Robust and Practical Approaches for Solar PV and Storage Sizing

Fiodar Kazhamiaka
University of Waterloo
fkazhami@uwaterloo.ca

Yashar Ghiassi-Farrokhfal
Erasmus University
y.ghiassi@rsm.nl

Srinivasan Keshav
University of Waterloo
keshav@uwaterloo.ca

Catherine Rosenberg
University of Waterloo
cath@uwaterloo.ca

ABSTRACT

We study the problem of optimally and simultaneously sizing solar photovoltaic (PV) and storage capacity in order to partly or completely offset grid usage. While prior work offers some insights, researchers typically consider only a single sizing approach. In contrast, we use a firm theoretical foundation to compare and contrast sizing approaches based on robust simulation, robust optimization, and stochastic network calculus. We evaluate the robustness and computational complexity of these approaches in a realistic setting to provide practical, robust advice on system sizing.

CCS CONCEPTS

• **Hardware → Batteries**; **Renewable energy**;

KEYWORDS

Energy Storage, Solar Power, Dimensioning

ACM Reference Format:
Fiodar Kazhamiaka, Yashar Ghiassi-Farrokhfal, Srinivasan Keshav, and Catherine Rosenberg. 2018. Robust and Practical Approaches for Solar PV and Storage Sizing. In *e-Energy '18: The Nineth International Conference on Future Energy Systems, June 12–15, 2018, Karlsruhe, Germany.* , 11 pages. https://doi.org/10.1145/3208903.3208935

1 INTRODUCTION

In the last few years, the prices of solar panels and storage have dropped dramatically, putting them in reach of many consumers. Companies such as Trina, Yingli, and Canadian Solar offer solar panels at a cost of less than USD 0.5/Watt, and companies such as Tesla, Sonnen, and Moixa provide off-the-shelf (albeit expensive) storage solutions.

Consider an entity that wants to purchase and install solar PV panels and storage in order to partly or completely offset grid usage[1]. How much of each should they buy? If the budget is not a constraint, then both can be generously sized, with ample slack capacity. However, given the high cost of storage, budget is often a binding constraint. Thus, we would like to provide practical guidance on the smallest possible *sizing*[2] to adequately meet the anticipated load. This is the subject of our work.

We expect many entities to face such a sizing problem in the future. These include individuals, small companies, and building operators faced with the rising cost of grid-provided electricity.

While prior work on this topic offers some insights, researchers typically consider only a single sizing approach [8, 9, 22, 25, 27, 34]. Moreover, the approaches advocated by some past researchers results in sizing decisions that may not be robust to perturbations in the inputs. In our work, we attempt to provide *practical*, *robust* advice on system sizing. To do so, we compare and contrast multiple sizing approaches, extending well-known approaches as necessary to reduce them to practice.

We make three key contributions:

- We provide a firm theoretical foundation for robust and practical sizing of both solar PV generation and storage based on three approaches: simulation, optimization, and stochastic network calculus
- We make contributions to the state-of-the-art both in stochastic network calculus and robust optimization
- We evaluate the robustness and computational complexity of these approaches in a realistic setting

2 RELATED WORK

Prior work on sizing approaches for energy storage in the presence of renewable energy sources can be grouped into three main classes: mathematical programming, simulation, and analytical methods. We sketch these approaches here, with a survey of representative work, deferring details of each approach to Section 5.

2.1 Mathematical Programming

There exist many methods for solving sizing optimization problems. In this paper, we focus on mathematical programming, which is a scenario-based approach. It requires modelling the system as a set of parameters and variables that are constrained to represent the capabilities of the physical system being modelled and an objective function representing the system target. Importantly, it typically does not model the operating policy; instead, the optimal operation is an output of the optimization program, and is dependent on the inputs. An algorithm, or *solver*, is used to search the space of feasible solutions to find the one which maximizes (or minimizes) the objective function for the given parameters. For example, in Reference [10], the problem of sizing a battery to meet the energy demands of a microgrid is formulated as a mixed-integer linear program. In Reference [16], the problem of sizing batteries and solar panels under a fixed budget to maximize the revenue of a solar

[1]The former case corresponds to that of an entity that remains grid-connected but wants to reduce its overall cost for electricity and the latter corresponds to an off-grid scenario. We treat them both identically in our work.

[2]By sizing, we refer to the power/energy size of the storage in kW/kWh and the size of solar generation in kWp.

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

ACM ISBN 978-1-4503-5767-8/18/06…$15.00
https://doi.org/10.1145/3208903.3208935

farm is formulated as a non-linear optimization problem, which is linearized to reduce the solving time.

Another notable optimization approach is to formulate a *robust optimization* problem [6], in which the objective function is optimized even when the inputs are perturbed. We do not cover robust optimization in this paper; rather, we present a simpler approach to dealing with uncertainties in the input parameters.

2.2 Simulation

Simulations are *scenario-based* sizing approaches that provide optimal system sizing for a given trajectory (i.e, a time series) for load and PV generation. They are versatile: a simulation program can evaluate different combinations of PV panel and battery sizes, calculating metrics such as loss of load probability (LOLP) [8], unmet load, and operating cost [2]. The simulated system can be operated using virtually any operating strategy, such as those proposed in [8, 9, 22, 25], and can implement complex battery models [2].

2.3 Analytical Methods

Inspired by the analogy between energy buffering by batteries and data buffering in computer networks, a variety of analytical methods have been proposed for storage capacity sizing in the literature. For example, in Reference [12] the system is modelled as a cyclic non-homogenous Markov chain, and the authors propose a steady-state analysis to determine whether a given system size is sufficient to meet a target LOLP. In Reference [3], the authors use a probabilistic tail bound on the aggregate of many regulated energy demand loads to jointly size the battery capacity and transformers for a certain LOLP in a residential setting.

Among existing analytical approaches, stochastic network calculus (SNC) [20] has shown great robustness and accuracy. This approach has been used in several applications: battery sizing to reduce reliance on diesel generators in rural areas with unreliable grid connections [30], energy demand management in a fleet of electric car charging stations [26], gaining energy flexibility through heating/cooling systems in data centres [4], supply-demand matching for prosumers [17, 29, 33], and profit maximization for renewables in electricity markets [18].

Applying stochastic network calculus to energy systems has some subtleties, due to the unique statistical properties of the underlying energy processes and the storage model in use. This has led to a series of incremental improvements in this field of research. The idea of using stochastic network calculus for energy systems was proposed in [33], where the authors assume ideal storage devices and use affine functions to separately model the long-term behavior of each of energy demand and energy supply. In Reference [17], the authors improve this approach by assuming a more realistic storage model and more complicated uni-variate envelopes for energy demand and supply. It is shown in [18] that uni-variate envelopes cannot properly capture the statistical properties of solar power due to its substantial seasonality; hence, introducing bi-variate envelopes to separately model the long term behavior of energy demand and supply. In this paper, we advance the state-of-the-art as discussed in Section 7.2.

3 GOAL

At a high level, the goal of our work is to provide robust, practical advice on how to size both solar panels and storage to partly or completely offset grid usage. This section discusses the inputs and objective of this sizing problem.

3.1 Inputs

It is reasonable to assume that an entity making a sizing decision would have access to a representative set of load traces, especially with the widespread deployment of smart meters that typically measure hourly load[3]. It is also possible to obtain hourly solar radiation traces in the geographical location of the entity, for most parts of the world [19]. In keeping with prior work, we make the assumption that these historical traces are generally representative of loads and generation. Nevertheless, the future will never exactly mimic the past; if it did, we would be able to make decisions with perfect information. Thus, the sizing decision must be robust to perturbations in the inputs, i.e., to 'small' changes in the solar irradiation or loads (we make this precise in Section 4).

In addition to generation and load traces, we need two other inputs. First, we need to know how a decision is made to either inject power into or withdraw power from the storage system. This *operating policy* can be quite complex, and is the subject of much research [7, 12, 16, 25]. Nevertheless, simple rules such as 'store excess solar energy' and 'discharge the store when solar generation is less than the load' are often adequate for most situations. We assume that, for the case of simulation and stochastic network calculus approaches, such an operating policy is provided to the sizing decision-maker. Second, it is necessary to model the behaviour of a storage system in response to power injection and discharge. We use a recently-proposed storage model in our work [23].

To summarize, we assume that the sizing decision-maker has access to the following inputs:

- A representative set of *solar traces* $S = \{S^i\}$ (for now, think of them as one trace per-year, but we discuss this point in more detail in Section 4.3).
- A representative set of load traces $D = \{D^j\}$ that constitute a set of *load scenarios*. Each load trace needs to be of the same time duration as the solar traces.
- An *operating policy*: for the simulation and stochastic network calculus approaches, the set of rules that determine when the store is charged or discharged.
- A *storage model*, along with all associated model parameters: given the current state of charge, and the applied power, this is a set of equations that computes the new state of charge.

3.2 Sizing Objective

Given the inputs in Section 3.1, our objective is to compute the "best" sizing for solar PV panels and the storage capacity. What constitutes the best choice will depend on the situation at hand. Several quality metrics are plausible[4]:

[3] Finer-grained traces would, of course, be good to have, but unlikely to be available in practice.

[4] For each application, one or multiple of these items can serve as objectives and one or multiple others as constraints.

- **Minimize LOLP:** This is the probability that the system is unable to meet the load from *solar* generation. This probability can be numerically estimated as the ratio of the time period during which the load is unmet from solar generation to the total time period under consideration.
- **Minimize unmet load (UL):** This is the total amount of load (energy) that cannot be delivered from solar generation during the period under consideration. If this load is not met from the grid, there will be user discomfort.
- **Minimize financial cost:** This is the dollar cost of purchasing the solar panel and storage system, as well as the cost of purchasing, as necessary, electricity from the grid, at its currently prevailing price. It can be viewed either as a one-time capital cost added to a periodical operational expense, due to potential purchases from the grid and the eventual degradation of the equipment from wear and tear. Note that if we can associate a cost to meeting unmet load from the grid or a diesel generator, then the cost-minimization objective incorporates the objective of minimizing unmet load.
- **Maximize robustness:** This is the degree of sensitivity of the sizing to perturbations in the input. Intuitively speaking, we wish to pick an approach such that small perturbations in the inputs result in only a small perturbation in the sizing [28]. We discuss this point in greater detail in Section 4.
- **Minimize computation time:** We expect that the sizing decision will be made on behalf of a system purchaser by a sizing decision maker. The computation cost of each such decision, therefore, should not be onerous.

In many cases, there will be a trade-off between cost on the one hand, and LOLP/UL and robustness on the other. Moreover, robustness and computation cost go hand-in-hand, since to get robust results we (generally) have to process more data and hence perform more computation. In this work, for concreteness, we focus on minimizing the cost of solar PV and storage, subject to meeting a certain LOLP constraint. Using other optimization objectives is also possible, and discussed at greater length in Section 7.

4 THE IMPACT OF NON-STATIONARITY

A key insight in our work is that the traces which serve as input to any sizing approach may neither be stationary nor representative of the future. We discuss this next.

4.1 Traces, Trajectories, and Stochastic Processes

A solar or load trace with T entries of the form $(time, value)$ is a trajectory instantiated from a stochastic process, which is defined as a set of random variables indexed by time. That is, $S^i(t)$, the t^{th} element of the i^{th} solar trace (resp. $D^j(t)$, the t^{th} element of the j^{th} load trace) is a value assumed by the random variable $S(t)$ (resp. $D(t)$) from a corresponding distribution. Hence, we can fully characterize the *historical* solar (resp. load) stochastic process by defining joint distribution of a set of T random variables, one for each time step. Assuming independence of each time step, we can decouple these distributions, allowing us to use the set S (resp. D) of solar generation (resp. load) traces to estimate parameters for each of the T distributions. For example, the numerical mean of the t^{th} time step of the set of traces can be viewed as an estimate of the mean of the t^{th} distribution and the sample variance of this set is an estimate of its variance. Thus, with sufficient data, we can use standard statistical techniques to find the best-fitting distributions that characterize a set of traces.

Given this characterization of historical stochastic processes, what can we say about the future? Suppose that the generation and load stochastic processes were time-invariant. Then, once the historical processes are characterized, the future is also 'known' in that we can generate potential future trajectories by generating a random value per time step from the corresponding distribution. We can then choose a sizing that meets our sizing objectives not just for historical trajectories, but also for potential future trajectories.

However, this naive approach has three problems. First, even assuming independence of time steps, it is onerous to define T separate distributions, since T can be very large, on the order of 10,000-100,000 values. Second, there is no guarantee that a stochastic process parametrized based on historical traces will adequately represent the future. Third, we do not have any guidelines on how much data is 'enough.' To solve these problems, we need to take a closer look at the generation and load stochastic processes.

4.2 Causes of Non-Stationarity

A key observation is that both the solar and load stochastic processes are *non-stationary*[5] due to three effects:

(1) **Diurnality.** For example, the distribution of the r.v. $S(t)$ corresponding to a time slot t at night will differ from the distribution of an r.v. corresponding to a time slot at mid-day.
(2) **Seasonality.** For example, the distribution of the r.v. $S(t)$ corresponding to a time slot t at mid-day in winter will differ from the distribution of an r.v. corresponding to a time slot at mid-day in summer.
(3) **Long-term trends.** For example, the distribution of the r.v.'s $D(t)$ and $S(t)$ corresponding to a time slot t at the start of a trace may differ from their distribution for a time slot later in the trace.

Given this non-stationarity, we need to be careful both in characterizing historical generation and load stochastic processes and extrapolating from the past to the future. We consider each in turn.

4.3 Stochastic Process Parametrization

Recall that the parameters of the stochastic process, i.e., corresponding to each of the T distributions constituting the process, are derived from solar and load traces. Given that the process has both diurnal and seasonal non-stationarity effects, the solar and load traces must be both *detailed* enough and *long* enough to capture all three effects. More precisely, :

- The traces should have sufficient temporal resolution to capture diurnal changes. That is, the time step should be sufficiently small to have an adequate number of values for each part of the day.
- The traces should be long enough to capture seasonality, i.e., at least one year in duration, if not longer.

[5]Roughly speaking, this means that statistics computed from two different random sub-samples of the traces can differ.

- The traces should be long enough to capture any long-term trends in load (we assume that solar generation is stationary at the time scale of a year).
- There should be enough traces in the set of traces so that there are sufficient samples to adequately estimate the parameters of each distribution.

Ideally, we would have access to per-minute load and generation traces spanning several decades. Then, setting $T = 60 * 24 * 365 = 525,600$, i.e., per minute of the year, we would obtain multiple sample values for each time step, allowing us to estimate, with adequate confidence, the parameters of each of the T solar generation distributions, and potentially long-term trends in the load (for example by fitting a linear regression to the residual after accounting for diurnal and seasonal effects).

In practice, it is unlikely that such data traces are available. Load and solar generation are often measured at a time scale of 30 minutes or longer, and it is rare to have more than a year or two of data. Given this, we propose the following pragmatic approach.

First, given the length of realistic data traces, we believe that mathematically well-founded long-term trend analysis is likely to be impossible. Thus, we model long-term trends with a simple approximation, as discussion in Section 4.4.

Second, we need to extract sub-processes of the solar and load processes that are stationary despite seasonality and diurnality. Consider the following: suppose we can assume that the solar generation process is stationary from 8am to 9am on any day of a given season[6]. Then, we can model this sub-process as being drawn from a set of i.i.d random variables, meaning we can model it completely by estimating the parameters of the corresponding distribution.

Thus, we divide the year into a small number of *seasons*, such as four (winter, spring, summer, and fall) and assume that seasonality can be ignored within each such interval. Moreover, to model diurnality, we estimate the parameters of a distribution for a set of time periods (of typically 30 minutes to one hour) for each season. Finally, recognizing that loads during working days and non-working days can differ significantly, we separately model their distributions.

As a concrete example, given a few years of solar and load traces with time granularity of one hour, we would estimate distribution parameters for solar generation for the time period 12:00am-1:00am for winter, 1:0am-2:00am for winter, ..., 11:00pm-12:00am for fall, for a total of 24*4 distributions and estimate distribution parameters for load for the period 12:00am-1:00am on working day for winter, 12:00am-1:00am on a non-working day for winter, ..., 11:00pm-12:00am on a working day for fall, 11:00pm-12:00am for a non-working day for fall, for a total of 2*24*4 distributions[7].

4.4 From Past to Future

Given a set of parametrized distributions, corresponding to the historical solar and load stochastic processes, we need to generate traces corresponding to potential future load and solar trajectories. If this set of distributions can be assumed to be time-homogeneous, i.e., without long-term trends, then we can generate as many trajectories as we wish from these distributions. To take long-term trends into account, we propose to use a simple expedient: to boost the mean of each distribution in our set of load distributions by a factor ρ_D which models long-term trends in load, and ρ_S for solar. To first-order, ρ_D models a fractional increase/decrease in mean load, which we anticipate to be the impact of appliance acquisition or disposal, while ρ_S models panels degradation.

We now discuss three robust sizing approaches that base their sizing decisions on 'traces of the future.' We defer a discussion on how to evaluate the robustness of these approaches to Section 6.

5 ROBUST SIZING APPROACHES

The sizing decision, that is, choosing the size of the store B (in kW/kWh) and of the solar panels C (in kW) to meet one or more of the objectives discussed in Section 3.2, can be made using many different approaches. In this section, we present three representative approaches.

We make the following assumptions:

- For simplicity, and for reasons of space, we assume that the goal is to find the minimum-cost storage and solar PV sizes that meet a certain LOLP criterion. Generalizing this approach to other criteria, such as minimizing unmet load, is discussed in Section 7.
- We assume that we have available m solar generation and n load traces $S^i, D^j, i, j \in [1, m], [1, n]$ respectively. We call each of the mn combinations of traces a 'trace pair.' Trace pairs correspond to potential futures and can either be measurements from the past or can be generated by a parametrized stochastic process, as discussed in Section 4.4.
- We only size the storage system for energy, not for power, since sizing for power is typically trivial (the power rating of the storage system must exceed the sum of power draws of the set of simultaneously active load components).
- For the simulation and stochastic network calculus approaches, we futher assume the storage system energy capacity can only take one of b different values and that the solar panel size can only take on one of c different values. Hence, these two approaches do a grid search through bc pairs of solar PV panel sizes and storage capacity sizes and return the optimum sizing with the minimum cost which also satisfies the target LOLP.
- We assume that if a certain combination of storage and PV values results in a certain LOLP, then larger values of either storage or PV will always result in lower values of LOLP. This allows us to use a greedy grid search heuristic.

We denote the number of time steps in the load and solar generation traces by T. π_B is the price for one unit of battery (i.e., 1 cell), π_C is the price for one unit of PV panel generation. We normalize the solar generation S^i trace, so that it represents the generation from a unit PV module. Finally, the LOLP target is denoted ϵ.

5.1 Optimization

In this approach, we formulate an optimization program for solar panel and battery sizing with the objective of minimizing the capital

[6] Given our independence assumption, we do not model the relationship between this sub-process and the sub-process corresponding to 9am-10am, or the same sub-process corresponding to the same time duration in another part of the year. That is, we do **not** use a time-series model, and hence do not directly model temporal effects that may persist across time steps.

[7] If traces have a time-granularity finer than one hour, and we expect the load/solar process to be stationary for one hour, then we can aggregate all the trace values within the stationary interval when estimating distribution parameters.

cost of the system, subject to physical system and LOLP constraints. We do not specify the operating policy, allowing it to be decided by the optimization solver. This allows us to compute the best possible sizing. In this sense, although the sizing decision made by the optimization program is a potentially-unattainable lower bound, it measures the level of sub-optimality in the operating policy used in the two other approaches.

Our approach has two phases. In the first phase, for each trace pair, we compute the optimal sizing, assuming optimal operation. This gives us *mn* sizings. In the second phase, we use a bivariate Chebyshev bound [31] to compute a *robust* sizing that is insensitive to the details of individual traces. We discuss each phase in turn.

5.1.1 Phase 1. For phase 1, define P_c to be the charging power, P_d to be the discharging power, P_{dir} to be the power that flows directly from solar to load, and b to be the energy content. The size of the battery is B and the generation capacity of the panels is C. Figure 1 shows a labelled system diagram.

The battery model used here is Model 1* from [23] with the following parameters: η_c (resp. η_d) the charging (resp. discharging) efficiency, α_c (resp. α_d) the charging (resp. discharging) rate limit, u_1, v_1, u_2, v_2 used to characterize the power-dependent lower and upper limits on the energy content (see constraint (6)). Recall that the data trace for solar generation S is for one unit of PV panel and that D denote the household electricity load. The duration of a time-step is T_u and the number of time-steps in a data trace is T.

Given a trace pair $(S(t))$, $(D(t))$, and storage parameters η_c, η_d, α_c, α_d, u_1, v_1, u_2, v_2, costs π_B, π_C, and trace parameters T_u and T, the problem can be formulated as:

$$\min_{\substack{B,C,P_c,P_d,\\ P_{dir},I,\gamma,b}} \pi_B B + \pi_C C \tag{1}$$

subject to

$$P_c(t) + P_{dir}(t) \le S(t)C \quad \forall t \tag{2}$$
$$P_{dir}(t) + P_d(t) = D(t) - \gamma(t) \tag{3}$$
$$b(0) = B/2 \tag{4}$$
$$b(t) = b(t-1) + P_c(t)\eta_c T_u - P_d(t)\eta_d T_u \quad \forall t \tag{5}$$
$$u_1 P_d(t) + v_1 B \le b(t) \le u_2 P_c(t) + v_2 B \quad \forall t \tag{6}$$
$$0 \le P_c(t) \le B\alpha_c \quad \forall t \tag{7}$$
$$0 \le P_d(t) \le B\alpha_d \quad \forall t \tag{8}$$
$$I(t) \in \{0,1\} \quad \forall t \tag{9}$$
$$B, C, P_{dir}(t), \gamma(t), b(t) \ge 0 \quad \forall t \tag{10}$$
$$1/T \sum_{t=1}^{T} I(t) \le \epsilon \tag{11}$$
$$I(t) \le \gamma(t) Z \quad \forall t \tag{12}$$
$$\gamma(t) \le I(t) D(t) \quad \forall t \tag{13}$$
$$P_c(t) P_d(t) = 0 \quad \forall t \tag{14}$$

where constraint (2) states that the sum of what goes into the battery and directly towards the load is bounded by the solar generation, $\gamma(t)$ is the load that is not met from solar generation at time t (it is always $\le D(t)$), constraints (5)-(8) represent the battery model, and $I(t)$ is an binary variable used to indicate if the load is met or not

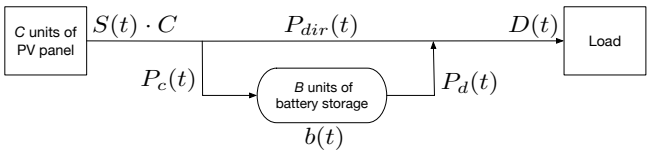

Figure 1: System diagram

in time-step t ($I(t) = 1$ means the load is not met). Constraint (12) ensures that $I(t)$ is zero if $\gamma(t) = 0$ (Z is a large positive constant), constraint (13) ensures that $I(t)$ is one if $\gamma(t) > 0$ and constraint (11) is the LOLP constraint. The last constraint forbids simultaneous charging and discharging and it was shown in [16] that it can be ignored which makes the problem a Linear Integer Program (LIP). Note that in this problem B and C are real numbers, i.e., they are not limited to the pre-defined values used for the other two approaches.

5.1.2 Phase 2. In phase 2, relying on a recent advance in the area of multivariate concentration bounds–specifically, Theorem 1 in Reference [31]– we use the set of $mn = N$ sizings to compute a multivariate Chebyshev bound. Specifically, we calculate the 2×2 empirical covariance matrix Σ_N and empirical means $\mu_{N,B}$ and $\mu_{N,C}$ of the N values of B and C, and then compute the ellipsoid of (B', C') points that bounds the 95% probability density mass of the (B, C) distribution as:

$$\begin{bmatrix} B' - \mu_{N,B} \\ C' - \mu_{N,C} \end{bmatrix}^T \Sigma_N^{-1} \begin{bmatrix} B' - \mu_{N,B} \\ C' - \mu_{N,C} \end{bmatrix} = \Lambda^2 \tag{15}$$

where Λ^2 satisfies the following equation:

$$\frac{2(N^2 - 1 + N\Lambda^2)}{N^2\Lambda^2} = 0.05 \tag{16}$$

We then select the least-cost (B', C') value on upper right quadrant[8] of this empirical ellipsoid boundary (see Figure 3 for an example) as the least-cost, robust sizing. To our knowledge, this approach to robust optimization is novel, differing from prior approaches that rely on formulating a 'robust counterpart' to an optimization problem [5].

If we are confident that the empirical means and covariance have converged to the population mean after N samples, we can use a tighter ellipsoid characterized by a Λ^2 that satisfies the following:

$$\frac{2}{\Lambda^2} = 0.05 \tag{17}$$

which is equivalent to Eq. 16 when $N \to \inf$, i.e. the analytical bound, in accordance with Theorem 2 in [31].

5.1.3 Computation cost. The inputs to the optimization program are the solar and load traces, each of size $O(T)$, for a total size of $O(T)$. Asymptotically, this is also the number of variables in the program. Denoting by $Q = O(T)$ the number of variables and $L = O(T)$ as the number of bits of input to the algorithm, even for a linear program, which is far more computationally efficient than a linear integer program, the best-known approach, the interior-point method,

[8]The upper right quadrant contains only (B', C') pairs which exceed $(\mu_{N,B}, \mu_{N,C})$ and represent robust sizing options.

requires a runtime of $O(Q^{3.5}L^2.logL.loglogL) \approx O(T^{5.5}.logT)$ [32]. Since we need *mn* such runs, and our problem is integer, the total complexity is lower bounded by $O(mnT^{5.5}.logT)$.

5.2 Simulation

In this approach, we run a total of up to *mnbc* simulations: one for each of the *bc* choices of storage and solar sizing, and for each of the *mn* trace pairs, to determine the corresponding LOLP.

Specifically, for each trace pair, for each potential sizing choice, and for each time step $t \in [1, T]$, we determine the availability of solar power $S(t)$ and the load $D(t)$. Depending on these values, the storage and PV sizes under test, and the given operating policy, we use the storage model to either charge or discharge the store, updating its SoC as in Eq. 5. If we find that there is a need to discharge the store, but its SoC is zero, then we mark that the load is unmet from solar generation for this time step. At the end of each simulation, we empirically compute the LOLP ϵ for this sizing.

To minimize the number of simulations, we use a simple greedy heuristic. Specifically, we order all *bc* choices of storage and solar generation in order of capital cost, lowest to highest. We then simulate the choices in order of increasing cost. The output of this approach is an set of *mn* sizing choices (one per trace pair) for solar and storage sizing that are the least-cost and also satisfy a given LOLP criterion.

As in Section 5.1.2, we view this set as a set of samples drawn from a bivariate (B, C) distribution, and compute a robust sizing from a Chebyshev bound.

Note that the computation cost of this approach is $O(mnbcT)$, since each time step takes $O(1)$ computation time, there are T steps per simulation, and *mmbc* simulations.

5.3 Stochastic Network Calculus (SNC)

In this approach, we characterize the net power arrival to the battery using lower and upper bounds computed on the ensemble of input traces. Then, we use stochastic network calculus to compute the LOLP for each choice of storage and solar panel size (B, C). The output is a statistical characterization of the LOLP ϵ as a function of the selected choices of (*B*,*C*). We then use the greedy heuristic discussed in Section 5.2 to compute the least-cost sizing that meets the LOLP criterion. Since SNC sizing is known to be robust to small perturbations in the input traces, we view this least-cost sizing as being robust to the input traces.

5.3.1 Details. Denote by $P_c(t)$ and $P_d(t)$, respectively, the charging and discharging power from and to the battery, given by an operating policy corresponding to:

$$P_c(t) = \min\left([S(t) - D(t)]_+, \alpha_c B\right) \tag{18}$$

$$P_d(t) = \min\left([D(t) - S(t)]_+, \alpha_d B\right) \tag{19}$$

That is, we assume that the operating policy is as follows: the battery is charged whenever the generation $S(t)$ exceeds the load, and discharged otherwise, with a bound $B\alpha_c$ on the charge power and a bound $B\alpha_d$ on the discharge power (matching Eq. 7-8). Different operating strategies will require these equations to be modified appropriately.

Define the net power inflow to the battery at any time as the overall net equivalent power injected to the battery, which is

$$P_{net}(t) = \eta_c P_c(t) - \eta_d P_d(t) \tag{20}$$

Note that at any time instant t, $P_c(t).P_d(t) = 0$, and $P_{net}(t)$ can be expressed as

$$P_{net}(t) = \begin{cases} \eta P_c(t) & \text{if } S(t) \geq D(t) \\ -\eta_d P_d(t) & \text{if } S(t) < D(t) \end{cases} \tag{21}$$

Please also note that while $P_c(t), P_d(t) \geq 0$ at any time t, $P_{net}(t)$ can be positive or negative at any time.

According to the battery model in Reference [23], the instantaneous available battery capacity is a function of charge/discharge power to/from the battery. The larger the charge/discharge power the lower the instantaneous available battery capacity. This means that apart from the power constraints discussed above, we also have energy constraints in battery operations. To be more precise, the battery state of charge $b(t)$ at any time t must satisfy $B_1(t) \leq b(t) \leq B_2(t)$, where

$$B_1(t) = u_1 P_d(t) + v_1 B \tag{22}$$

$$B_2(t) = u_2 P_c(t) + v_2 B \tag{23}$$

With the above notation, the state of charge of a battery $b(t)$ at any time t can be, recursively, expressed by

$$b(t) = [b(t-1) + P_{net}(t)T_u]_{B_1(t)}^{B_2(t)} \tag{24}$$

where $[.]_{B_1(t)}^{B_2(t)}$ truncates the inner expression to a lower bound $B_1(t)$ and an upper bound $B_2(t)$, or equivalently

$$b(t) = \begin{cases} B_1(t) & \text{if } b(t-1) + P_{net}(t)T_u < B_1(t) \\ B_2(t) & \text{if } b(t-1) + P_{net}(t)T_u > B_2(t) \\ b(t-1) + P_{net}(t)T_u & \text{otherwise} \end{cases} \tag{25}$$

Recall that LOLP is the probability that at time t the energy to be withdrawn from the battery reaches the lower battery capacity boundary (i.e., $B_1(t)$) and hence the demand cannot be met at that time. Mathematically speaking, this means that

$$LOLP = \mathbb{P}\{b(t-1) + P_{net}(t)T_u < B_1(t)\} \ \forall t \tag{26}$$

where $b(t-1)$ can be computed recursively according to Equation (24). This recursive equation can be turned into a complicated min-max non-recursive equation. At any time t, the min-operand searches, in the range of $[0, t-1]$, for the last reset time before t, which is the last occurrence of a loss of load event. As shown in Reference [30], instead of applying the min-operand to t scenarios in $[0, t-1]$, we can highly accurately approximate LOLP by only accounting for only two scenarios: (I) the reset time occurs at the last time slot $t-1$ and (II) there has been no reset time $t = 0$.

Hence, define $LOLP^{\mathrm{I}}$ and $LOLP^{\mathrm{II}}$ representing LOLP, respectively under the two scenarios mentioned above and LOLP can be approximated by

$$LOLP \approx \min\left(LOLP^{\mathrm{I}}, LOLP^{\mathrm{II}}\right) \tag{27}$$

Under scenario I, the last reset time always happens at the previous time slot. The LOLP under this scenario can be closely approximated by a battery-less scenario. This means that $LOLP^{\mathrm{I}}$ can be

approximated by the likelihood that the instantaneous demand is larger than the instantaneous supply, or mathematically speaking:

$$LOLP^{\mathrm{I}} \approx \mathbb{P}\{D(t) > S(t)\} \tag{28}$$

Under scenario II, when there is no reset time until time t, then the battery state of charge has never reached its lower boundary and $LOLP^{\mathrm{II}}$ is given by:

$$LOLP^{\mathrm{II}} = \mathbb{P}\left\{B_2(0) - \sup_{0\le s\le t}(-P_{net}(s,t))T_u < B_1(t)\right\} \ \forall t \tag{29}$$

$$= \mathbb{P}\left\{v_2 B - \sup_{0\le s\le t}(-P_{net}(s,t))T_u < B_1(t)\right\} \ \forall t \tag{30}$$

$$= \mathbb{P}\left\{\sup_{0\le s\le t}\left(\frac{u_1 P_d(t)}{T_u} - P_{net}(s,t)\right) < \frac{v_2 - v_1}{T_u}B\right\} \ \forall t \tag{31}$$

where in the second line we assume that the battery is initially full and $P_{net}(s,t)$ is defined as

$$P_{net}(s,t) = \sum_{k=s+1}^{t} P_{net}(k) \tag{32}$$

Suppose that $\mathcal{G}(s,t;\sigma)$, given by

$$\mathcal{G}(s,t;\sigma) = \frac{u_1 \mathbf{E}[P_d(t)]}{T_u} - \sum_{k=s+1}^{t} \mathbf{E}[P_{net}(k)] + \sigma \tag{33}$$

for some $\sigma \ge 0$ is a sample path envelope on $\frac{u_1 P_d(t)}{T_u} - P_{net}(s,t)$. This means that for any $\delta \ge 0$

$$\mathbb{P}\left(\sup_{0\le s\le t}\left(\frac{u_1 P_d(t)}{T_u} - P_{net}(s,t) - \mathcal{G}(s,t;\sigma)\right) > \delta\right) \approx p_1 e^{-\lambda\delta} \tag{34}$$

For any choice of σ, define and compute $\omega(\sigma)$ to be

$$\omega(\sigma) = \left[\sup_{s,t}\mathcal{G}(s,t;\sigma)\right]_+ \tag{35}$$

Combining sample path definition with Equation (31), we have

$$LOLP^{\mathrm{II}} \le$$

$$\mathbb{P}\left\{\sup_{0\le s\le t}\left(\frac{u_1 P_d(t)}{T_u} - P_{net}(s,t) - \mathcal{G}(s,t;\sigma)\right) > \frac{v_2 - v_1}{T_u}B - \omega(\sigma)\right\} \ \forall t \tag{36}$$

While Equation (36) holds as an upper bound on $LOLP^{\mathrm{II}}$ for any choice of σ, we can make this upper bound tighter and use it as an approximation, when the right-hand side is minimized over all choices of $\sigma \ge 0$; i.e.,

$$LOLP^{\mathrm{II}} \approx \min_{\sigma}\mathbb{P}\left\{\sup_{0\le s\le t}\left(\frac{u_1 P_d(t)}{T_u} - P_{net}(s,t) - \mathcal{G}(s,t;\sigma)\right) > \frac{v_2 - v_1}{T_u}B - \omega(\sigma)\right\} \ \forall t \tag{37}$$

Denote σ^* the argmin value of the above minimization. Then, combining Equation (37) and Equation (34), yields

$$LOLP^{\mathrm{II}} \approx p^{\mathrm{II}} e^{-\lambda^{\mathrm{II}}(\frac{v_2-v_1}{T_u}B - \omega(\sigma^*))} \tag{38}$$

and finally LOLP can be computed, by inserting Equations (28-38) into Equation (27). There are three unknowns in Equation (27) to be evaluated: $LOLP^{\mathrm{I}}$, p^{II}, and λ^{II} that can be computed as discussed next.

5.3.2 Algorithm. We now translate this mathematical presentation into an algorithm. Suppose that we have m solar generation traces and n load traces; i.e., $S^i, D^j, i, j \in [1,m], [1,n]$ respectively. This leads to mn different sample paths of the stochastic processes $P_d^{i,j}$ and $P_{net}^{i,j}$ for a time horizon of length T. We can compute LOLP, using stochastic network calculus, following these steps in turn:

Step 1: Compute $LOLP^{\mathrm{I}}$: This is a point-wise probability, expressed in Equation (28) and can be computed as

$$LOLP^{\mathrm{I}} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{t=0}^{T}\mathbb{I}\left(D^j(t) > S^i(t)\right)}{mnT} \tag{39}$$

where $\mathbb{I}(x)$ is the indicator function, which is 1 if x is true and 0, otherwise.

Step 2: Initialize $LOLP = LOLP^{\mathrm{I}}$ and $\sigma = 0$.

Step 3: Construct $Y^{i,j,t}(\sigma)$ for the given σ: To compute p^{II} and λ^{II}, we should first construct the set of all $Y^{i,j,t}(\sigma)$ for any ensemble trace $i,j \in [1,m],[1,n]$ and any time $t \le T$, defined as:

$$Y^{i,j,t}(\sigma) = \sup_{0\le s\le t}\left(\frac{u_1 P_d^{i,j}(t)}{T_u} - P_{net}^{i,j}(s,t) - \mathcal{G}(s,t;\sigma)\right) \tag{40}$$

It can be shown that $Y^{i,j,t}(\sigma)$ can be expressed, recursively, by

$$Y^{i,j,1}(\sigma) = \frac{u_1}{T_u}(P_d^{i,j}(1) - \mathbf{E}[P_d(1)]) - (P_{net}^{i,j}(1) - \mathbf{E}[P_{net}(1)]) - \sigma \tag{41}$$

$$Y^{i,j,t}(\sigma) = \frac{u_1}{T_u}(P_d^{i,j}(t) - \mathbf{E}[P_d(t)]) - (P_{net}^{i,j}(t) - \mathbf{E}[P_{net}(t)]) + \max\left(Y^{i,j,t-1}(\sigma) - \frac{u_1}{T_u}(P_d^{i,j}(t-1) - \mathbf{E}[P_d(t-1)]),\ -\sigma\right) \tag{42}$$

Step 4: Compute p^{II} and λ^{II}: Using $Y^{i,j,t}(\sigma)$ from the previous step, p^{II} is the likelihood of $Y^{i,j,t}$, being positive, or

$$p^{\mathrm{II}} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{t=1}^{T}\mathbb{I}\left(Y^{i,j,t}(\sigma) > 0\right)}{mnT} \tag{43}$$

and λ^{II} can be obtained as the exponent of fitting an exponential distribution to the following set

$$\lambda^{\mathrm{II}} \sim \text{Exponential}\left(\left\{Y^{i,j,t}(\sigma) \mid Y^{i,j,t}(\sigma) > 0\right\}\right) \tag{44}$$

Step 5: Update LOLP: Compute $LOLP^{\mathrm{II}}$, according to Equation (38). Set if $LOLP > LOLP^{\mathrm{II}}$, update it to $LOLP = LOLP^{\mathrm{II}}$.

Step 6: Vary σ and iterate: Increment σ. If $\sigma < \frac{v_2-v_1}{T_u}B$ then, go back to Step 3. Otherwise, terminate.

5.3.3 Computational complexity. There is a single free parameter σ over which we minimize the LOLP bound. Define s to be the number of σ values we test. Each test requires $O(T)$ time to construct the set Y and calculate the parameters of $LOLP^I$ and $LOLP^{II}$ for each of the mn trace combinations. This needs to be repeated over all combinations of battery and PV panel sizes, hence the total complexity is $O(mnbcsT)$.

6 NUMERICAL EVALUATION

For concreteness, we numerically evaluate our approaches using four years of data collected from a number of homes in the Pecan Street Dataport [1], and present the results for one typical home in

Table 1: Battery model parameters

Parameter	α_c	α_d	u_1	u_2	v_1	v_2	η_c	η_d
Value	1	1	0.053	-0.125	0	1	0.99	1.11*

*includes inverter inefficiencies of ~10%

the dataset. To evaluate the cost of a particular sizing, we set π_C, the installed cost of solar panels, to be USD 2.50/W and π_B, the cost of storage, to be USD 460/kWh[9], with battery parameters corresponding to a Lithium-Nickel-Manganese-Cobalt battery chemistry [13, 24] as summarized in Table 1. The battery model is Model 1* in Reference [23] and we use the simple operating policy of charging the battery when solar generation exceeds the load, and discharging the battery when load exceeds solar generation.

Due to the small size of the trace data set (four years of data, yielding 16 trace pairs), we find the sizing to be quite loose (we discuss this further in Section 6.3). In order to synthetically increase the number of available traces, we use the following boosting procedure: we find the best-fit Gaussian Mixture Model using up to four components, evaluated using the Bayesian Information Criterion, for each hour of trace data for each season, giving us 24*4 PV generation and 48*4 load models. We then generate 10 solar and 10 load synthetic traces from these models, giving us 100 trace pairs.

6.1 Convergence

Recall that when using either simulations or optimization, the recommended sizing is based on a statistical measure (i.e., the Chebyshev bound) on the underlying sample of computed sizings. Hence, it is progressively refined as we evaluate more scenarios (trace pairs). In contrast, when using SNC, we compute a single envelope on the entire ensemble of scenarios. Nevertheless, as we add more scenarios to the ensemble, this results in a convergence of the sizing determined by the envelope computed from this ensemble.

We illustrate this convergence by showing the mean sizing of storage B and PV generation C, as a function of the number of scenarios, for $\epsilon = 0.1$ in Figure 2. We find that in all three approaches, mean B (blue) and C (orange) values converge after about 25 scenarios. For smaller ϵ, more scenarios are required for convergence. In the rest of our evaluation we compute results using 25 scenarios.

6.2 Sizing

We now compare the sizing that results from the three approaches in Figure 3. For the simulation and optimization approaches, we show the Chebyshev bounds both for the 25 samples that we use to compute the bound (the outer ellipsoid) and the asymptotic limit (the inner green ellipsoid). The black dot in the ellipsoid (pointed to by an arrow) is the least-cost sizing that lies on the upper-right quadrant of the bounding ellipsoid, and is the sizing recommended by that approach. The sizing obtained using SNC is a single point, not a Chebyshev bound, since it uses an envelope that is already robust to perturbations in the trace.

We see that sizing obtained using optimization is always much smaller than with the other two approaches. This is not surprising, given that the optimization approach chooses the optimal operating strategy, rather than the naïve strategy used by the other

[9]Source: https://www.tesla.com/powerwall

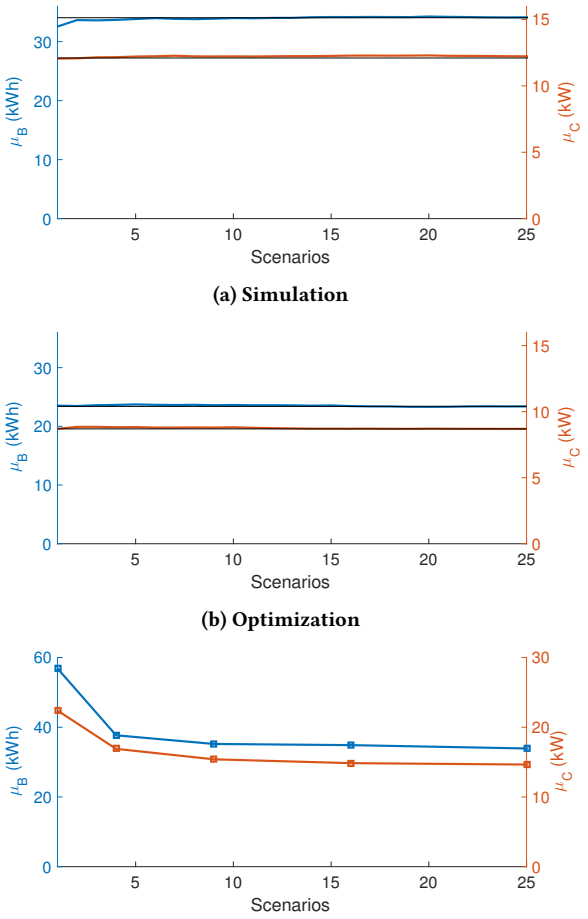

Figure 2: Convergence of mean B and C for different approaches.

approaches. It is also clear that as the LOLP target decreases, the least-cost sizing increases. Again, this is to be expected. Finally, the sizing using SNC is comparable, though with a higher value of C, than that obtained with simulations.

The gap between the two ellipsoids can be viewed as the gains from increasing the number of scenarios being evaluated. The gap is fairly small for high values of ϵ. The gap increases as ϵ decreases, indicating a higher variability in the sizing and hence a larger number of traces needed to find a better sizing.

6.3 Robustness

Although our approach to optimization-based sizing is robust, in that it is insensitive to small perturbations in the input trace, it is often not possible to use optimization-based sizing in real life, because it relies on optimal operation of the storage system, using a policy that cannot be determined in advance. Thus, we only evaluate robustness of our sizing for the simulation and SNC approaches. To do so, we first compute the B, C sizing using each of the approaches and using three years of traces, then test if the fourth year meets the

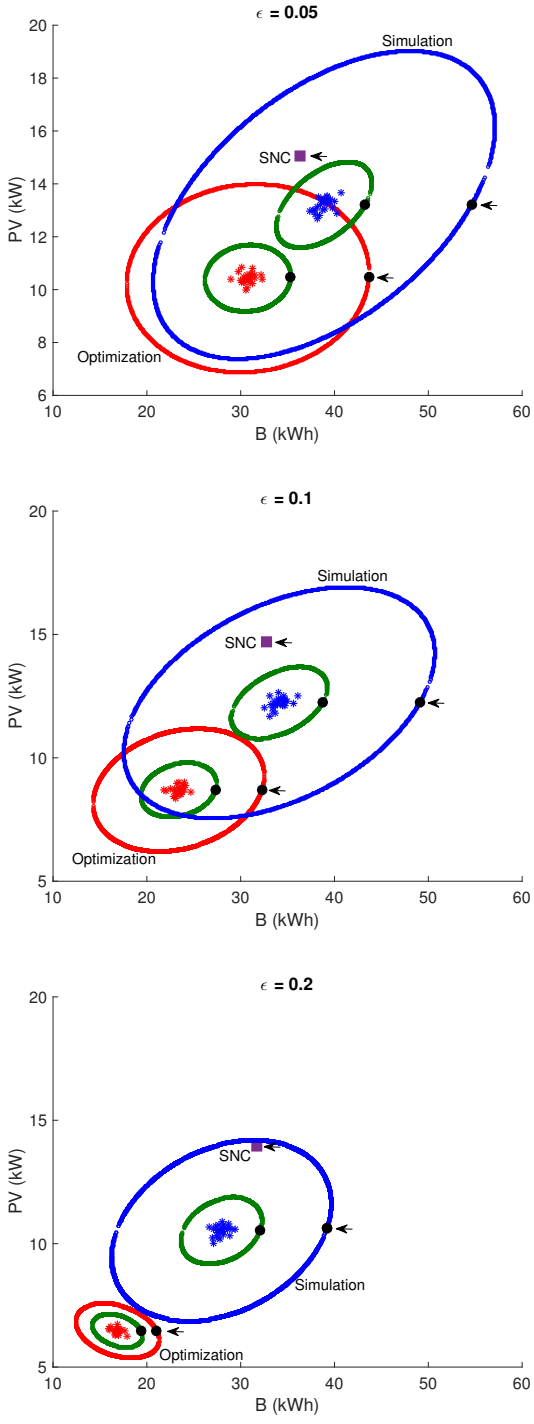

Figure 3: Comparing the three approaches for a single household for three values of ϵ. SNC results are shown as a purple square. Optimization and simulation results are shown as red and blue asterisks, respectively, and the ellipsoids represent sample (outer) and analytical (inner, green) Chebyshev 95% confidence bounds [31]. The black dot on each ellipsoid represents the lowest-cost system size in the upper-right ellipsoid quadrant.

Table 2: Mean computation time (Linux user time)

Method	Mean CPU time per scenario (m:s)
Simulation	1:50
Optimization	357:21
SNC	5:28

bound ('leave-one-out' analysis). Note that we do not use boosting in this evaluation, to be as realistic as possible. We summarize the results of this evaluation in Table 3.

Note that in all cases, the bound of $\epsilon = 0.1$ is met. However, the sizing is much higher than with boosting (compare these sizes with the arrows in Figure 3). This is because with fewer scenarios to evaluate, the Chebyshev bound as well as the SNC envelope, is quite loose. The sizing can be reduced by increasing the number of traces in the input trace set (by collecting more data).

Note also that each of the four sizings obtained using a subset of three years of data results in a substantially different sizing. Even though this sizing does not result in a violation of the ϵ bound, this indicates that solar generation and load is highly variable, and also that many pairs of B, C values are equally effective in meeting the desired bound.

Though we do not have space to show the results, we find that with our larger set of synthetic traces, and a similar leave-one-out-analysis, the sizing computed using SNC or simulations does not result in any violations of the ϵ bound, further indicating that these approaches are robust.

6.4 Computation time

Recall that the asymptotic complexity of the optimization approach is lower bounded by $O(mnT^{5.5}logT)$. The computational complexity of simulation is $O(mnbcT)$ and of SNC is $O(smnbcT)$. Thus, for large values of T, which is typical, the best approach is simulation.

To determine the impact of asymptotic complexity on typical computation times, in Table 2 shows the mean computation time per scenario using the three approaches on a 2.7GHz Intel Xeon CPU. It is clear that simulations take the least time, with optimization taking two orders of magnitude more.

7 DISCUSSION

7.1 Comparison of the three approaches

Our work evaluates three distinct approaches to robust and practical sizing of solar and storage systems. Over and above the numerical comparison in Section 6, we now make some qualitative observations about their relative merits.

Unlike some prior work [14, 16, 21] which solve the joint problem of optimal sizing and optimal operation, in this work, we study only sizing. However, with optimization, the operation rules are a free variable, in that the output of the optimization program is also the optimal charge/discharge schedule. Note that these operation rules cannot be used in practice, because the rules depend in detail on the traces, and the future is unknown. Moreover, the sizing obtained from optimization may be unrealistically small. If we could encode operation rules into the optimization program, we would be able to come up with a sizing that did not have this coupling. Unfortunately, it is non-trivial, perhaps impossible, to

Table 3: Leave-one-out analysis

Leave out:	Year 1			Year 2			Year 3			Year 4		
	B	C	ϵ	B	C	ϵ	B	C	ϵ	B	C	ϵ
SNC	44.6	23	0.021	48.8	21.1	0.017	45.136	22	0.024	45.8	21.2	0.033
Sim.	66.1	14	0.041	75.8	13.9	0.051	41	19.7	0.053	96	13.8	0.042

encode arbitrary operation rules in an optimization program. For instance, consider the operation rule "Charge the store from the grid if the SoC is below 0.3 and the grid price is lower than \$0.12/kWh." This couples P_c to the SoC, not just complicating the program but also making it non-linear, making it difficult to solve efficiently.

There is a similar problem with stochastic network calculus, where encoding complex charge/discharge operations rules into Equations (18) and (19) may result in greatly complicating the subsequent analysis. In contrast, the simulation approach is not only the one with the least time complexity ($O(mnbc)$ vs. $O(smnbc)$ and $O(mnT^{5.5})$), and also the least running time per scenario, it can be used with any operating strategy. Thus, from a qualitative perspective, the simulation approach is perhaps the best one, especially when combined with a *post hoc* Chebyshev bound.

7.2 Contributions of our work

Our work makes multiple contributions. To begin with, it is the first work, to our knowledge, that provides robust and practical advice on sizing by comparing multiple approaches.

Second, our use of a multivariate Chebyshev bound in combination with optimal sizing for multiple trace pairs is innovative, and can be generalized to other robust optimization problems.

Third, the LOLP formulation using SNC in Section 5.3 considerably advances the state-of-the-art in SNC analysis of battery-equipped systems, such as in References [17, 18]. This is because the battery model we use in this work is more realistic (and more complicated). Additionally, we take a different approach in characterizing energy profiles, by modelling net energy demand directly instead of modelling supply and demand separately. Finally, we use a bi-variate sample path envelope in this work instead a uni-variate function as used in Reference [17]. While this may seem like a small change, a uni-variate sample path identifies any time interval only by the length of the interval, while a bi-variate function indexes each time interval both by its beginning and ending point. For stochastic processes with substantial seasonality, such as solar power, a bi-variate function is a more accurate choice, as it captures properties that vary over time intervals that have the same size, but occur at a different time of the day or in different seasons.

7.3 Limitations and future work

Our work suffers from some limitations, as discussed next.

First, we did not consider unmet load as an optimization criterion. It is straightforward to characterize the unmet load using simulations. Moreover, the optimization problem for unmet load is a linear program, that is easily obtained as follows: Let δ be the fraction that we can tolerate, and constrain $\gamma(t)$ by

$$\sum_{t=1}^{T} \gamma(t) \leq \sum_{t=1}^{T} D(t)\delta. \tag{45}$$

This constraint, when added to Equations (1)-(8), results in the computation of the optimal sizing assuming optimal operation, with the goal of minimizing the unmet load. However, using stochastic network calculus to compute a sizing under the constraint that the probability of unmet load is bounded is quite complex. Hence, we defer it to future work.

Second, we have assumed that the load is not under our control. In some cases, it is possible to ask the energy consumers to modify their behaviour, using a control signal. Thus, for example, a home owner may be asked to defer a heavy load if the state of charge of the storage was particularly low. In this situation, it is obvious that the system sizing can be much smaller. However, sizing a system in the presence of load control is a much more complex problem, in that it requires jointly optimizing the storage operation as well as the load control actions. We intend to explore this in future work.

Third, the computation times presented in this paper are only indicative. For example, both simulations and the stochastic network calculus algorithm can be tuned, or re-coded in a more efficient high-level language to improve computation times. Similarly, it is well known that choice of optimization meta-parameters can also significantly impact the computation time. Nevertheless, given the substantial differences in performance, we believe our results are representative of expected outcomes in practical scenarios.

Fourth, our approach to generating data traces can be improved. For example, trends in historical solar irradiation can be found in NREL's NSRDB [15], and historical weather data on related variables such as cloud cover could be integrated to refine the generated values. A model-free approach such as a Generative Adversarial Network [11] is another alternative to data generation. Refining the generation of future load and solar power traces will improve the accuracy of our two-step sizing approach.

Finally, due to reasons of space, we have presented results only from a single home and from a single location. We will present more extensive evaluations in an extended version of this paper.

8 CONCLUSION

We evaluate and compare three state-of-the-art approaches to size solar generation and storage in a realistic setting. Unlike prior work, which evaluates a single approach and does not evaluate the robustness of the resulting solution, we compare the sizing results from these approaches on identical inputs, permitting a fair comparison. We find that, due to both qualitative and quantitative reasons, simulation appears to be the best tool for sizing in a realistic setting. In carrying out our work, we have made contributions to the state of the art both in the area of stochastic network calculus and in the use of multivariate Chebyshev bounds to obtain a novel technique for robust optimization and simulation.

REFERENCES

[1] 2018. Pecan Street Inc. Dataport. (2018).

[2] Andreas Aichhorn, Michael Greenleaf, H Li, and J Zheng. 2012. A cost effective battery sizing strategy based on a detailed battery lifetime model and an economic energy management strategy. In *Power and Energy Society General Meeting, 2012 IEEE*. IEEE, 1–8.

[3] Omid Ardakanian, Catherine Rosenberg, and S. Keshav. 2012. On the Impact of Storage in Residential Power Distribution Systems. *SIGMETRICS Perform. Eval. Rev.* 40, 3 (Jan. 2012), 43–47.

[4] Robert Basmadjian, Yashar Ghiassi-Farrokhfal, and Arun Arvishwa Keshav. 2018. Hidden Storage in Data Centers: Gaining Flexibility Through Cooling Systems. In *19th International GI/ITG Conference on "Measurement, Modelling and Evaluation of Computing Systems*.

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

[6] Dimitris Bertsimas, David B Brown, and Constantine Caramanis. 2011. Theory and applications of robust optimization. *SIAM review* 53, 3 (2011), 464–501.

[7] Arnab Bhattacharya, Jeffrey Kharoufeh, and Bo Zeng. 2016. Managing energy storage in microgrids: A multistage stochastic programming approach. *IEEE Transactions on Smart Grid* (2016).

[8] Ted KA Brekken, Alex Yokochi, Annette Von Jouanne, Zuan Z Yen, Hannes Max Hapke, and Douglas A Halamay. 2011. Optimal energy storage sizing and control for wind power applications. *IEEE Transactions on Sustainable Energy* 2, 1 (2011), 69–77.

[9] Edgardo D Castronuovo and Joao A Pecas Lopes. 2004. Optimal operation and hydro storage sizing of a wind–hydro power plant. *International Journal of Electrical Power & Energy Systems* 26, 10 (2004), 771–778.

[10] SX Chen, Hoay Beng Gooi, and MingQiang Wang. 2012. Sizing of energy storage for microgrids. *IEEE Transactions on Smart Grid* 3, 1 (2012), 142–151.

[11] Yize Chen, Yishen Wang, Daniel S Kirschen, and Baosen Zhang. 2018. Model-free renewable scenario generation using generative adversarial networks. *IEEE Transactions on Power Systems* (2018).

[12] Jiaojiao Dong, Feng Gao, Xiaohong Guan, Qiaozhu Zhai, and Jiang Wu. 2017. Storage Sizing With Peak-Shaving Policy for Wind Farm Based on Cyclic Markov Chain Model. *IEEE Transactions on Sustainable Energy* 8, 3 (2017), 978–989.

[13] EEMB 2016. *LIR18650 cell*. EEMB. Li-NMC cell specifications.

[14] Samira Fazlollahi, Pierre Mandel, Gwenaelle Becker, and Francois Maréchal. 2012. Methods for multi-objective investment and operating optimization of complex energy systems. *Energy* 45, 1 (2012), 12–22.

[15] Ray George, Steve Wilcox, Mary Anderberg, and Richard Perez. 2007. *National Solar Radiation Database (NSRDB)–10 Km Gridded Hourly Solar Database*. Technical Report. National Renewable Energy Laboratory (NREL), Golden, CO.

[16] Yashar Ghiassi-Farrokhfal, Fiodar Kazhamiaka, Catherine Rosenberg, and Srinivasan Keshav. 2015. Optimal design of solar PV farms with storage. *IEEE Transactions on Sustainable Energy* 6, 4 (2015), 1586–1593.

[17] Yashar Ghiassi-Farrokhfal, Srinivasan Keshav, and Catherine Rosenberg. 2015. Toward a realistic performance analysis of storage systems in smart grids. *IEEE Transactions on Smart Grid* 6, 1 (2015), 402–410.

[18] Yashar Ghiassi-Farrokhfal, Srinivasan Keshav, Catherine Rosenberg, and Florin Ciucu. 2015. Solar power shaping: An analytical approach. *IEEE Transactions on Sustainable Energy* 6, 1 (2015), 162–170.

[19] Paul Gilman, Nate Blair, Mark Mehos, Craig Christensen, Steve Janzou, and Chris Cameron. 2008. Solar advisor model user guide for version 2.0. *National Renewable Energy Laboratory, Golden, CO, Technical Report No. NREL/TP-670-43704* (2008).

[20] Yuming Jiang and Yong Liu. 2008. *Stochastic network calculus*. Vol. 1. Springer.

[21] ND Kaushika, Nalin K Gautam, and Kshitiz Kaushik. 2005. Simulation model for sizing of stand-alone solar PV system with interconnected array. *Solar Energy Materials and Solar Cells* 85, 4 (2005), 499–519.

[22] Fiodar Kazhamiaka, Catherine Rosenberg, and Srinivasan Keshav. 2016. Practical strategies for storage operation in energy systems: design and evaluation. *IEEE Transactions on Sustainable Energy* 7, 4 (2016), 1602–1610.

[23] Fiodar Kazhamiaka, Catherine Rosenberg, Srinivasan Keshav, and Karl-Heinz Pettinger. 2016. Li-ion storage models for energy system optimization: the accuracy-tractability tradeoff. In *Proceedings of the Seventh International Conference on Future Energy Systems*. ACM, 17.

[24] Kokam 2016. *Li-ion Polymer Cell*. Kokam. Li-NMC cell specifications.

[25] Magnus Korpaas, Arne T Holen, and Ragne Hildrum. 2003. Operation and sizing of energy storage for wind power plants in a market system. *International Journal of Electrical Power & Energy Systems* 25, 8 (2003), 599–606.

[26] Pratyush Kumar. 2016. A Network Calculus Foundation for Smart-grids Where Demand and Supply Vary in Space and Time. In *Proceedings of the Seventh International Conference on Future Energy Systems (e-Energy '16)*. ACM, New York, NY, USA, Article 4, 11 pages.

[27] Joydeep Mitra. 2010. Reliability-based sizing of backup storage. *IEEE Transactions on Power Systems* 25, 2 (2010), 1198–1199.

[28] John M Mulvey, Robert J Vanderbei, and Stavros A Zenios. 1995. Robust optimization of large-scale systems. *Operations research* 43, 2 (1995), 264–281.

[29] Majid Raeis, Almut Burchard, and Jörg Liebeherr. 2017. Analysis of the Leakage Queue: A Queuing Model for Energy Storage Systems with Self-discharge. *CoRR* abs/1710.09506 (2017). http://arxiv.org/abs/1710.09506

[30] S. Singla, Y. Ghiassi-Farrokhfal, and S. Keshav. 2014. Using Storage to Minimize Carbon Footprint of Diesel Generators for Unreliable Grids. *IEEE Transactions on Sustainable Energy* 5, 4 (Oct 2014), 1270–1277.

[31] Bartolomeo Stellato, Bart PG Van Parys, and Paul J Goulart. 2017. Multivariate Chebyshev inequality with estimated mean and variance. *The American Statistician* 71, 2 (2017), 123–127.

[32] Pravin M Vaidya. 1989. Speeding-up linear programming using fast matrix multiplication. In *Foundations of Computer Science, 1989., 30th Annual Symposium on*. IEEE, 332–337.

[33] K. Wang, F. Ciucu, C. Lin, and S. H. Low. 2012. A Stochastic Power Network Calculus for Integrating Renewable Energy Sources into the Power Grid. *IEEE Journal on Selected Areas in Communications* 30, 6 (July 2012), 1037–1048.

[34] Hongxing Yang, Wei Zhou, Lin Lu, and Zhaohong Fang. 2008. Optimal sizing method for stand-alone hybrid solar–wind system with LPSP technology by using genetic algorithm. *Solar energy* 82, 4 (2008), 354–367.