Optimal Matching of Stochastic Solar Generators to Stochastic Loads

Sun Sun, S. Keshav, Catherine Rosenberg
University of Waterloo, Canada
sun.sun,keshav,cath@uwaterloo.ca

Matthew Peloso
Sun Electric Pte Ltd, Singapore
mattpeloso@gmail.com

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 two approaches 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, the approach that assumes Gaussian mixture models for solar and loads best fits our design requirements.

CCS CONCEPTS

• **Mathematics of computing → Probability and statistics**;

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. 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 [1, 2, 4, 6].

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. A recent trend in such systems is to allow generators and consumers to directly enter into peer-to-peer contracts [1, 3–5]. In such a system, the VPP acts essentially as a market maker and supply contracts for the next billing period are 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 is met. 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. The main challenge is that the actual future values of both solar generation and energy consumption are highly variable and unknown.

In this paper, 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. Our key contributions are as follows: First, 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. Second, to solve the stochastic optimization problem, we propose two approaches based on different assumptions on the distributions of solar and loads. Third, we compare these algorithms in a realistic setting.

The remainder of the paper is organized as follows. In Section 2, we present our system model. Section 3 considers two matching approaches based on stochastic programming. Numerical results are presented in Section 4 and we conclude in Section 5.

2 SYSTEM MODEL

Consider a time slotted system with each day indexed by d. 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

e-Energy '18, June 12–15, 2018, Karlsruhe, Germany

ACM ISBN 978-1-4503-5767-8/18/06…$15.00
https://doi.org/10.1145/3208903.3208930

multiple customers, i.e., a generator can split its generation without any constraint.

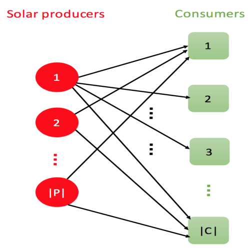

Figure 1: System model.

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 [2], a typical solar VPP, we model the consumers as follows. 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. The objective of this paper is to design the matching parameters for admitted consumers. Due to space limitation, the design of admission control strategies for consumers is deferred to Section 5 of Reference [7].

3 MATCHING DESIGN

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 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. 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:

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

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

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

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

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. Note that constraint (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.

Next, we propose two 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.

3.1 Stochastic Optimization Using Gaussian Mixture Models (GMMs)

In this section, we use the Gaussian mixture models to approximate the real distributions of solar and energy consumption. In particular, we make the following assumptions:

(1) the distribution of the solar generation and energy consumption is modeled by a GMM;
(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.

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. From Assumption (4), it suffices to 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 $\beta_1 = 1$. The value of other β_is can be evaluated by comparing the solar generation of producer i and producer 1 based on historical data.

To facilitate derivation, 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

$$\mathbf{x}_j \triangleq [1, -m_{1,j}, -m_{2,j}, \cdots, -m_{|\mathcal{P}|,j}]^T, \tag{4}$$

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

$$\mathbf{y} \triangleq [\mathbf{x}_1^T, \mathbf{x}_2^T, \cdots, \mathbf{x}_{|C|}^T]^T. \tag{5}$$

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 $m_{i,j} = \mathbf{c}_i^T \mathbf{x}_j = \mathbf{c}_i^T \mathbf{E}_j \mathbf{y}$.

Under the approximation of a GMM, we denote the distribution of the reference solar generation as $p \sim \sum_r^R \pi_r \mathcal{N}(\overline{p}_r, \sigma_r^2)$, where R is the number of Gaussian components, π_r is the probability associated with the r-th Gaussian component, $\overline{p}_r$ 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 $c_j \sim \sum_l^{L_j} \pi_{j,l} \mathcal{N}(\overline{c}_{j,l}, \sigma_{j,l}^2)$. where $\pi_{j,l}$ is the probability associated with the l-th Gaussian component, and L_j is the number of Gaussian components.

Under the approximation of a GMM, we propose a new deterministic optimization problem **GmmSP-P2**. In particular, given the statistics of GMMs introduced above, the linear coefficients β_i, and the target probability α, we have the following formulation:

$$\begin{aligned}
\textbf{GmmSP-P2:}\ \min_{\mathbf{y},\epsilon_{j,h}} \quad & \left(|\mathcal{D}| \sum_r \pi_r \overline{p}_r\right) \sum_{i\in\mathcal{P}} \sum_{j\in C} \beta_i \mathbf{c}_i^T \mathbf{E}_j \mathbf{y} \\
\text{s.t.} \quad & \mathbf{c}_i^T \mathbf{E}_j \mathbf{y} \geq 0, \forall i \in \mathcal{P}, \forall j \in C, \\
& \sum_{j\in C} \mathbf{c}_i^T \mathbf{E}_j \mathbf{y} \leq 1, \forall i \in \mathcal{P}, \\
& \sum_{h=1}^{H_j} \pi_{j,h} \epsilon_{j,h} \geq \alpha, \forall j, && (6) \\
& \epsilon_{j,h} \leq \Phi\left(\frac{-\boldsymbol{\mu}_{j,h}^T \mathbf{B} \mathbf{E}_j \mathbf{y}}{\|\Sigma_{j,h}^{\frac{1}{2}} \mathbf{B} \mathbf{E}_j \mathbf{y}\|}\right), \forall j, \forall h, && (7)
\end{aligned}$$

where $\epsilon_{j,h}$ are new optimization variables, $\Phi(\cdot)$ is the CDF of the standard Gaussian distribution, and (6) and (7) are the new introduced constraints. It can be shown that **GmmSP-P2** and the original matching problem are equivalent (see Section 4.2 of our technical report[7] for the proof), thus we can focus on **GmmSP-P2** instead.

The optimization problem **GmmSP-P2** is deterministic, which is desirable. However, **GmmSP-P2** is a non-convex optimization problem due to constraint (7). 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 (6) (e.g., α).
(2) Given $\epsilon_{j,h}$, solve the following second-order cone optimization problem with the optimal solution $\hat{\mathbf{y}}$, and record $\hat{\mathbf{y}}$:

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

(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 (5).

Note that constraints (6) always hold if we initialize $\epsilon_{j,h}$ to α.

3.2 Distributionally-Robust Stochastic Optimization

A chance constraint is generally hard to deal with except in some specific cases, e.g., when the distribution of the uncertain parameter is Gaussian (see Section 4.1 of our technical report [7] for a matching design under Gaussian assumption). In this section, we adopt a distributionally-robust approach by considering a family of distributions with some common statistical information (e.g., mean and variance). 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 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 $\Pr(\boldsymbol{\omega}_j^T \mathbf{x}_j \leq 0) \geq \alpha$, where $\mathbf{x}_j$ is defined in (4).

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

$$\inf_{\boldsymbol{\omega}_j \sim \mathcal{F}(\boldsymbol{\mu}_j, \Sigma_j)} \Pr(\boldsymbol{\omega}_j^T \mathbf{x}_j \leq 0) \geq \alpha, \tag{8}$$

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 (8) holds then the chance constraint holds. It can be shown that (8) is equivalent to a second-order cone constraint (see Theorem 1 in Section 4.3 of our technical report [7]).

Therefore, 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:

$$\begin{aligned}
\textbf{DRSP:}\ \min_{\mathbf{y}} \quad & \sum_{i\in\mathcal{P}} \sum_{j\in C} |\mathcal{D}| \beta_i \overline{p} \mathbf{c}_i^T \mathbf{E}_j \mathbf{y} \\
\text{s.t.} \quad & \mathbf{c}_i^T \mathbf{E}_j \mathbf{y} \geq 0, \forall i \in \mathcal{P}, \forall j \in C, \\
& \sum_{j\in C} \mathbf{c}_i^T \mathbf{E}_j \mathbf{y} \leq 1, \forall i \in \mathcal{P}, \\
& \|\Sigma_j^{\frac{1}{2}} \mathbf{E}_j \mathbf{y}\| \leq -\frac{1}{\sqrt{\alpha/(1-\alpha)}} \boldsymbol{\mu}_j^T \mathbf{E}_j \mathbf{y}, \forall j \in C, && (9)
\end{aligned}$$

where constraint (9) is a second-order cone constraint equivalent to (8). With sufficient historical data, the empirical estimates of the defined statistics can be derived with a reasonably high accuracy.

4 EXPERIMENTAL RESULTS

In this section, we evaluate and compare the performance of the algorithms proposed in Section 3 using real-world data.

4.1 Data Description

We have access to one year (Sept. 2016 to Aug. 2017) half-hourly solar generation and load data from 2 geographically-close solar panels and 15 customers of SunElectric, which is a VPP based in Singapre. Note that the weather in Singapore is similar throughout the year 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, which 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 pre-selected linear coefficients. In our experiment, we generate a matching for generation and load considering only one time slot each day, i.e., 12pm – 12:30pm.

4.2 Experiment Description

In the experiment we focus on answering 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 question 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]. The target α is the same for all consumers.

4.3 Algorithm Comparison

In this section, we compare the two algorithms proposed in Section 3.1 and 3.2, respectively. Under the GMM assumption, both solar generation and energy consumption are fit by the 'best' GMM model with the number of the Gaussian components up to a predetermined threshold (e.g., 5). We call this **GmmSPb**.

It is obvious that, good algorithms should be feasible in both the training and testing phases. However, infeasibility can happen as explained below. First, in the training phase, the optimization problem 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 hence infeasible. 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
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, **GmmSPb** is feasible in the training phase for all values of the target α; and second, **DRSP** is only feasible for relatively small values of α. To see if this behavior of **DRSP** was due to inadequate supply of solar energy, we further increase the linear coefficients of the 8 synthetic solar producers. We found that even with very high values for solar genearation, **DRSP** was still found to be 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.

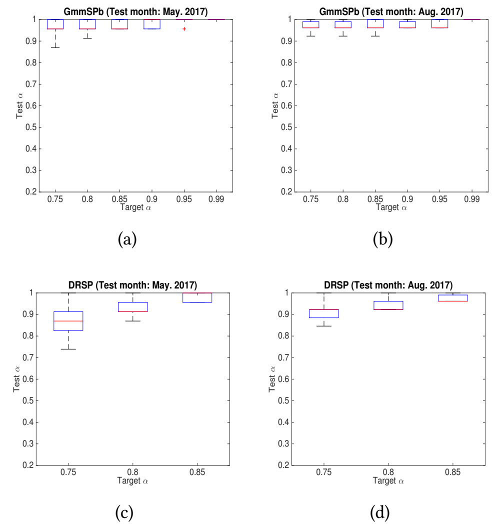

Figure 2: GmmSPb and DRSP in test phase.

Next we consider feasibility in the test phase. In Fig. 2, we use box plots to show results on the target α as a function of the test α for all 15 consumers. Due to space limitation, we only show two representative 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.

We can see that, **GmmSPb** ensures feasibility in all cases; **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**. We further compare **GmmSPb** with an oracle algorithm, in which the future information is assumed to be perfectly known. As an example for Aug. 2017, we find that with the target α equal to 0.99, the total amount of the allocated solar generation under **GmmSPb** is 1.9 times that under Oracle. On the other hand, with the same amount of solar as that under Oracle, **GmmSPb** can achieve the target α equal to 0.9. See Section 6.3.3 of our technical report [7] for details.

5 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 consumers, we propose two approaches that 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.

REFERENCES

[1] LO3 Energy Inc. 2018. LO3 Exergy Technical White Paper. https://exergy.energy/wp-content/uploads/2017/12/Exergy-Whitepaper-v8.pdf.

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

[3] 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*.

[4] 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.

[5] 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.

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

[7] Sun Sun, keshav Srinivasan, Rosenberg Catherine, and Peloso Matthew. 2018. *Optimal Matching of Stochastic Solar Generators to Stochastic Loads*. Technical Report. Waterloo, ON, Canada. https://uwaterloo.ca/computer-science/sites/ca.computer-science/files/uploads/files/cs-2018-02.pdf