.

[15] Lingwen Gan, Adam Wierman, Ufuk Topcu, Niangjun Chen, and Steven H Low. 2013. Real-time deferrable load control: handling the uncertainties of renewable generation. In *Proceedings of the fourth international conference on Future energy systems*. ACM, 113–124.

[16] Michael C Grant and Stephen P Boyd. 2018. CVX: Matlab Software for Disciplined Convex Programming. (2018). http://cvxr.com/cvx/

[17] Qilong Huang, Qing-Shan Jia, Zhifeng Qiu, Xiaohong Guan, and Geert Deconinck. 2015. Matching EV charging load with uncertain wind: A simulation-based policy improvement approach. *IEEE Transactions on Smart Grid* 6, 3 (May 2015), 1425–1433.

[18] Martin Hupez, Jean-François Toubeau, Zacharie De Grève, and François Vallée. 2016. SARMA time series for microscopic electrical load modeling. In *International Work-Conference on Time Series Analysis*. Springer, 133–145.

[19] LO3 Energy Inc. [n. d.]. LO3 Exergy Technical White Paper. https://exergy.energy/wp-content/uploads/2017/12/Exergy-Whitepaper-v8.pdf. ([n. d.]).

[20] SunElectric Inc. 2018. Electricity Services. http://sunelectric.com.sg/clean-energy-buyer/. (2018).

[21] M Jurado. 1996. Statistical distribution of the clearness index with radiation data integrated over five minute intervals. In *Fuel and Energy Abstracts*, Vol. 37. Elsevier, 201.

[22] Peng Kou, Deliang Liang, Lin Gao, and Feng Gao. 2016. Stochastic Coordination of Plug-In Electric Vehicles and Wind Turbines in Microgrid: A Model Predictive Control Approach. *IEEE Transactions on Smart Grid* 7, 3 (May 2016), 1537–1551.

[23] Chao Long, JZ Wu, CH Zhang, Lee Thomas, Meng Cheng, and Nick Jenkins. 2017. Peer-to-peer energy trading in a community microgrid. In *IEEE PES General Meeting*.

[24] James Luedtke and Shabbir Ahmed. 2008. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization* 19, 2 (2008), 674–699.

[25] Yuan Luo, Satoko Itaya, Shin Nakamura, and Peter Davis. 2014. Autonomous cooperative energy trading between prosumers for microgrid systems. In *Local Computer Networks Workshops (LCN Workshops), 2014 IEEE 39th Conference on*. IEEE, 693–696.

[26] Esther Mengelkamp, Johannes Gärttner, Kerstin Rock, Scott Kessler, Lawrence Orsini, and Christof Weinhardt. 2018. Designing microgrid energy markets: A case study: The Brooklyn Microgrid. *Applied Energy* 210 (2018), 870–880.

[27] NC Clean Energy Technology Center. 2018. Database of State Incentives for Renewables and Efficiency. http://www.dsireusa.org/. (2018).

[28] Hieu T Nguyen, Duong T Nguyen, and Long B Le. 2015. Energy management for households with solar assisted thermal load considering renewable energy and price uncertainty. *IEEE Transactions on Smart Grid* 6, 1 (2015), 301–314.

[29] Michiel Nijhuis, Madeleine Gibescu, and Sjef Cobben. 2017. Gaussian mixture based probabilistic load flow for LV-network planning. *IEEE Transactions on Power Systems* 32, 4 (2017), 2878–2886.

[30] Seyyed Mostafa Nosratabadi, Rahmat-Allah Hooshmand, and Eskandar Gholipour. 2017. A comprehensive review on microgrid and virtual power plant concepts employed for distributed energy resources scheduling in power systems. *Renewable and Sustainable Energy Reviews* 67 (2017), 341–363.

[31] Danny Pudjianto, Charlotte Ramsay, and Goran Strbac. 2007. Virtual power plant and system integration of distributed energy resources. *IET Renewable Power Generation* 1, 1 (2007), 10–16.

[32] Ravindra Singh, Bikash C Pal, and Rabih A Jabr. 2010. Statistical representation of distribution system loads using Gaussian mixture model. *IEEE Transactions on Power Systems* 25, 1 (2010), 29–37.

[33] sonnen GmbH. 2018. sonnenCommunity. https://www.sonnenbatterie.de/en/node/169. (2018).

[34] Sun Sun, Min Dong, and Ben Liang. 2016. Distributed real-time power balancing in renewable-integrated power grids with storage and flexible loads. *IEEE Transactions on Smart Grid* 7, 5 (2016), 2337–2349.

[35] Joakim Widén, Nicole Carpman, Valeria Castellucci, David Lingfors, Jon Olauson, Flore Remouit, Mikael Bergkvist, Mårten Grabbe, and Rafael Waters. 2015. Variability assessment and forecasting of renewables: A review for solar, wind, wave and tidal resources. *Renewable and Sustainable Energy Reviews* 44 (2015), 356–375.

[36] Yuanming Zhang and Qing-Shan Jia. [n. d.]. A Simulation Based Policy Improvement Method for Joint-Operation of Building Microgrids with Distributed Solar Power and Battery. *IEEE Transactions on Smart Grid* ([n. d.]). accepted.

A PROOF OF PROPOSITION 2

Define a new random variable $\theta_j \in \{1, 2, \cdots, H_j\}$. Suppose that the realization of θ_j is h. Then we draw a sample of $\boldsymbol{\omega}_j$ from the h-th Gaussian component. After introducing θ_j, we have the following equivalence:

$$\begin{aligned}
&\Pr(\boldsymbol{\omega}_j^T \mathbf{z}_j \leq 0) \geq \alpha \\
\iff &\sum_{h=1}^{H_j} \Pr(\boldsymbol{\omega}_j^T \mathbf{z}_j \leq 0, \theta_j = h) \geq \alpha \\
\iff &\sum_{h=1}^{H_j} \pi_{j,h} \Pr(\boldsymbol{\omega}_j^T \mathbf{z}_j \leq 0 | \theta_j = h) \geq \alpha \\
\iff &\sum_{h=1}^{H_j} \pi_{j,h} \Phi\left(\frac{-\mathbf{z}_j^T \boldsymbol{\mu}_{j,h}}{\|\Sigma_{j,h}^{\frac{1}{2}} \mathbf{z}_j\|} \right) \geq \alpha,
\end{aligned} \tag{19}$$

where the equivalence in (19) holds because given $\theta_j = h$, $\boldsymbol{\omega}_j^T \mathbf{z}_j$ is Gaussian distributed with the mean $\boldsymbol{\mu}_{j,h}^T \mathbf{z}_j$ and variance $\mathbf{z}_j^T \Sigma_{j,h} \mathbf{z}_j$. Note that $\mathbf{z}_j$ can be represented as a linear function of the optimization variable y (defined in (6)) as $\mathbf{z}_j = \mathbf{BE}_j \mathbf{y}$. Then the equivalence is obvious [8] .