Solar Power Shaping: An Analytical Approach

Y. Ghiassi-Farrokhfal, *Member, IEEE,* S. Keshav, *Member, IEEE,* C. Rosenberg, *Fellow, IEEE,* and F. Ciucu, *Member, IEEE*

***Abstract*—The focus of our work is the use of an energy storage system (ESS) to integrate solar energy generators into the electrical grid. Although, in theory, an ESS allows intermittent solar power to be shaped to meet any desired load profile, in practice, parsimonious ESS dimensioning is challenging due to the stochastic nature of generation and load and the diversity and high cost of storage technologies. Existing methods for ESS sizing are based either on simulation or analysis, both of which have shortcomings. Simulation methods are computationally expensive and depend on the availability of extensive data traces. Existing analytical methods tend to be conservative, overestimating expensive storage requirements. Our key insight is that solar power fluctuations arise at a few distinct time scales. We separately model fluctuations in each time scale, which allows us to accurately estimate ESS performance and efficiently size an ESS. Numerical examples with real data traces show that our model and analysis are tight.**

I. Introduction

Traditionally, energy generators are finely controlled to match the fluctuations in aggregate demand. Unfortunately, due to their intrinsic stochastic nature, solar energy generators cannot be controlled in this way, making it difficult to integrate them into the grid. Specifically, solar fluctuations can harm power quality, increase the need for regulation, and complicate load following and unit commitment [4]. Hence, these fluctuations must be mitigated [9], [10], [21], [14].

Several approaches to integrating stochastic energy generators have been proposed in prior work [8]: geographical diversity, complementary energy sources (e.g., wind and solar), demand response, oversizing the capacity of the renewable energy sources, forecasting generator variations, and using energy storage systems (ESS) [22]-Ch.12. Using an ESS is interesting, in that an ESS provides flexibility to meet several of the integration challenges listed above [2]. Thus, the focus of our work is in ESS dimensioning to match stochastic supply with stochastic demands.

In theory, energy from solar generators can be stored in an ESS and withdrawn as necessary to match any desired demand with only a small probability of *loss of power*. In practice, however, this solar shaping is challenging due to the diverse physical constraints of storage technologies and their uniform high cost. For each ESS technology, we want to compute the minimum size which can meet demands with an acceptable risk of loss of power [6], [16], [23].

There are two existing approaches to size an ESS for solar energy generators.

Y. Ghiassi-Farrokhfal (yghiassi@uwaterloo.ca) and S. Keshav (keshav@uwaterloo.ca) are with the Department of Computer Science, University of Waterloo. C. Rosenberg (cath@ecemail.uwaterloo.ca) is with the Department of Electrical and Computer Engineering, University of Waterloo. F. Ciucu (florin@dcs.warwick.ac.uk) is with the Department of Computer Science, University of Warwick.

- **Simulations**: Given datasets of solar irradiance and demand profiles, it is possible to simulate an ESS of a particular size that is based on a particular storage technology to determine the probability of loss of power [24]. For a given ESS technology, this simulation must be repeated for each storage size until the minimum ESS size that meets the requirements is obtained. Although this method is widely used due to its simplicity and precision, it has two drawbacks. First, to obtain small values of loss of power probabilities, the simulation must incorporate sufficiently large datasets[1]. Second, simulations need to be re-run for each choice of parameter values, such as the storage size, which is computationally expensive and cumbersome.
- **Analytical methods**: They mathematically model a stochastic solar generator, an ESS, and stochastic demands to estimate the loss of power probability. The existing analytical methods [11], [26], [27], [28] are more efficient than repeated simulations, but have two drawbacks. First, they are based on strong assumptions, throwing into question the validity of the obtained results. Second, they are often too conservative, resulting in oversizing of the ESS.

What is desirable, therefore, is an analytical model that makes few assumptions and closely approximates the results from simulations, thus providing the best of both worlds.

Our key insight is that an analytical model for an ESS used for solar power shaping must take into account the three intrinsic time scales over which solar power fluctuates. First, at the time scale of a day, solar power varies due to the position of the sun in the sky. Second, long-term cloudiness causes power fluctuations at time scales ranging from a few hours to about 10 minutes. Third, there is a high-frequency power modulation due to clouds at time scales faster than about 10 minutes. This is demonstrated vividly by the power spectral density of solar power (Fig. 3) [3] [12]. Therefore, we model the impact of clouds by multiple stochastic processes at different time scales.

We have used this analytical model, along with a unified analytical ESS model from our prior work [11], to accurately estimate loss of power performance and optimally size an ESS. Our key contributions are:

1) We provide a new analytical model for solar power shaping which characterizes both the short-term and long-term variations in daily solar power.
2) Given any feasible pair of target output power shape and allowable loss of power threshold, we provide techniques to compute a near-optimal ESS size using our unified

[1]For instance, if the target loss of power probability is 1 day in 10 years, the datasets must be for a period of at least 10 years, if not longer.

Name	*Description*
$I_g(t)$	The actual solar insolation at time t (W/m^2)
$I_{CS}(t)$	The clear sky insolation at time t (W/m^2)
α_{pv}	Photovoltaic coefficient
$D(t)$	ESS output power at time t (W)
$D^*(t)$	ESS target output power at time t (W)
$S(t)$	Solar power at time t (W)
$S_T(t)$	Solar power aggregate in time intervals of size T (W)
$G_T(t)$	Ensemble average of $S_T(t)$ (W)
$\mathcal{G}$	Statistical sample path envelope on S (W)
$\mathcal{G}_d$	Dynamic sample path upper envelope on S (W)
CSI(t)	Clear sky index at time t ($\frac{I_g(t)}{I_{CS}(t)}$)
B	Storage size (Wh)
$\alpha_c(\alpha_d)$	Storage charging (discharging) power limit (W)
η	Storage efficiency
DoD	Storage depth of discharge
ε^*	The allowable risk of loss of power

TABLE I: Notations

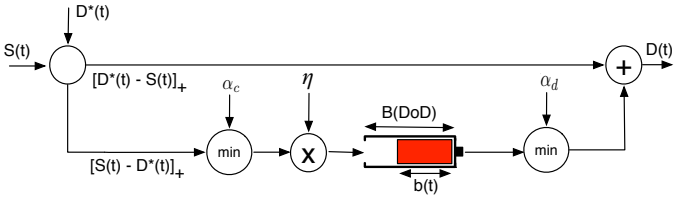


Fig. 1: A renewable energy source S equipped with a storage to provide a target output power D^*. The storage physical constraints are α_c, α_d, η, and DoD. The actual output power is D.

ESS model.

3) Using real datasets, we show that the results from our analytical model reasonably match those from a simulation-based approach where the complete dataset is known ahead of time (we call it offline optimal) and considerably outperform prior analytical models.

The rest of the paper is organized as follows. In Section II, we discuss our system model and define the problem. We review the existing solar power models which can be used for solar shaping in Section III. We describe our solar power model in Section IV. We then formulate and solve the solar power shaping in Section V. We evaluate our approach in Section VI, and conclude the paper in Section VII.

II. Problem Definition

The stochastic process under study here is *solar power*, which is a fluid-flow (i.e., can take any value) process. We assume a discrete-time model, where time is slotted $t = 0, T_u, 2T_u, \ldots$, with T_u being the time unit. To simplify notation, we drop T_u from our formulation by assuming $T_u = 1$. Generalizing the formulas for any T_u is a matter of additional notations. The goal is to shape the fluctuating solar power to a target output power using a storage system. Denote by $S(t)$ and $D^*(t)$, respectively, the available solar power and the system target output power at time t. To simplify notation, we write $D^*(s,t)$ and $S(s,t)$ to, respectively, mean $\sum_{\tau=s+1}^{t} D^*(\tau)$ and $\sum_{\tau=s+1}^{t} S(\tau)$ (e.g., $S(t-1,t) = S(t)$).

We illustrate our energy storage system (ESS) model in Fig. 1 (Please see [11] for the details). Solar power is used to serve the target output power directly as much as possible (i.e., it departs the system without going through the storage). If, in a given time slot t, the available solar power is insufficient (i.e., $S(t) < D^*(t)$), the energy stored, if any, can be used to make up the difference. Moreover, if the available solar power in a time slot t is larger than the target output power (i.e., $S(t) > D^*(t)$), then the surplus energy ($(S(t) - D^*(t))T_u$) is stored in the storage, if it is not yet full. All incoming power exceeding storage's charging rate limit α_c is dropped. The discharging rate is α_d. Moreover, the storage loses a fraction of $1-\eta$ of the total energy being stored in the storage due to storage inefficiency. Finally, the storage lifetime constraint is met if only a DoD fraction ≤ 1 of the entire storage is used[2].

Let the actual output power from ESS at any time t be $D(t)$. Then, $D(t) \leq D^*(t)$ due to the ESS limitations such as inefficiency, leakage, and finite size. An important research problem is to find the minimum storage size and the best storage technology to guarantee the target output power at any time t (i.e., $D(t) = D^*(t)$) with an allowable risk ε^*. In other words, at any time t, we must have

$$\Pr\{D(t) < D^*(t)\} \leq \varepsilon^*. \quad (1)$$

There are three ways to reach this goal:

- **Direct simulation**: Given a large enough solar power measurement trace, one can simulate the charging/discharging process in the ESS system in Fig. 1 for each storage technology and iterate on the storage size to find the minimum size which satisfies Eq. (1).
  Direct simulation is the most accurate method, but has major weaknesses. First, very large datasets are required for small values of ε^*. Second, the complexity of this method is cumbersome given that for each storage technology and size, we need to repeat the simulation for the entire large data trace.
- **Simulation with a generated data-trace**: If the available dataset is not large enough, a large data trace can be generated by simulating the statistical properties of the given data trace. Then, the rest is identical to the direct simulation method. This method is not as accurate as the direct simulation due to possible inaccuracy of the generated data trace.
- **Analysis**: This method computes an analytical upper bound on $\Pr\{D(t) < D^*(t)\}$. This upper bound is a function of the statistical properties of the solar power, the type, and the size of storage. This method does not suffer from the shortcomings of the direct simulation method. However, it might lead to oversizing if the model and formulations are not tight.

In this paper, we assume that the given measurement dataset of solar power is not large enough to use direct simulation for the target loss of power probabilities. Hence, we must use either of the two other methods (listed above) to model solar power based on the limited measurement set. We call

[2]The energy stored in the storage decreases due to self-discharge. However, for battery storage systems this value can be safely neglected.

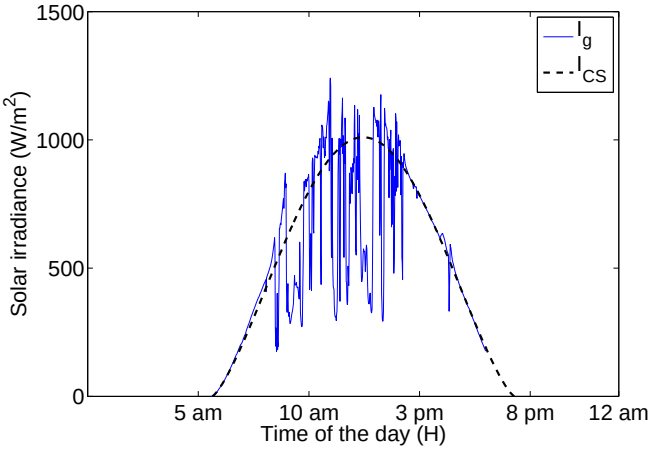


Fig. 2: Illustrating the global and the clear sky irradiance for a sample day.

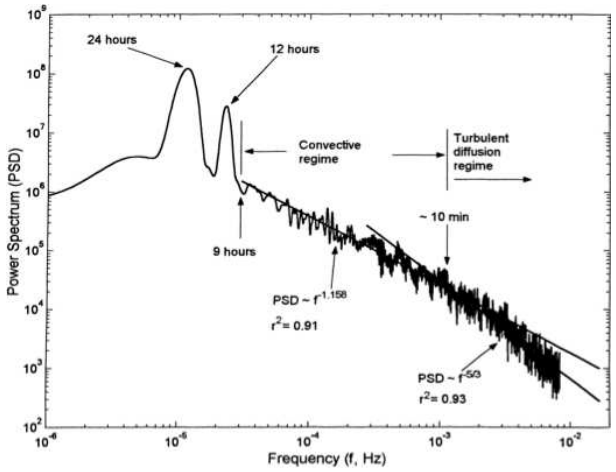


Fig. 3: Power spectrum of the 1min values of the global solar irradiance [12].

solar power models which are used to generate traces (second method) as *feature models* and those to compute performance bounds (third method) as *statistical models*. We discuss the existing models of each category in the following section.

III. Background and Related Work on Solar Power Modelling

The *clear sky irradiance*, denoted I_{CS}, is the amount of power received from the sun per square meter (W/m^2) in the absence of clouds, shadows, and atmospheric particulates. It is easy to mathematically model this value at any point on the surface of the globe for a given time of day, day of year, and surface tilt angle [5]. In contrast, the *global* irradiance, denoted I_g, is the amount of power that is actually received by a photovoltaic (PV) panel. PV output power is almost linearly proportional to the global irradiance, so that solar power S generated by a panel of unit size at any time t is given by

$$S(t) = \alpha_{pv} \times I_g(t) \ , \tag{2}$$

where α_{pv} is the efficiency of the panel.

The ratio of the global irradiance to the clear sky irradiance (typically, but not always, smaller than 1) is called the *Clear Sky Index* or *CSI*.

A. Frequency-domain analysis

Solar power fluctuations arise from three stochastic processes, each operating in a different time scale [12] as shown in Fig. 3:

1) **Short time scale variations:** This process models fluctuations in irradiance when the direct solar beam is blocked by clouds. This typically results in attenuation that changes rapidly over time, but, in some cases, can actually lead to an enhancement of solar power [19]. It has been found that solar power fluctuations due to inhomogeneous small clouds happen on a time scale shorter than 10 minutes. This corresponds to the rightmost linear section in Fig. 3.
2) **Large time-scale variations:** Long-term cloudiness that occurs for periods between 10 minutes and 9 hours are modeled by this stochastic process. Long-term solar power fluctuations have substantially different statistical properties than the short-term ones because larger time intervals reflect the aggregate impacts of all attenuations and enhancements [25], [19]. The larger the size of the cloud, the larger the size of the time interval. This corresponds to the middle linear section in Fig. 3.
3) **Diurnal:** This stochastic process corresponds to the 24-hour and 12-hour time scales and is due to the daily transit of the sun in the sky, with the morning and evening solar power being roughly equal[3]. This corresponds to the two sharp peaks in Fig. 3.

This discussion indicates that an accurate solar power model must separately characterize diurnal, short-term, and long-term variations. We now present the existing models.

B. Existing models for solar power

We categorize the existing models into *feature models* and *statistical models*. Feature models extract certain features of solar power production from a dataset. These traces can then be used for performance analysis through simulation. Statistical models, on the other hand, are not used to generate traces. They only extract some statistical characteristics from the dataset and use those characteristics to formulate performance metrics. In the following, we review some of the most-widely used models of each category.

1) Clear Sky Index-based model (feature model): It has been shown that at large time scales (larger than hourly), the global irradiance can be modelled accurately [15] by separately modelling clear sky irradiance (I_{CS}) [5], [13] and the clear sky index (CSI) (the ratio of the global irradiance to the clear sky irradiance) [17], [20][4]. Clear sky index is

[3]The second peak corresponds to the length of a typical day, which is 12 hours for the location studied in this graph.

[4]A similar, older model characterizes the *clearness number* (instead of CSI), which is the ratio between the global irradiance and the extraterrestrial irradiance (the irradiance measured outside the atmosphere by satellites).

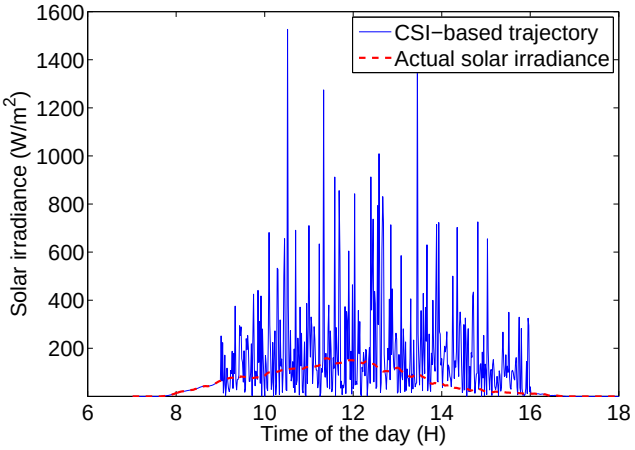


Fig. 4: Comparing the actual solar irradiance with the CSI-based generated trajectory in a cloudy winter day.

characterized under the assumptions that it is a Lévy process (stationary and independent increment with a continuous probability) with bi-modal distribution at any time instant [19], [25]. Then, the global irradiance at any given time is given by

$$I_g(t) = CSI \times I_{CS}(t). \tag{3}$$

Solar power modelling for smart grid applications deals with higher time resolutions than hourly (e.g., every minute). Unfortunately, CSI-based models are not accurate at these time scales due to the assumption of CSI being a Lévy process, which becomes less accurate in smaller time scales. Fig. 4 shows how erroneous a trajectory generated based on CSI method can be, compared to the actual solar irradiance data trace.

2) General envelope-based model (an analytical model): In this model, the important statistical properties of solar power which impact the charging/discharging processes are represented by envelopes. To see the nature of the envelopes required, consider an ideal ESS with infinite storage size, solar power S, and target output power D^*. Using the mapping between the state of charge (SoC) in such a system and the buffer content in packet switch networks, we have from the well-known Reich's equation:

$$SoC(t) = \max_{0\le s\le t} \left(S(s,t) - D^*(s,t)\right). \tag{4}$$

From Eq. (4), if we know a lower bound on $\max_{0\le s\le t} S(s,t)$, then we can compute a lower bound (envelope) on the SoC. Several recent papers [26], [27], [28], use the well-known *statistical sample path envelopes* [18] for this matter; solar power S is characterized by a statistical sample path lower envelope $\mathcal{G}$ and a bounding function ε such that at any time $t \ge 0$ and for any σ

$$\Pr\Big\{\max_{s\le t} \left(\mathcal{G}(t-s) - S(s,t)\right) > \sigma\Big\} \le \varepsilon(\sigma). \tag{5}$$

In this model, $\mathcal{G}$ characterizes the underlying deterministic behaviour of the cumulative solar power and ε characterizes the stochastic variations (i.e., the likelihood that cumulative solar power is less than $\mathcal{G}$).

This model suffers from the following shortcomings:

1) Not accounting for the diurnal effect: Unlike the CSI-based model, this model cannot suitably accommodate the diurnal effect as described above. Due to the fact that $\mathcal{G}$ is a uni-variate function, this model can be a good candidate *only* for a stochastic process which modulates a function which is only dependent on the size of the time interval (and not the position), or equivalently, a constant rate r at any time instant, i.e., $\mathcal{G}(t) = rt$. Therefore, this model is not a good candidate for solar power as the deterministic diurnal effect of solar power is a function of both the length and the position of each time interval.

2) No separation between the short-term and long-term variations (the knee point at 10min in Fig. 3): The general envelope model characterizes both the long-term and the short-term variations, simultaneously, through ε and for this reason cannot be precise.

IV. A NEW ANALYTICAL MODEL FOR SOLAR POWER SHAPING

In this section, we propose a new envelope model. This model adapts the general envelope model to enable a separate characterization of the three underlying processes of solar power (diurnal, long-term, and short-term variations).

To account for the diurnal effect, we replace the uni-variate sample path envelope function in the general envelope model by a bivariate envelope called the *dynamic sample path envelope*. A bi-variate function $\mathcal{G}_d(s,t)$ is a dynamic sample path envelope on solar power S, if it satisfies

$$\Pr\Big\{\max_{s\le t} \left(\mathcal{G}_d(s,t) - S(s,t)\right) > \sigma\Big\} \le \varepsilon(\sigma) \tag{6}$$

at any time t and for any $\sigma \ge 0$.

As observed in Fig. 3, the statistical properties of short time scale (1min-10min) is different from large time scale (10min-9h). Hence, we separate the two regimes: short time scale with time unit of 1 and large time scale with time unit $T > 1$. In large time scale regime, solar power S_T is the average of S in time slots of size T (see Fig. 5a), i.e., for any integer j

$$S_T(t) = \frac{\sum_{\tau=jT+1}^{(j+1)T} S(\tau)}{T} \qquad \forall t: \quad jT < t \le (j+1)T. \tag{7}$$

The variations in S_T reflect the large time scale variations (10min-9h in Fig. 3) and the variations in $S - S_T$ reflect the variations in short-time scale (1min-10min in Fig. 3).

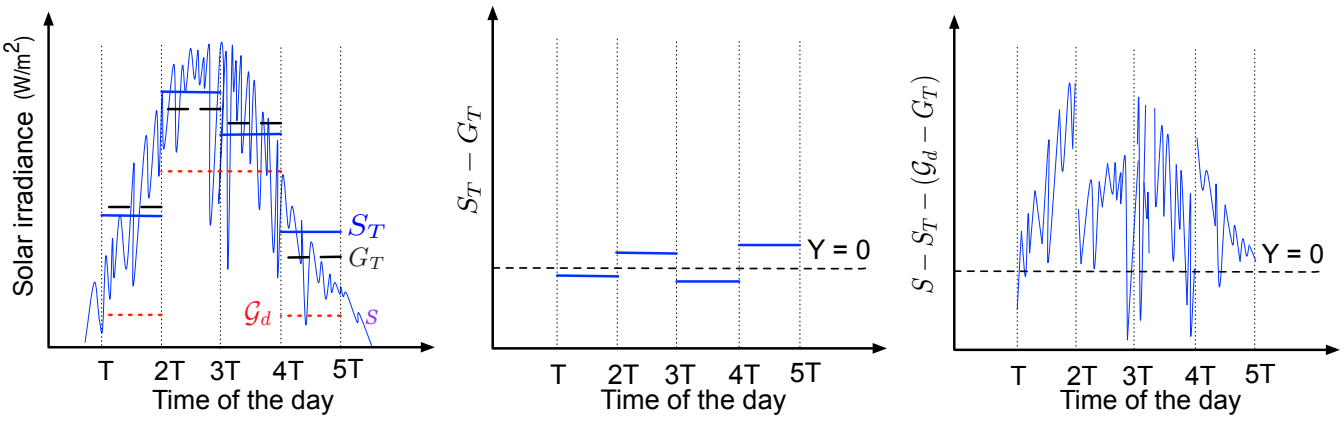
We characterize the variations in the large time scale by a dynamic envelope $G_T(s,t)$ with bounding function[5] ε_T, satisfying

$$\Pr\Big\{\max_{s\le t} \left(G_T(s,t) - S_T(s,t)\right) > x\Big\} \le \varepsilon_T(x). \tag{8}$$

for any $x \ge 0$. In addition, suppose that $\mathcal{G}_d(s,t) - G_T(s,t)$ is the lower dynamic envelope which characterizes the short-term variations with bounding function ε_1. Thus, for any $x \ge 0$

$$\Pr\Big\{\max_{s\le t} \left(\left(\mathcal{G}_d(s,t) - G_T(s,t)\right) - \left(S(s,t) - S_T(s,t)\right)\right) > x\Big\} \le \varepsilon_1(x). \tag{9}$$

[5] G_T is defined such that it can vary only at time instants $t = jT$ for any $j = 0, 1, \ldots$.



(a) Terminology (b) T-minute solar power modelling (c) 1-minute solar power modelling

Fig. 5: **Our solar power model:** Our terminology and the two-stage break-down in our model.

Then, we can obtain a bounding function corresponding to the dynamic envelope $\mathcal{G}_d$ on S as follows:

$$\begin{aligned}
&\Pr\Big\{\max_{s\le t}\left(\mathcal{G}_d(s,t)-S(s,t)\right)>\sigma\Big\}\\
&\le \inf_{x_1+x_T=\sigma}\Big(\Pr\Big\{\max_{s\le t}(G_T(s,t)-S_T(s,t))>x_T\Big\}\\
&\qquad+\Pr\Big\{\max_{s\le t}((\mathcal{G}_d(s,t)-G_T(s,t))\\
&\qquad\qquad-(S(s,t)-S_T(s,t)))>x_1\Big\}\Big), \quad (10)\\
&\le \underbrace{\inf_{x_1+x_T=\sigma}(\varepsilon_T(x_T)+\varepsilon_1(x_1))}_{:=\varepsilon(\sigma)}, \quad (11)
\end{aligned}$$

where in the second line, we use the fact that $\Pr(A+B>a+b)\le\Pr(A>a)+\Pr(B>b)$ for any random variables A and B and any constants a and b. In the last line, we use Eqs. (8)-(9).

The first term in Eq. (10) is in the large time scale regime as both S_T and G_T can vary only at time instants $t=jT$ for any $j=0,1,\ldots$. The second term in Eq. (10) is in the short time scale regime. Thus, Eq. (10) breaks the model into long-term and short-term variations.

In summary, we model solar power by a dynamic envelope $\mathcal{G}_d$ and a bounding function ε in terms of Eq. (6). The strength of our analytical model is that, unlike previous models, it can separately characterize the three underlying time-scales of solar power by:

1) **Accounting for the diurnal effect:** using a bi-variate envelope function $\mathcal{G}_d$.
2) **Separate characterization of the short-term and long-term variations**: using two time scales in describing the bounding function ε as in Eq. (10).

There are two free parameters in our solar power model: T and G_T. Proper choices of these parameters lead to an accurate solar power characterization.

A. How to choose the free parameters?

Minimizing the modelling error over T and G_T is a non-convex problem. However, here are some good choices:

• **The choice of G_T:** G_T represents the deterministic trend (diurnal effect) of solar power. Two good candidates are $G_T=I_{CS}$ and G_T being the historical average of S_T. The latter one is a great choice when a large dataset is available so that the historical average can be estimated precisely. If large datasets are not available, then the former one is a great choice as I_{CS} can be estimated quite precisely using the existing models.

• **The choice of T:** The right choice of T is crucial as it determines the boundary between the small and large time scales in our model. Indeed, the optimal value of T must coincide with the knee point in Fig. 3 (around 10min) and depends on the location, PV panel sizes, and the cloud type/speed in that location.

B. How to characterize the bounding function ε

For a given solar power dataset S, an envelope $\mathcal{G}_d$, and for fixed choices of G_T and T, we want to characterize a bounding function ε satisfying Eq. (6). We do so for a specific dataset, i.e., the U.S. Department of Energy's Atmospheric Radiation Measurement (ARM) [1] dataset, which is freely available. We characterize ε_T, ε_1, and eventually ε.

The bounding function ε_T is characterized as follows: We want to construct a dynamic sample path envelope on solar power in each day. Hence, solar power in each day in our dataset can be considered as a trajectory. Construct a set Y with elements $Y^{i,t}$ chosen at time $t\ge 0$ corresponding to a trajectory i such that

$$Y^{i,t}=\max_{0\le s\le t}(G_T(t-s)-S_T^i(s,t)),$$

where we use superindex i to refer to the trajectory i. From Eq. (8), ε_T can be chosen to be the Complementary Cumulative Distribution Function (CCDF) of any distribution that fits Y. We found (through QQ-plots which are not shown here due to lack of space) that fitting a distribution of the form $f_Y(x)=(1-p_T^0)\delta_0(x)+p_T^0\alpha_T e^{-\alpha_T x}$ is a good choice[6]; p_T^0 is the ratio of the non-zero elements in Y and α_T is the parameter

[6] $\delta_0(x)=1$ if $x=0$ and $\delta_0(x)=0$ if $x\neq 0$.

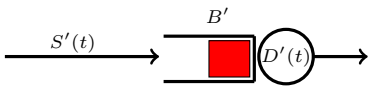


Fig. 6: Converting a non-ideal storage model to an ideal one by defining virtual source S', demand D', and storage size B'.

obtained when fitting an exponential distribution to the non-zero elements of Y.This means that $\varepsilon_T(x) = p_T^0 e^{-\alpha_T x}$ for any $x > 0$. Similar results have been obtained for ε_1, i.e.,

$$\varepsilon_T(x) = p_T^0 e^{-\alpha_T x}; \qquad \varepsilon_1(x) = p_1^0 e^{-\alpha_1 x}, \tag{12}$$

where p_1^0 and α_1 are, respectively, computed in a similar way as for p_T^0 and α_T, by replacing S_T by $S - S_T$ and G_T by $\mathcal{G}_d - G_T$ and repeating all the above steps.

Combining Eqs. (11)-(12), for any $\sigma \geq 0$, we have

$$\begin{aligned}
&\Pr\{\max_{s\leq t}(\mathcal{G}_d(s,t) - S(s,t)) > \sigma\} \\
&\quad \leq \inf_{x_1+x_T=\sigma} \left(p_T^0 e^{-\alpha_T x_T} + p_1^0 e^{-\alpha_1 x_1}\right) \qquad (13)\\
&\quad = \underbrace{(\alpha_T+\alpha_1)\left(\frac{p_1^0}{\alpha_T}\right)^{\frac{\alpha_T}{(\alpha_1+\alpha_T)}}\left(\frac{p_T^0}{\alpha_1}\right)^{\frac{\alpha_1}{(\alpha_1+\alpha_T)}} e^{-\frac{\alpha_1\alpha_T}{\alpha_1+\alpha_T}\sigma}}_{:=\varepsilon(\sigma)}, \qquad (14)
\end{aligned}$$

where we use Lemma 3 from [7] to obtain the last line.

V. Solar shaping analysis

In this section, we use our solar power model to compute the risk of power unavailability when shaping solar power to a target output function. We adopt the performance bound formulation from our earlier work in [11]. We have shown in that work that the system model in Fig. 1 is equivalent to the simple scenario depicted in Fig. 6, where S', D', and B' are called *virtual processes*, given by

$$\begin{aligned}
S'(t) &= S(t) - (1-\eta)[S(t) - D^*(t)]_+ \\
&\qquad - \eta[S(t) - D^*(t) - \alpha_c]_+, \qquad (15)\\
D'(t) &= D^*(t), \qquad (16)\\
B' &= B \times \text{DoD}, \qquad (17)
\end{aligned}$$

where we have safely assumed that $[D^*(t) - S(t) - \alpha_d]_+ = 0$ in Eq. (16), knowing that the storage discharge rate limits α_d is not typically a constraint in the existing technologies. This simple model leads to a loss of power formulation given in Theorem 1 in [11], which we use in the following corollary for solar power shaping (Please refer to the appendix for the proof):

Corollary 1 (Solar power shaping)**.** *Suppose that a solar PV panel, equipped with a storage of size B and parameters (α_c, α_d, η, DoD), is used to provide a target output power $D^*(t)$ at any time slot t (Fig. 1). Let the virtual power source S' and the virtual output target D' be as defined in Eqs. (15)-(16). Suppose that ε_s is the bounding function of the dynamic sample path envelope $\mathcal{G}_d(s,t) = D'(s,t)$ on S' for any s and t. Let ε_0 be a constant satisfying*

$$\Pr\{D'(t) > S'(t)\} \leq \varepsilon_0 \tag{18}$$

at any $t \geq 0$[7] then, this system can provide the target output D^* with the risk of failing ε^*, given by

$$\varepsilon^* = \min\left(\varepsilon_0, \varepsilon_s(B_0)\right), \tag{19}$$

where B_0 is the initial state of charge.

Remark: Corollary 1 can also be used for storage sizing: Given a target output power D^* and a maximum allowable violation probability $\bar{\varepsilon}$, we can use Corollary 1 to minimally size an ESS so that the target output power D^* can be guaranteed with a failure probability of less than $\bar{\varepsilon}$. To do so, we fix the storage size B (starting from $B = 0$) and for the given D^*, we compute the corresponding violation probability ε^* from Eq. (19). If $\varepsilon^* > \bar{\varepsilon}$, we increase the size of storage B with a small value[8] and repeat the same step until we find the minimum value of B for which $\varepsilon^* \leq \bar{\varepsilon}$. It is important to note that depending on the values of D^* and $\bar{\varepsilon}$, there might not exist any ESS size (even infinite) to guarantee $\varepsilon^* \leq \bar{\varepsilon}$. A trivial example could be when $D^*(s,t)$ is always larger than the average of the solar input power $S(s,t)$ in any time interval $[s,t]$. In this case, we say the pair (D^*, ε^*) is *infeasible*, meaning that even an ESS of infinite size is not helpful[9]

VI. Evaluation

In this section, we evaluate our model with respect to some existing models and also to the simulation results given the whole data trace in advance. We use the dataset from the ARM website [1] from $C1$ in SGP permanent site and for a large time interval of 10 years (from 2002 to 2011). We use 5 years of this dataset (2002-2006) to extract the statistical properties of solar power for all of the models including ours. Then, we use the next 5 years (2007-2011) of this dataset for evaluation.

We categorize the annual data into four seasons (spring, summer, fall, winter) and the days in each season into three classes based on their sky cover: *sunny, partly-cloudy, and cloudy*. As a result, each day in a year belongs to one of 12 *profiles* based on its season and sky cover.

We use the solar shaping scenario illustrated in Fig. 1 with a simple On-Off target output power: For each (season-sky cover) profile p, $D^*(t) = K_p$ in the middle of the day for an interval of size T_{on} hours (On period), where K_p is a constant, and $D^*(t) = 0$ for the rest of the day (Off period). We choose $\mathcal{G}_d(s,t) = D'(s,t)$ as suggested in Corollary 1. Moreover, we choose G_T to be the average power over the entire On period T_{on}. The rest of the parameters are computed as described in Sections IV, V.

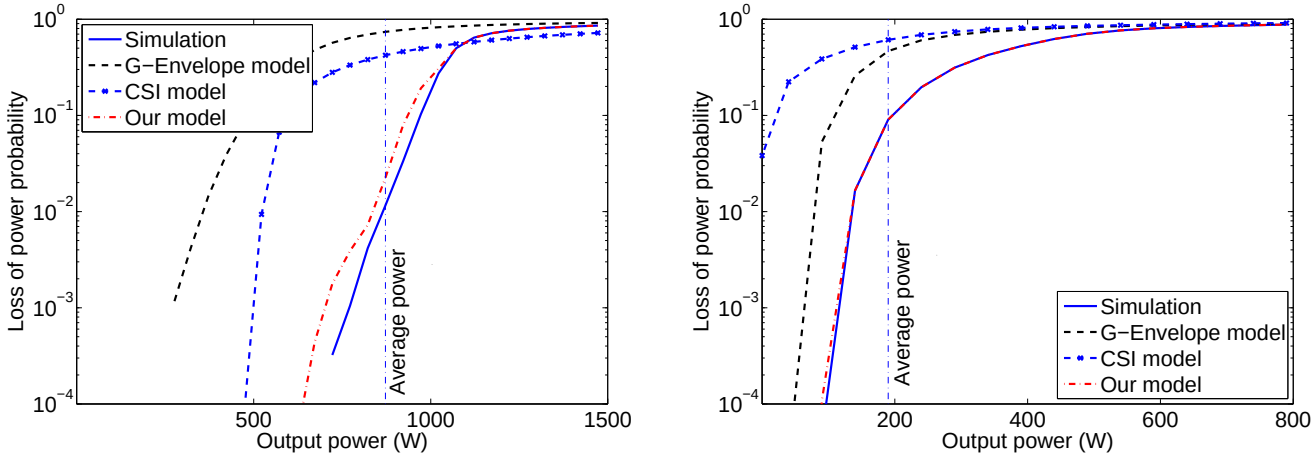
In all examples, we assume a solar PV panel of size $5m^2$ with PV efficiency $\alpha_{pv} = 0.20$. The storage technology is chosen to be a Lithium-ion battery of size 500KWh. From [11], we know that $\alpha_c(W) = B(Wh)/3(h)$, $\alpha_d(W) = 5\alpha_c(W)$,

[7]Note that this constant exists, because $\varepsilon_0 = 1$ always satisfies Eq. (18).

[8]From Eq. (19), we observe that ε^* is monotonically decreasing in B_0 and consequently in B (B_0 is non-decreasing in B).

[9]From Eq. (19), we observe that ε^* is monotonically increasing in D^*..

When we talk about storage sizing for a target output power D^* with an allowable violation probability $\bar{\varepsilon}$ in the rest of the paper, we implicitly assume that $(D^*, \bar{\varepsilon})$ is a feasible pair.



(a) Sunny summer days (b) Cloudy winter days

Fig. 7: **Evaluating solar power models:** Probability of loss of power as a function of output power in the interval of 10am-2pm.

$\eta = 0.85$, and DoD = 0.8 for Lithium-ion batteries. The state-of-charge at the beginning of each day is computed by emulating the battery state of charge in the previous days.

We include four curves in most of the plots:

1) **Simulation:** We use direct simulation assuming that we have the entire trace in advance. This provides *offline optimal* values for that given trace and holds as a benchmark to evaluate other methods. Indeed, with simulating the entire trace we can exactly compute the loss of power probability for any target output power and the minimum storage size for a given output power and a violation probability (for a given trace).
2) **CSI-based model:** We have access to both I_g and I_{CS} in our dataset; hence we can compute CSI at any time. We collect the values of CSI for the first half of our dataset and fit a hyper-exponential distribution to CSI measurements at any time instant. Assuming that CSI is a Lévy process, and using Eq. (3), we simulate I_g from Eq. (3) for the second half of our dataset. Then, using this generated trace, we compute the quantiles (for ε^*) using simulation on the generated trace.
3) **G-Envelope model:** Solar power is characterized by a sample path envelope $\mathcal{G}_d(t) = \min_{0\le s} D'(s, s+t)$ and a hyper-exponential distribution as the violation probability in the sense of Eq. (5). We then use the uni-variate special case of our analysis from Section V to compute an upper bound on ε^*.
4) **Our model:** As described in Sections IV, V.

A. The accuracy of our model

In this section, we evaluate the accuracy of our model with respect to the other models and to the simulation results. We compare the loss of power probability estimated using each of the models. In Fig. 7 we choose the On period to be 10am-2pm. This graph shows that our model outperforms the other existing ones significantly, and it achieves tight bounds, as it closely follows the offline optimal values (labelled simulation in the figure). The second interesting observation here is that both the CSI-based model and the general envelope-based model in a sunny summer day underestimate the available power as we consider the time interval 10am-2pm. This is the result of an inaccurate characterization of the diurnal effect in these two models. They cannot completely capture the fact that the solar irradiance is maximized in the time interval of 10am-2pm. In addition, the significant difference between Fig. 7a and Fig. 7b illustrates the existence of a strong seasonality in the solar power process and this in turn reveals the importance of having different profiles for each season and sky cover.

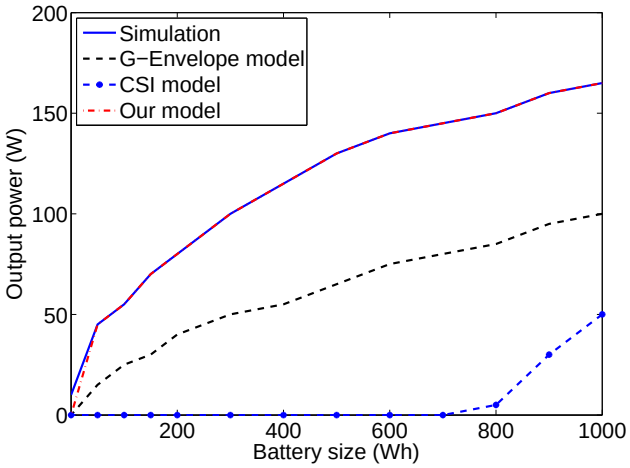


Fig. 8: **Storage sizing:** The target output power for a given storage size with probability larger than 99% and in a cloudy winter day.

B. Storage sizing

In this section, we study the accuracy of our analysis in terms of storage sizing in a cloudy winter day. Similar to the previous example, we assume that T_{on} is the time interval 10am-2pm. We fix the violation probability ε^* to be 10^{-2}. Then, we compute the minimum storage size which satisfies

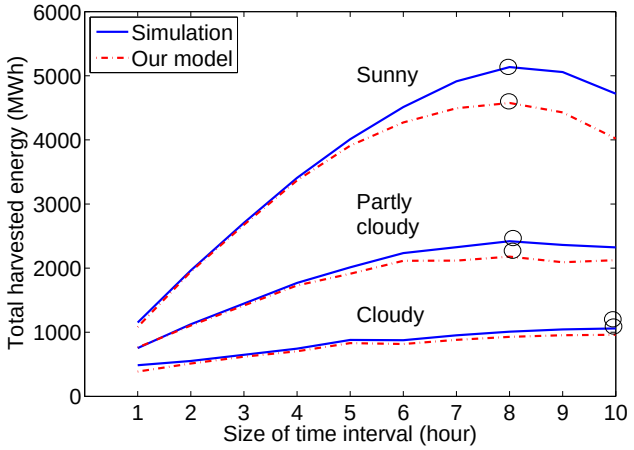


Fig. 9: **The optimal size of the time interval:** The guaranteed total energy harvested with probability larger than 99% as a function of total size of the time interval in a summer day. The small circles show the size of the time intervals for which the total harvesting energy is maximized.

this constraint using our method and compare it with other exiting methods and the offline optimal (please see the remark following Corollary 1 to see the steps of storage sizing using our method). Fig. 8 shows that storage sizing using our model is near-optimal, meaning that it almost matches the *offline optimal*, computed assuming that we have the entire future solar power trace in advance (tagged as simulation). In contrast, we observe that the existing models (G-model and CSI model) significantly overestimate the required size of the ESS. For example, for a constant target output power of 50W (during the On period), both the offline optimal and our model compute the required size of the battery to be 100Wh, which is substantially smaller than 300Wh obtained with the G-Envelope model, or 1000Wh obtained with the CSI model.

Fig. 8 also shows that the benefit of increasing the size of the storage (in terms of increasing the output power) is much larger for small values of storage sizes. This is because by adding ESS to the system the best we can expect is to ideally have the average power as the output power. Thus, there will be a point after which increasing the size of ESS is not useful anymore.

C. The optimum time interval

In this section, we study how the size of the *On* period (T_{on}) affects the total energy harvested from the solar PV panels. There is an inherent tradeoff in choosing T_{on}, as the total energy harvested is the product of T_{on} and the available power (which is inversely proportional to T_{on}). Fig. 9 illustrates the total harvested energy as a function of T_{on}. The *On* period is chosen to be of size T_{on} and is symmetric over 12noon. The optimum time interval varies as a function of the sky cover and the season as illustrated in Fig. 9. The graph shows that our model can estimate the total harvested energy and the optimal T highly accurately. We observe a higher accuracy of our model for smaller values of T in Fig. 9 and we start to lose the accuracy as T increases. The reason is that we have simply assumed that the border time scale T is equal to the size of the time interval and obviously we lose accuracy as the size of the time interval increases.

VII. Conclusion

Integrating solar power in the grid requires solar power to be shaped to a manageable form. One of the best methods to convert the intermittent solar power to a reliable output power is to use energy storage devices. An important research question is how to find the minimum ESS size of a system to ensure that a target output function can be provided, while keeping the loss of power risk below an allowable threshold ε^*. Simulating the SoC analysis can be used to estimate the risk of loss of power using quantiles; however, this method needs a large data trace and it has a huge time complexity as it must be repeated for any change of parameters. To account for these shortcomings, analytical methods can be alternatively used. Unfortunately, the existing analytical methods cannot compete with the simulation methods as they are overly conservative. A tight analytical framework needs a meticulous solar power modelling. In this paper, we propose a precise analytical solar power model by carefully studying and separating the underlying processes of solar power and describing each, individually. Using this model, we are able to size the storage for *any* target output function; something which was not efficiently doable with the state-of-the-art. Our numerical examples illustrate that our solar power model is precise, closely following the simulation results, and considerably outperforming the previous ones.

References

[1] http://www.archive.arm.gov.

[2] C. Abbey and G. Joos. Supercapacitor energy storage for wind energy applications. *Industry Applications, IEEE Transactions on*, 43(3):769 – 776, 2007.

[3] H. F. Assuno, J. F. Escobedo, and A. P. Oliveira. Modelling frequency distributions of 5 minute-averaged solar radiation indexes using beta probability functions. *Theoretical and Applied Climatology*, 75(3):213 – 224, 2003.

[4] M. Beaudin, H. Zareipour, A. Schellenberglabe, and W. Rosehart. Energy storage for mitigating the variability of renewable electricity sources: An updated review. *Energy for Sustainable Development*, 14(4):302 – 314, 2010.

[5] R. E. Bird and R. Hulstrom. *A Simplified Clear Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces*. Solar Energy Research Institute, 1981.

[6] H. Chen, T. N. Cong, W. Yang, C. Tan, Y. Li, and Y. Ding. Progress in electrical energy storage system: A critical review. *Progress in Natural Science*, 19(3):291 – 312, 2009.

[7] F. Ciucu, A. Burchard, and J. Liebeherr. Scaling properties of statistical end-to-end bounds in the network calculus. *IEEE Transactions on Information Theory*, 52(6):2300 – 2312, June 2006.

[8] M. A. Delucchi and M. Z. Jacobson. Providing all global energy with wind, water, and solar power, part II: Reliability, system and transmission costs, and policies. *Energy Policy*, 39(3):1170 – 1190, 2011.

[9] R. Doherty and M. O'Malley. A new approach to quantify reserve demand in systems with significant installed wind capacity. *IEEE Transactions on Power Systems*, 20(2):587 – 595, 2005.

[10] G. Energy and A. Truewind. Ontario wind integration study. tech. rep., ontario power authority, 2006. available [on line] at: http://www.ieso.ca/imoweb/pubs/marketreports/ OPA-Report-200610-1.pdf.

[11] Y. Ghiassi-Farrokhfal, S. Keshav, and C. Rosenberg. Towards a realistic storage modelling and performance analysis in smart grids. Technical Report CS-2013-18, Department of Computer Science, University of Waterloo, Nov 2013.

[12] L. Gu, J. D. Fuentes, M. Garstang, J. T. da Silva, R. Heitz, J. Sigler, and H. H. Shugart. Cloud modulation of surface solar irradiance at a pasture site in southern Brazil. *Agricultural and Forest Meteorology*, 106(2):117 – 129, 2001.
[13] C. Gueymard. Critical analysis and performance assessment of clear sky solar irradiance models using theoretical and measured data. *Solar Energy*, 51(2):121 – 138, Auguest 1993.
[14] H. Holttinen. *The impact of large scale wind power production on the Nordic electricity system*. PhD thesis, Helsinki University of Technology, Finland, 2004.
[15] R. Hulstrom. *Solar Resources*. MIT press, 1989.
[16] H. Ibrahim, A. Ilinca, and J. Perron. Energy storage systemscharacteristics and comparisons. *Renewable and Sustainable Energy Reviews*, 12(5):1221 – 1250, 2008.
[17] M. D. Islam and R. H. B. Exell. Solar radiation mapping from satellite image using a low cost system. *Solar Energy*, 56(3):225 – 237, 1996.
[18] Y. Jiang and Y. Liu. *Stochastic Network Calculus*. Springer Verlag, 2008.
[19] M. Jurado, J. Caridad, and V. Ruiz. Statistical distribution of the clearness index with radiation data integrated over five minute intervals. *Solar Energy*, 55(6):469 – 473, 1995.
[20] A. Mefti, A. Adane, and M. Y. Bouroubi. Satellite approach based on cloud cover classification: Estimation of hourly global solar radiation from meteosat images. *Energy Conversion and Management*, 49(4):652 – 659, 2008.
[21] F. Musgens and K. Neuhoff. Modelling dynamic constraints in electricity markets and the costs of uncertain wind output. cambridge working papers in economics; 2006. Available [on line] at: http://www.dspace.cam.ac.uk/handle/1810/131648.
[22] A. E. Sarasua, M. G. Molina, and P. E. Mercado. *Energy Storage - Technologies and Applications*. InTech, 2013.
[23] S. Shoenung. Characteristics and technologies for long- vs. short-term energy storage. United States Department of Energy; 2001. Marc.
[24] A. K. Sinhaa and P. Bajpai. Swarm intelligence based optimal sizing of solar pv, fuel cell and battery hybrid system. In *Proc. of International Conference on Power and Energy Systems Lecture Notes in Information Technology*, pages 467 – 473, October 2012.
[25] H. Suehrcke and P. McCormick. The frequency distribution of instantaneous insolation values. *Solar Energy*, 40(5):413 – 422, 1988.
[26] K. Wang, F. Ciucu, C. Lin, and S. H. Low. A stochastic power network calculus for integrating renewable energy sources into the power grid. *IEEE Journal on Selected Areas in Communications*, 30(6):1037 – 1048, July 2012.
[27] K. Wang, S. Low, and L. Chuang. How stochastic network calculus concepts help green the power grid. In *Proc. of IEEE SmartGridComm*, pages 55 – 60, October 2011.
[28] J. Xie and Y. Jiang. A temporal network calculus approach to service guarantee analysis of stochastic networks. In *Proc. of VALUETOOLS*, pages 408 – 417, October 2011.

APPENDIX
PROOF OF COROLLARY 1

The existing state of charge recursion equations assume that storage is initially fully charged. This, however, is not the case for our application as we discussed. Suppose that B_0^d ($= B - B_0$) represents the initial deficit charge of the battery. The existing recursive equations assume that $B_0^d = 0$. To account for non-zero B_0^d, one can add B_0^d as additional discharge at time zero, i.e., $D(s,t) + B_0^d I_{s=0}$ instead of $D(s,t)$, and then safely set $B_0^d = 0$. Thus, given an ideal storage (α_c, $\alpha_d = 0$, $\eta = 1$), the exact loss of power formulation from [27] can be refined as follows to account for the initial state of charge

$$l(t) = \min_{0\le u<t}\Bigg(\max_{u\le s<t}\big([D(s,t) - S(s,t) - k(t) + B_0^d I_{s=0}]_+, \\ D(u,t) - S(u,t) + k(u) - k(t) + B_0^d I_{u=0}\big)\Bigg), \tag{20}$$

where

$$k(t) = \begin{cases} B & t > 0 \\ 0 & t = 0, \end{cases} \tag{21}$$

and $I_{expr} = 1$ if $expr$ is true and $I_{expr} = 0$, otherwise.

Eq. (20) can be extended to account for a non-ideal storage system using virtual processes S', D', and B' as described in Eqs. (15-19). By picking two specific values for u ($= t-1$ and $= 0$) in the minimization in Eq. (20), we have the following inequality for any $t > 1$

$$l(t) \le \min\Big([D'(t-1,t) - S'(t-1,t)]_+, \\ \max_{0\le s<t}\left([D'(s,t) - S'(s,t) - B_0]_+\right)\Big), \tag{22}$$

where in the second term, we use the fact that $k(t) - B_0^d I_{s=0} \ge B_0$ for any $s \ge 0$ and $t > 0$. Using Eq. (22), we can prove the corollary as follows:

$$\begin{aligned}
&\Pr\{l(t) > 0\} \\
&\le \Pr\Bigg\{\min\Bigg([D'(t) - S'(t)]_+, \\
&\qquad \max_{0\le s<t}\left([D'(s,t) - S'(s,t) - B_0]_+\right)\Bigg) > 0\Bigg\} && (23) \\
&\le \min\Bigg(\Pr\left\{[D'(t) - S'(t)]_+ > 0\right\}, \\
&\qquad \Pr\left\{\max_{0\le s<t}[D'(s,t) - S'(s,t) - B_0]_+ > 0\right\}\Bigg) && (24) \\
&\le \min\left(\varepsilon_0, \Pr\left\{\max_{0\le s<t}\left(\mathcal{G}_d(s,t) - S'(s,t)\right) > B_0\right\}\right) && (25) \\
&\le \min\left(\varepsilon_0, \varepsilon_s\left(B_0\right)\right) && (26) \\
&= \varepsilon^*, && (27)
\end{aligned}$$

where we use Eq. (22) in the second line. Eq. (24) is an upper bound on Eq. (23) using the fact that $P(X \cap Y) \le \min(P(X), P(Y))$ for any events X and Y. We use Eq. (18) to obtain the first term in Eq. (25). To obtain the second terms in Eq. (25) and Eq. (26), we use the corollary assumption that $\mathcal{G}_d(s,t) = D'(s,t)$ is a dynamic lower envelope on S' with bounding function ε_s in the sense of Eq. (6). We finally, use the definition of ε^* from Eq. (19) to obtain the last line.