



# A Smart Grid Energy Management Problem for Data-driven Design with Probabilistic Reachability Guarantees (Benchmark Proposal)

Daniele Ioli<sup>1</sup>, Alessandro Falsone<sup>1</sup>, Marianne Hartung<sup>2</sup>, Axel Busboom<sup>2</sup>, and  
Maria Prandini<sup>1</sup>

<sup>1</sup> Politecnico di Milano, Milano, Italy

{daniele.ioli,alessandro.falsone,maria.prandini}@polimi.it

<sup>2</sup> GE Global Research, Munich, Germany

{hartungm,busboom}@ge.com

## Abstract

In this paper we describe an energy management benchmark problem for a smart grid where electrical energy is supplied to a load via local power production from a solar PhotoVoltaic (PV) installation. The smart grid is connected with the main grid, which can eventually provide the energy needed for balancing demand and generation. The goal is to set the battery energy flow so as to keep the energy exchange with the main grid as close as possible to a nominal profile, within certified bounds, avoiding the fluctuations caused by the local PV energy production. Some energy production profiles of the PV installation and environmental data on irradiation and temperature are available for the design of the energy management strategy, together with a hybrid model for the battery and the electrical load profile. We describe a data-driven solution, pointing out its limits and providing some hint on possible direction for improvement.

## 1 Introduction

This paper proposes a benchmark problem in energy management of a smart grid connected to the main grid. The considered set-up is quite simple since it comprises only a few elements, that is photovoltaic (PV) solar panels, a simulated electrical load, a battery, and related interfaces to the control unit, with inverters and metering systems. Yet it presents some of the key difficulties (control of a hybrid system modeling the battery, presence of stochastic uncertainty due to renewable energy sources) encountered in the energy management of a larger scale smart grid.

Electrical energy storage systems are becoming more and more widely used, in particular, as a means to deal with the intermittent and non-controllable nature of the power production by renewable energy supply systems such as solar PV installations and wind turbines, since they can serve the purpose of storing energy when production exceeds demand and releasing

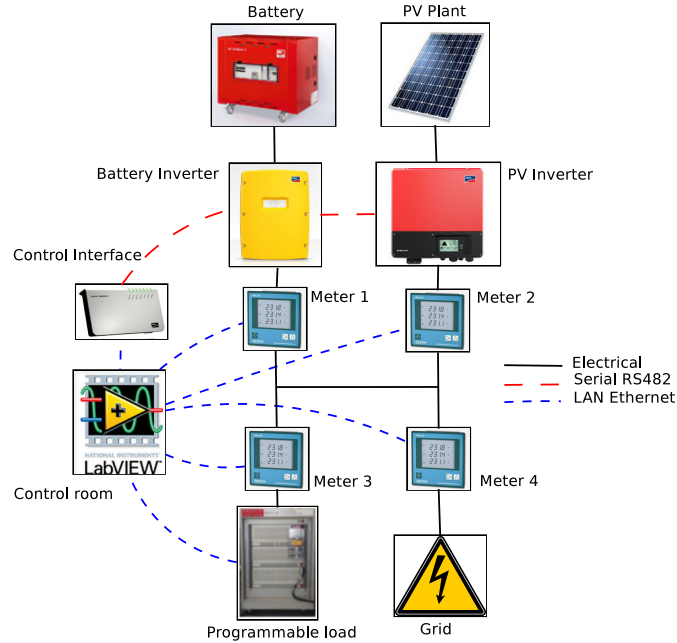


Figure 1: Smart grid configuration.

it when load request exceeds generation, see e.g. [8] and the references therein. Studies on electrical storage systems suitable for energy management in small grids or even single building applications, [15], and even stand-alone electricity generation systems relying to a significant extent on the supply of solar or wind energy, have been developed in the literature, [12].

The goal here is to define a strategy for the battery usage so as to compensate for the PV energy production fluctuations and guarantee that the energy exchange with the main grid is as close as possible to a nominal profile, within certified bounds, with reference to a one-day time horizon. This will reduce the uncertainty that the main grid is experiencing and, in a more global perspective of distributed energy resource systems, simplify the task of the main grid operator when defining the reserves and ancillary services needed to safely operate the grid, avoiding service disruption.

Due to the stochastic nature of the PV energy production, the problem is naturally formulated in a stochastic framework and the challenge becomes providing the probabilistic guarantees on the region where the energy request to the main grid is confined, based on a limited number of PV energy production profiles only, and some measurements of environmental data on irradiance and air temperature. A data-driven solution resting on the so-called *scenario theory*, [7], originally introduced in [4, 5] to address robust design and then extended to chance constrained optimization in [6], is proposed and possible directions for improvement are discussed.

All available data are described at the end of Section 4.2 and can be downloaded from the website <http://cps-vo.org/group/ARCH/benchmarks> to allow other approaches to be developed and compared.

The rest of the paper is organized as follows. We describe the smart grid configuration in Section 2. The benchmark energy management is presented in Section 3. A data-driven solution is proposed in Section 4 and its performance is assessed in Section 5. Concluding remarks are drawn in Section 6.

## 2 Description of the smart grid configuration

We consider the smart grid testbed in Figure 1, which has been set up at *General Electric Global Research Center (GERC)* in Munich, Germany, in the *Hertz Lab*.

The testbed consists of a PV installation with the associated PV inverter, a programmable load simulator, a battery storage unit with its inverter, and a grid interface. Four smart meters provide measures of the involved energy flows, with the meters of the PV installation and battery located downstream the corresponding inverters so as to include power conversion losses. A control interface is connected to the inverters via a serial RS-485 interface and can collect from them monitoring data and also set some of their configuration parameters. This latter functionality, in particular, allows to control the energy exchange with the battery.

The control strategy for the battery is implemented via NI Labview (<http://www.ni.com/labview/>), which runs on a desktop personal computer located in the control room. An Ethernet based Local Area Network (LAN) makes the control room communicate with the control interface, the meters, and the programmable load, via the MODBUS protocol.

### 2.1 Model of the electrical storage unit

A Li-ion battery produced by the company IBC-Solar is adopted for electrical energy storage in the considered testbed. The battery is combined with an SMA SunnyIsland inverter. A Battery Management System (BMS) at the interface of the battery with the inverter prevents its overheating (for safety) and deep discharge (to avoid shortening the battery life). The main role of the BMS is driving the inverter and selecting the working point on the voltage-current curve so as to track the desired power flow while pursuing the maximal efficiency for the battery.

From now on we will refer to the battery together with its inverter as the *storage*, and we shall derive a model of the storage for the design of the smart grid energy management strategy.

The plot on the left of Figure 2 represents the results of a power flow tracking experiment. Note that the power flow can be controlled with a high precision (of the order of 10 W) and that the storage system presents a fairly symmetric behavior when charging and discharging, with the power dynamics overshooting when the operational mode changes from charging to discharging or viceversa. Since the transient duration when commuting is of the order of 10 seconds and we are interested in a time scale of minutes, we shall neglect this fast dynamics in the model of the storage.

Figure 2 shows on the right a cycle where the storage is first discharged and then charged at a maximum constant power rate of about  $\pm 3.8$  kW when its state of charge (SOC) is between 5% and 95%. As the SOC grows beyond 95%, the maximum charging power rate decreases exponentially. Other experimental tests suggest not to reduce the SOC of the storage below 5% since the power flow control could be overtaken by the BMS to the purpose of preventing deep discharge. It is then convenient to limit the operational SOC range of the storage system between 5% and 95%.

The maximum power rates in charging and discharging that were tested are of about 3.797 kW and 3.822 kW, respectively. Charging and discharging rates depend on the SOC and the battery temperature. Empirical insight suggests to take 3.5 kW as maximum power rate in both the charging and discharging phases since this is the (maximum) value that is possible to track in all SOC and temperature operating conditions. The round trip efficiency is 4.07% and it is calculated as the energy mismatch for discharging the storage from 76% to 5% SOC and then recharging it back to 76% SOC as shown in the experiment reported on the right plot of Figure 2. The same experiment allows to estimate the capacity of the storage that is 24.780MJ (22.680 MJ according to the datasheet).

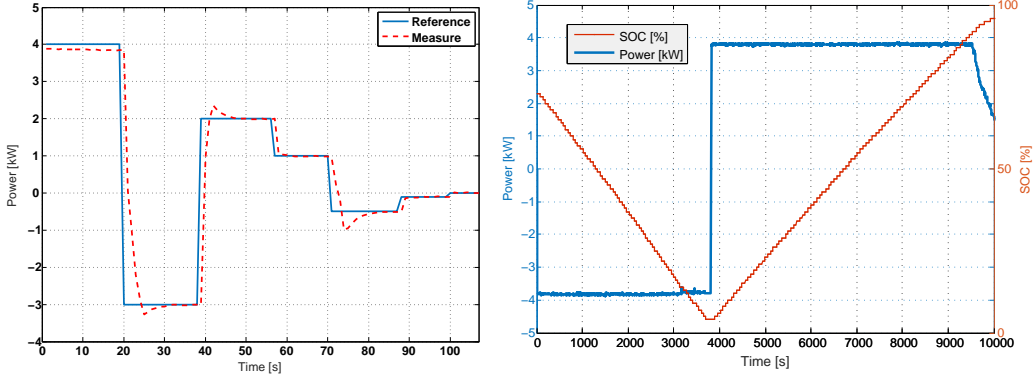


Figure 2: Power flow tracking experiment (left) and a cycle where the storage is discharged and charged at maximal rate (right).

Based on the experimental observations described above, we can introduce a sampling time interval  $\Delta$  and model the storage system as a discrete time hybrid system where the stored energy  $x$  evolves according to

$$x(k+1) = ax(k) + b(u(k))u(k), \quad (1)$$

where  $u(k)$  represents the energy fed into ( $u(k) > 0$ ) or drawn from ( $u(k) \leq 0$ ) the storage in the  $k$ -th time slot of duration  $\Delta$ , and it is re-scaled by coefficient

$$b(u) = \begin{cases} 1 - \varepsilon_u & u > 0 \\ 1 + \varepsilon_u & u \leq 0, \end{cases}$$

with  $\varepsilon_u \in (0, 1)$ . This coefficient accounts for the round trip efficiency: if we feed the storage with an amount of energy  $E_{in}$ , then the storage content increases of  $(1 - \varepsilon_u)E_{in}$ , and if we draw an amount of energy  $E_{out}$ , the storage content decreases of  $(1 + \varepsilon_u)E_{out}$ . Given that in a cycle where the storage is charged and discharge we waste a percentage of energy equal to 4.07%, then we can easily compute  $\varepsilon_u$  as follows  $0.0407E_{in} = E_{in} - (1 - \varepsilon_u)E_{in}/(1 + \varepsilon_u)$ , thus getting  $\varepsilon_u \simeq 0.02$ . Notice that we assume exchange losses to be symmetric between charging and discharging. Parameter  $a$  in (1) depends on  $\Delta$  and accounts for the self discharging of the storage (Faradaic efficiency), which in the considered storage system turns out to be quite small. More precisely, if we set the sampling time equal to  $\Delta = 1$  minute, then  $a = 0.9998$ .

Model (1) is valid when the SOC of the storage is within 5% and 95%, which translates into the constraint

$$x_{\min} \leq x(k+1) \leq x_{\max}, \quad k = 0, 1, \dots, k_f - 1, \quad (2)$$

where  $x_{\min} = 1.239$  MJ (5% SOC) and  $x_{\max} = 23.54$  MJ (95% SOC). The energy exchanged with the storage is subject to the following constraint:

$$u_{\min} \leq u(k) \leq u_{\max}, \quad k = 0, 1, \dots, k_f - 1, \quad (3)$$

where  $u_{\max} = -u_{\min} = \bar{u}\Delta$ ,  $\bar{u} = 3.5$  kW being the maximum power rate in both the charging and discharging phases. If  $\Delta = 1$  minute, then,  $u_{\max} = -u_{\min} = 0.21$  MJ.

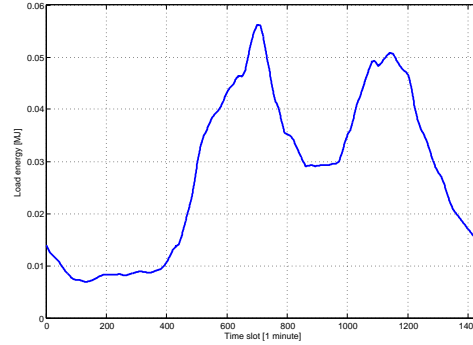


Figure 3: Simulated load profile. Values refer to electrical energy request per minute.

## 2.2 Electrical load

Prodigit 32612A is a device which can simulate both AC and DC electrical loads, with a maximum voltage of 300 V and maximum power of 5.4 kW, and is commonly used for inverter testing. In the smart grid testbed, it simulates the purely resistive load in Figure 3 that represents a typical household application rescaled so that its daily integral is comparable with the PV installation average daily energy production.

Notice that the simulated load is deterministic. The stochastic ingredient is brought into the problem by the PV energy production which is added to the load demand in the grid energy balance equation. The proposed approach to the design of an energy management strategy can be easily extended to the case when both the PV energy production and the load are stochastic.

## 2.3 PV installation & Inverter

The testbed is equipped with a PV installation composed of 10 Hanwha HSL72 Polycrystalline silicon modules 300 W each for an overall nominal power of 3 kW. These panels are placed on the roof of the *Hertz lab* and connected to the smart grid via a SMA Sunny Boy Inverter. The rack is facing south and is tilted of 28 degrees.

The SMA Sunny Boy converts the direct current of the PV array to grid-compliant alternating current and feeds it into the utility grid. The inverter can communicate with the control interface via a serial RS-485 interface.

A weather station equipped with irradiance and temperature sensors is placed on the roof of the *Hertz lab* (see Figure 4). Tilted global irradiance measures are made available through an irradiance silicon sensor tilted of 28 degrees and oriented to the south so that its measures are consistent with the radiation received by the array of PV panels. The temperature of each PV module is measured by a thermocouple placed on its backside, as shown in Figure 4. Measures are sampled every thirty seconds and stored in a database.



Figure 4: Weather station (on the left) and module temperature sensor (on the right).

### 3 Energy management benchmark problem

Consider a one-day time horizon, discretized into  $k_f = 1440$  time steps of length  $\Delta = 1$  minute. The energy  $g$  exchanged with the main grid at step  $k \in \{0, 1, \dots, k_f - 1\}$  corresponding to the time frame  $[k\Delta, (k+1)\Delta)$  is given by the following energy balance equation

$$g(k) = \ell(k) - d(k) + u(k), \quad k = 0, 1, \dots, k_f - 1, \quad (4)$$

where  $g(k) > 0$  if the main grid provides energy to the smart grid in the  $k$ -th time frame, and  $g(k) < 0$  viceversa. All quantities  $\ell(k)$ ,  $d(k)$ ,  $u(k)$  refer to the  $k$ -th time frame and represent, respectively, the electrical load energy request, the energy produced by the PV installation, and the energy exchanged with the storage.

The load request  $\ell$  is set equal to the (deterministic) profile in Figure 3. The energy exchanged with the storage  $u$  affects the storage content according to the model (1) in Section 2.1, and is subject to the constraints (2) and (3). The PV energy production profiles are uncertain and we assume that they are realizations of an underlying stochastic process, distributed according to some probability measure  $\mathbb{P}$ . Probability  $\mathbb{P}$  is not known explicitly but only indirectly through a dataset of PV energy production profiles.

The goal is to appropriately set the energy flow with the battery so as to keep the energy exchange with the main grid along the one-day time horizon as close as possible to some nominal profile avoiding high fluctuations due to the energy production from renewable solar power. The nominal profile can be optimized so as to minimize the variability range of the energy exchanged with the main grid, while providing certified bounds on its extent. Given that the PV energy production is uncertain, the bounds are certified in probability.

More precisely, let  $\bar{g}(k)$ ,  $k = 0, \dots, k_f - 1$ , be the nominal profile to be jointly optimized with the energy flow with the battery  $u(k)$ ,  $k = 0, \dots, k_f - 1$ . Then, the energy management problem can be formulated as the following chance-constrained optimization program

$$\begin{aligned} & \min_{\substack{h_g, \bar{g}(k), k=0, \dots, k_f-1 \\ u(k), k=0, \dots, k_f-1}} h_g & (5) \\ \text{subject to: } & \mathbb{P} \left\{ \left| \sum_{k=jw+1}^{(j+1)w} \delta g(k) \right| \leq h_g, \quad j = 0, 1, \dots, \lfloor k_f/w \rfloor - 1, \right. \\ & \left. x_{\min} \leq x(k+1) \leq x_{\max}, u_{\min} \leq u(k) \leq u_{\max}, k = 0, \dots, k_f - 1 \right\} \geq 1 - \epsilon \end{aligned}$$

where the extent of the fluctuations  $\delta g(k) = g(k) - \bar{g}(k)$ ,  $k = 0, \dots, k_f - 1$ , integrated over time windows of length  $w$ ,  $1 \leq w \leq k_f$ , is minimized through its upper bound  $h_g$ , while imposing the actuation constraints (2) and (3). Note that the minimization of  $h_g$  is performed jointly with the constraint enforcement over a set of realizations of the PV energy production of probability larger than or equal to  $1 - \epsilon$ , where the violation probability  $\epsilon \in (0, 1)$  is a design parameter and modulates the reachability region extent: the smallest  $\epsilon$ , the larger the reachability region. The presence of the time window length parameter  $w$  adds some flexibility to the problem: if  $w = 1$ , then, fluctuations over a time scale of length  $\Delta = 1$  minute are minimized, whereas choosing  $w > 1$  can accommodate for a larger time scale of operation at the main grid level.

The challenge is now to solve the chance constrained optimization program (5) – which is a difficult problem on its own, [17, 18] – knowing  $\mathbb{P}$  only indirectly, through a limited number of PV energy production profiles and some measurements of environmental data on irradiance and air temperature.

We shall next describe a data-driven solution to this problem. The proposed approach can also be applied to variants of the problem like, e.g., the case when the nominal profile is assigned and given by the main grid operator, or when both PV generation and load are stochastic and  $\mathbb{P}$  represents their joint probability distribution.

## 4 Proposed solution

### 4.1 Control policy parametrization

Given that the PV energy production  $d$  is available through measurements and the goal is to limit its impact on the energy exchange with the main grid, we shall adopt a disturbance compensation scheme where the control input  $u$  is chosen as follows

$$u(k) = \gamma_k + \sum_{j=0}^{k-1} \theta_{k,j} (d(j) - \bar{d}(j)), \quad (6)$$

$\bar{d}$  being the expected value of the PV energy production  $d$ . The parameters  $\gamma_k \in \mathbb{R}$  and  $\theta_{k,j} \in \mathbb{R}$  are decision variables to be optimized.

Let

$$\mathbf{u} = \begin{bmatrix} u(0) \\ u(1) \\ \vdots \\ u(k_f - 1) \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} d(0) \\ d(1) \\ \vdots \\ d(k_f - 1) \end{bmatrix} \quad \boldsymbol{\ell} = \begin{bmatrix} \ell(0) \\ \ell(1) \\ \vdots \\ \ell(k_f - 1) \end{bmatrix} \quad \mathbf{g} = \begin{bmatrix} g(0) \\ g(1) \\ \vdots \\ g(k_f - 1) \end{bmatrix}$$

be the vectors collecting all relevant quantities to the energy balance equation (4) along the reference one-day time-horizon. Then, equation (4) can be rewritten in vectorial form as

$$\mathbf{g} = \boldsymbol{\ell} - \mathbf{d} + \mathbf{u} \quad (7)$$

and the following expression can be derived for  $\mathbf{u}$  as a function of  $\mathbf{d}$

$$\mathbf{u} = \boldsymbol{\Gamma} + \boldsymbol{\Theta}(\mathbf{d} - \bar{\mathbf{d}}), \quad (8)$$

where  $\Gamma$  and  $\Theta$  are given by

$$\Gamma = \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{k_f-1} \end{bmatrix} \quad \Theta = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \theta_{1,0} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \theta_{k_f-1,0} & \cdots & \theta_{k_f-1,k_f-2} & 0 \end{bmatrix}.$$

By plugging equation (8) into equation (7) and reordering some terms, we obtain

$$\mathbf{g} = \ell - \bar{\mathbf{d}} + \Gamma + (\Theta - I)(\mathbf{d} - \bar{\mathbf{d}}). \quad (9)$$

We can then set

$$\bar{\mathbf{g}} = \ell - \bar{\mathbf{d}} + \Gamma \quad (10)$$

$$\delta\mathbf{g} = (\Theta - I)(\mathbf{d} - \bar{\mathbf{d}}) \quad (11)$$

and rewrite (9) as  $\mathbf{g} = \bar{\mathbf{g}} + \delta\mathbf{g}$ . The quantities  $\bar{\mathbf{g}}$  and  $\delta\mathbf{g}$  in (11) represent, respectively, the nominal exchange profile with the main grid and its fluctuations. Note that  $\delta\mathbf{g}$  depends only on  $\Theta$  and cannot be set equal to zero by imposing  $\Theta = I$  since  $\Theta$  is lower triangular due to the compensation scheme causality. Also,  $\Theta$  cannot be set arbitrarily since the actuation constraints on the storage capacity and energy exchange rate (see (2) and (3)) have to be satisfied. The nominal profile  $\bar{\mathbf{g}}$  depends only on  $\Gamma$  and not on  $\Theta$ . However, the choice of  $\Gamma$  is coupled with that of  $\Theta$  due to the actuation constraints, which are next written in vectorial form for uniformity. This is straightforward for the constraint on the energy exchange rate:

$$\mathbf{1}u_{\min} \leq \mathbf{u} \leq \mathbf{1}u_{\max}, \quad (12)$$

where  $\mathbf{1}$  is a vector with all elements equal to 1 of the same dimension of  $\mathbf{u}$ . As for the storage capacity constraint, in order to make it explicit as a function of  $\mathbf{u}$ , we need first to define the stored energy vector  $\mathbf{x} = [x(1) \ x(2) \ \dots \ x(k_f)]^\top$ . Then, it is easily seen from (1) that

$$\mathbf{x} = Fx_0 + G(I - \varepsilon_u \text{diag}(\text{sign}(\mathbf{u}))) \mathbf{u} \quad (13)$$

where  $x_0$  is the initial storage content,  $\text{diag}(\text{sign}(\mathbf{u}))$  is the diagonal matrix with the sign of the elements of  $\mathbf{u}$  on the diagonal, and matrices  $F$  and  $G$  are given by

$$F = \begin{bmatrix} a \\ a^2 \\ \vdots \\ a^{k_f} \end{bmatrix} \quad G = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a^{k_f-1} & \cdots & a & 1 \end{bmatrix}.$$

This finally leads to the following vectorial form for (2)

$$\mathbf{1}x_{\min} \leq Fx_0 + G(I - \varepsilon_u \text{diag}(\text{sign}(\mathbf{u}))) \mathbf{u} \leq \mathbf{1}x_{\max},$$

which – by observing that the inequality  $|G\varepsilon_u \text{diag}(\text{sign}(\mathbf{u}))\mathbf{u}| \leq \varepsilon_u G\mathbf{1} \max_{k=0,1,\dots,k_f-1} |u_k|$  holds componentwise – can be enforced by adopting the conservative approximation

$$\mathbf{1}x_{\min} + \Xi h(\mathbf{u}) \leq Fx_0 + G\mathbf{u} \leq \mathbf{1}x_{\max} - \Xi h(\mathbf{u}), \quad (14)$$



where  $\Xi = \varepsilon_u G \mathbf{1}$  and  $h(\mathbf{u}) = \max_{k=0,1,\dots,k_f-1} |u_k|$ . This approximation in practice reduces the model of the storage from hybrid to linear, and simplifies the design of the energy management strategy.

We can now finally formulate the energy management problem in terms of the following chance-constrained optimization program

$$\begin{aligned} & \min_{\Gamma, \Theta, h_g, h_u} h_g + \rho_g \|\bar{\mathbf{g}}\|_2 + \rho_u h_u & (15) \\ \text{subject to:} & \quad \mathbb{P}\left\{ \left| \sum_{k=jw+1}^{(j+1)w} \delta g(k) \right| \leq h_g, j = 0, 1, \dots, \lfloor k_f/w \rfloor - 1, |\mathbf{u}| \leq \mathbf{1}h_u, \right. \\ & \quad \left. \mathbf{1}x_{\min} + \Xi h_u \leq Fx_0 + G\mathbf{u} \leq \mathbf{1}x_{\max} - \Xi h_u, \mathbf{1}u_{\min} \leq \mathbf{u} \leq \mathbf{1}u_{\max} \right\} \geq 1 - \epsilon \end{aligned}$$

which differs from (5) in that the penalization terms  $\rho_g \|\bar{\mathbf{g}}\|_2$  and  $\rho_u h_u$  with weights  $\rho_g, \rho_u > 0$  are added to  $h_g$ . This is in order to have an uniquely defined solution for  $\Gamma$ , and to reduce the conservatism of the approximated storage capacity constraint (14), respectively.

## 4.2 Scenario solution

Given that the probability  $\mathbb{P}$  is not known and only a set of realizations  $\{\mathbf{d}^{(i)}, i = 1, \dots, N\}$  extracted from  $\mathbb{P}$  is available, we shall formulate the following scenario optimization problem with constraint removal to approximate the original chance-constrained program:

$$\begin{aligned} & \min_{\Gamma, \Theta, h_g, h_u} h_g + \rho_g \|\ell - \bar{\mathbf{d}} + \Gamma\|_2 + \rho_u h_u & (16) \\ \text{subject to:} & \quad \begin{cases} \left| \sum_{k=jw+1}^{(j+1)w} (\Theta - I)(\mathbf{d}^{(i)}(k) - \bar{\mathbf{d}}(k)) \right| \leq h_g, j = 0, 1, \dots, \lfloor k_f/w \rfloor - 1 \\ \mathbf{1}x_{\min} + \Xi h_u \leq Fx_0 + G\Gamma + \Theta(\mathbf{d}^{(i)} - \bar{\mathbf{d}}) \leq \mathbf{1}x_{\max} - \Xi h_u \\ \mathbf{1}u_{\min} \leq \Gamma + \Theta(\mathbf{d}^{(i)} - \bar{\mathbf{d}}) \leq \mathbf{1}u_{\max} \\ |\Gamma + \Theta(\mathbf{d}^{(i)} - \bar{\mathbf{d}})| \leq \mathbf{1}h_u \\ i \in \{i_1, i_2, \dots, i_{N-\lfloor \eta N \rfloor}\}, \end{cases} \end{aligned}$$

where  $\eta \in [0, \epsilon)$  is the empirical violation parameter and  $\{i_1, i_2, i_{N-\lfloor \eta N \rfloor}\} \subseteq \{1, 2, \dots, N\}$  contains the indices of  $N - \lfloor \eta N \rfloor$  out of the  $N$  available PV energy production realizations, which should be selected so as to best improve the cost. Each one of the removed  $\lfloor \eta N \rfloor$  realizations should change the solution if included in the scenario program (16).

If  $N$  is appropriately selected, then, the solution to the scenario program (16) is feasible for the original chance constrained problem (15). This is stated in the following proposition whose proof follows directly from [6] with the only difference that the cost is not linear and one has then to apply the considerations stated in the proof of Theorem 1 in [10].

**Proposition 1.** *Select a confidence parameter  $\beta \in (0, 1)$  and an empirical violation parameter  $\eta \in [0, \epsilon)$ . If  $N$  satisfies*

$$\binom{\lfloor \eta N \rfloor + n - 1}{\lfloor \eta N \rfloor} \sum_{i=0}^{\lfloor \eta N \rfloor + n - 1} \binom{N}{i} \epsilon^i (1 - \epsilon)^{N-i} \leq \beta,$$

where  $n$  denotes the number of optimization variables in (16), then, with probability no smaller than  $1 - \beta$ , the solution to the scenario optimization problem with constraint removal (16) satisfies the probabilistic constraint in (15).

Not surprisingly, feasibility of the scenario solution holds with a certain confidence  $1 - \beta$ . This is because the scenario solution is a random quantity that depends on the extracted multi-sample  $d^{(1)}, d^{(2)}, \dots, d^{(N)}$ . If a bad multi-sample is extracted, then, the feasibility property does not hold, but this event becomes more and more unlikely as  $N$  increases. The dependence on  $\beta$  is logarithmic (see [1, 16]) so that it can be set quite small without growing too much  $N$ . Though the feasibility of the randomized solution is guaranteed for every  $\eta \in [0, \varepsilon)$ , the closer  $\eta$  is to the desired violation probability  $\varepsilon$ , the better the randomized solution approximates the actual solution to the chance-constrained problem. At the same time, however,  $N$  grows to infinity as  $O(\frac{1}{\varepsilon - \eta})$  when  $\eta \rightarrow \varepsilon$ , [6], so that one should choose  $\eta$  based on the available computational resources.

The proposed data-driven approach requires a relatively large collection of electrical energy production profiles from the PV installation in the smart grid. 30 day of measurements were performed in August 2016. Only 28 out of the 30 PV power production profiles turned out to be not corrupted. PV module temperature, irradiation, and air temperature profiles were also collected. A further dataset of 220 irradiation and ambient temperature profiles is available. Data were collected during July, August, September of 2010, 2011, 2015, and 2016. The idea is then to reconstruct further PV power production profiles from this latter dataset. This is described in the following Section 4.3.

### 4.3 PV energy production profiles reconstruction

Reliable profiles can be reconstructed from environmental data on air temperature and irradiance. This topic is well investigated in literature (see e.g. [19]) and the use of photovoltaic power models is common practice in applied power forecasting, plant performance evaluation and design, [9]. The module power production  $P_M(t)$  at time  $t$  can be expressed as a static function of the irradiance on the tilted surface  $I(t)$  and the temperature of the module  $T_M(t)$  as follows

$$P_M(t) = \bar{P} \frac{I(t)}{\bar{I}} \left( 1 - \frac{\alpha}{100} (T_M(t) - \bar{T}) \right) \quad (17)$$

where  $\bar{T} = 25^\circ\text{C}$  and  $\bar{I} = 1 \text{ kW/m}^2$  are the standard test temperature and irradiance,  $\bar{P}$  represents the nominal power production of the panel, measured in the testing conditions of  $T_M = \bar{T}$  and  $I = \bar{I} = 1 \text{ kW/m}^2$ . The temperature coefficient  $\alpha$  is introduced to take into account the performance degradation of the panel due to overheating. In normal working conditions a module can heat up to  $70/80^\circ\text{C}$  (how much depends on the panel technology and application), and this affects significantly the solar cell efficiency, [11]. This is the case, in particular, for the polycrystalline solar panels in the testbed, which have slightly higher temperature coefficients than panels built with other technologies (e.g. Cadmium Telluride CdTe solar panels).

We shall next use the dataset of August 2016 to determine the parameters in model (17) (see Section 4.3.1). We shall then identify a model that provides the temperature of the module, based on the irradiance and air temperature in Section 4.3.2. By means of the two identified models we shall be able to reconstruct the PV energy production profiles from the dataset containing only irradiance and air temperature profiles.

#### 4.3.1 Power model identification

Equation (17) can be rewritten in regression form as follows

$$P_M(t) = \nu^\top \phi(t), \quad (18)$$

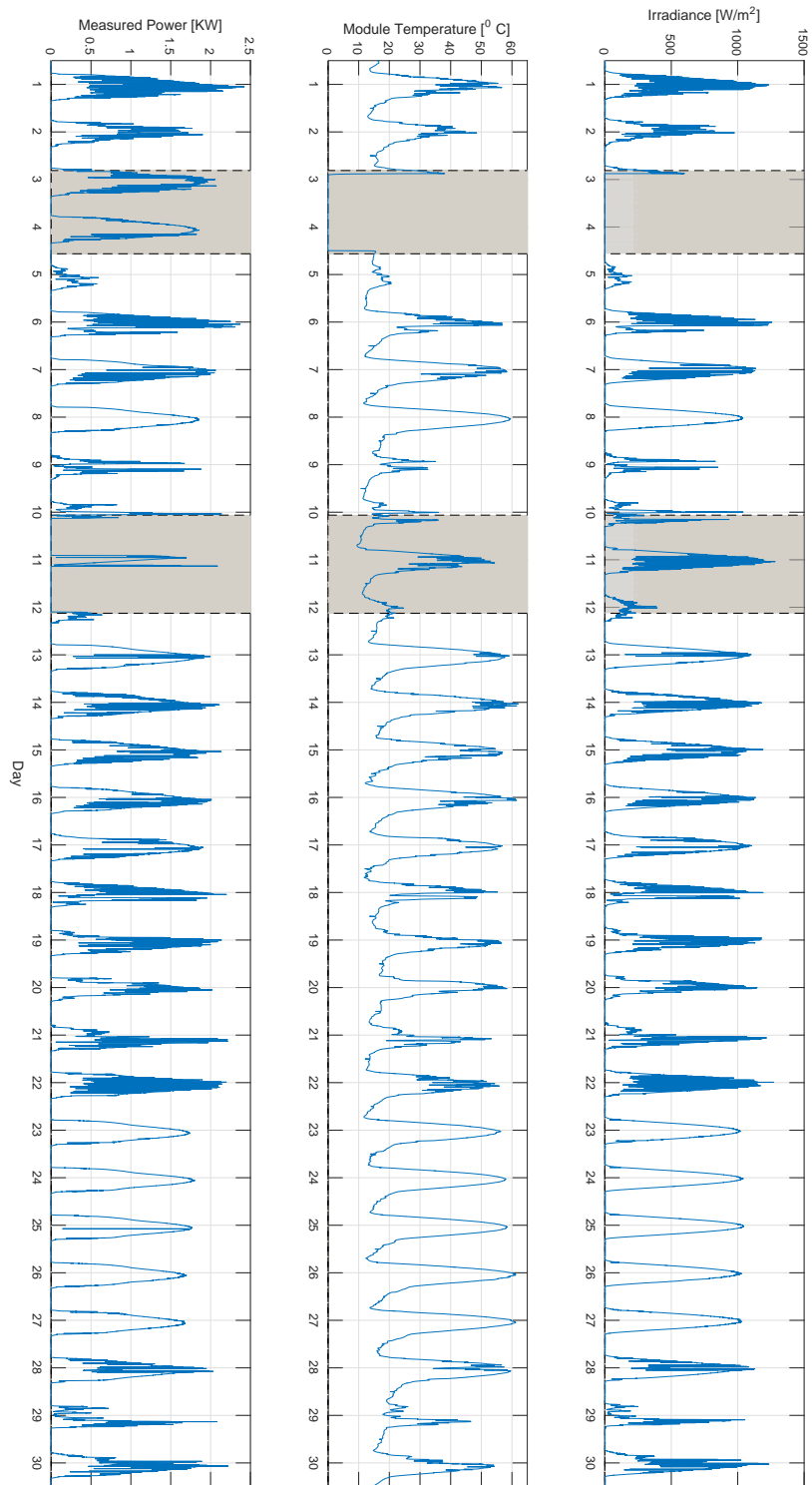


Figure 5: PV energy profiles and environmental data collected in August 2016. Gray areas correspond to days where some of the measurements were corrupted.

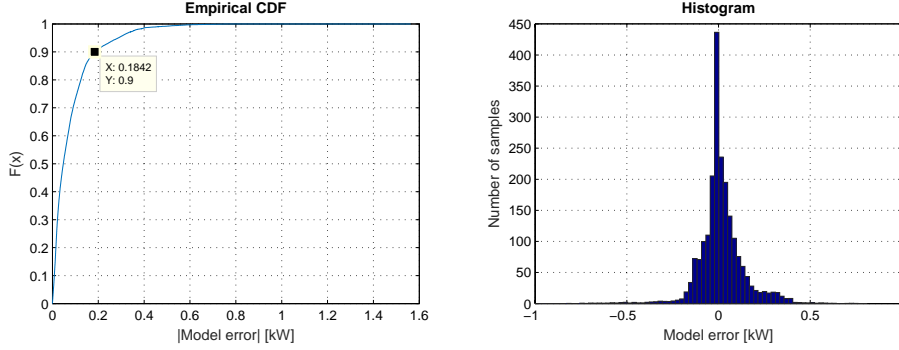


Figure 6: Power model error statistics

where  $\nu^\top = [\bar{P} \quad \alpha\bar{P}]$  and  $\phi(t)^\top = [I(k)/\bar{I} \quad -I(k)10^{-2}/\bar{I}(T_M(t) - \bar{T})]$ .

Data of PV power production, irradiance, and module temperature measured over 30 days in August 2016 are reported in Figure 5, with a sampling time of 1 minute for the power production and 30 seconds for irradiance and temperature. Over 43200 collected samples, corrupted data due to inactivity of the test plant (i.e. maintenance) and/or missing data from the weather or temperature measurement device and night time values are discarded.  $n_d = 19337$  samples, corresponding to 90% of the available data, are used to identify the parameters vector  $\nu$  by the least squares method, i.e.,

$$\nu = \left( \sum_{k=1}^{n_d} \phi(k)\phi(k)^\top \right)^{-1} \sum_{k=1}^{n_d} \phi(k)P_M(k).$$

The obtained parameters are  $\bar{P} = 2.1096$  kW and  $\alpha = 0.6566^\circ\text{C}^{-1}$ , which are similar to the values in the literature, [20]. The remaining 10% of the data are used for model validation. Statistics of the power model error are presented in Figure 6. From the empirical cumulative distribution function of the absolute value of the power error, we can observe that with 90% probability it is lower than 0.1842 kW, which is less than 8% of the nominal power  $\bar{P}$ . The expected value for the power error is 0.0084 kW and from the histogram we can see that the error is quite concentrated around the mean.

### 4.3.2 Thermal model identification

In order to reconstruct the module temperature from environmental measures of air temperature and irradiance, we propose a variant of the following model, that is proposed by Sandia National Laboratories, a multi-program laboratory managed and operated by Sandia Corporation operating for the U.S. Department of Energy's National Nuclear Security (<http://pvpmc.sandia.gov/modeling-steps/2-dc-module-iv/module-temperature/>):

$$T_M(t) = c_T I(t) e^{b_1 + b_2 W(t)} + T_A(t) \quad (19)$$

where  $c_T = 1^\circ\text{C}W^{-1}\text{m}^2$ , and  $b_1$  and  $b_2$  are parameters that depend on the module construction and materials and on its mounting configuration and their values can be found in technical references and manuals for various common module types and configurations. Model (19) is static and expresses the temperature of the module  $T_M(t)$  at time  $t$  as a function of the air

temperature  $T_A(t)$ , wind speed  $W(t)$  and irradiance  $I(t)$ . It is conceived for the estimation of the average module temperature over relatively long time periods (i.e., with a duration larger than 10 minutes). In our case, we need a smaller time resolution and, hence, we introduce a first order dynamics modeling the thermal transients of the module, thus replacing the static model (19) with the following discrete time dynamic model:

$$T_M(k) = a_1 T_M(k-1) + a_2 \left( c_T I(k) e^{b_1 + b_2 \bar{W}} + T_A(k) \right). \quad (20)$$

Note that in (20) the wind speed is replaced with its average value  $\bar{W}$ , which will then be estimated from data together with  $b_1$  and  $b_2$ .

In order to retain the physical features of the original static model, parameters  $a_1$  and  $a_2$  are chosen so that in stationary conditions (i.e., when  $T_M(k) = T_M(k-1)$ ) model (20) matches the original static model

$$\frac{a_2}{1 - a_1} = 1 \Leftrightarrow a_2 = 1 - a_1. \quad (21)$$

Finally, by setting  $a_3 = (1 - a_1) e^{b_1 + b_2 \bar{W}}$  and replacing  $a_2$  with its expression as a function of  $a_1$  in (21), model (20) is re-parameterized as follows

$$T_M(k) = a_1 T_M(k-1) + a_3 c_T I(k) + (1 - a_1) T_A(k). \quad (22)$$

Parameters  $a_1$  and  $a_3$  are then identified by minimizing the mean square prediction error

$$J(a_1, a_3) = \frac{1}{n_d} \sum_{k=1}^{n_d} \left( T_M(k) - a_1 T_M(k-1) + a_3 c_T I(k) + (1 - a_1) T_A(k) \right)^2,$$

subject to the physical constraints  $0 \leq a_1 < 1$  and  $a_3 > 0$ . Data referring to day time only were used in the identification since nocturnal radiative cooling effects are not captured in the proposed model. Identification is performed on  $n_d = 5436$  samples, roughly 70% of daytime data measured every 30 seconds in a 6 days long record, the remaining 2300 (30% of data) are then used for validation.

The identified parameters values are  $a_1 = 0.8682$  and  $a_3 = 0.0046$ . The mean value of the prediction error (considering day time only) is  $0.049^\circ C$ . The empirical cumulative distribution function of the absolute value of the prediction error plotted in Figure 7 shows that for 90% of the samples the magnitude of the prediction error is smaller than  $2.227^\circ C$  for the dynamical model (20). This value roughly doubles in the case we use the original static model. This error magnitude appears acceptable based on some studies in the literature (see e.g., [13]).

### 4.3.3 Power data reconstruction

We use irradiance and air temperature data records for the period of July, August, September of the years 2010, 2011, 2015, and 2016 for an overall number of 220 daily realizations to first reconstruct the module temperature (Section 4.3.2) and then to obtain realizations of PV power production (Section 4.3.1). Some post processing to lower the noise on top of the tails for the reconstructed realizations due to the irradiance sensor was then performed. In Figure 8 the daily mean profile of the reconstructed data set and that of the real measurements are compared: the latter is more oscillating and noisy because of the limited number of realizations (only 28). The main differences can be found at the very beginning and at the end of the production period, which are most probably due to some model error for the power production at low value of irradiance.

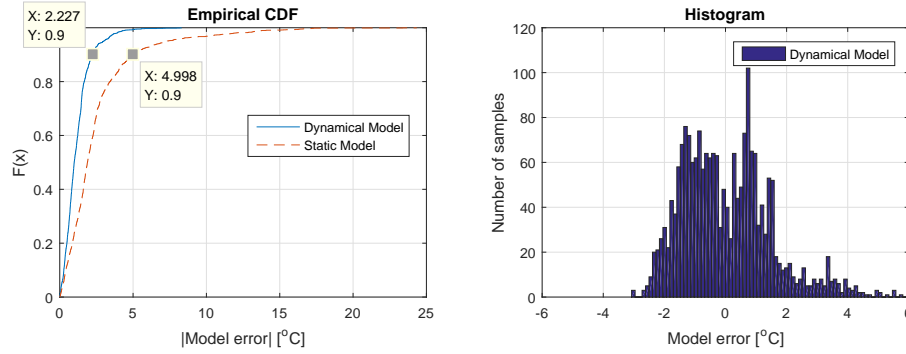


Figure 7: Module temperature error statistics.

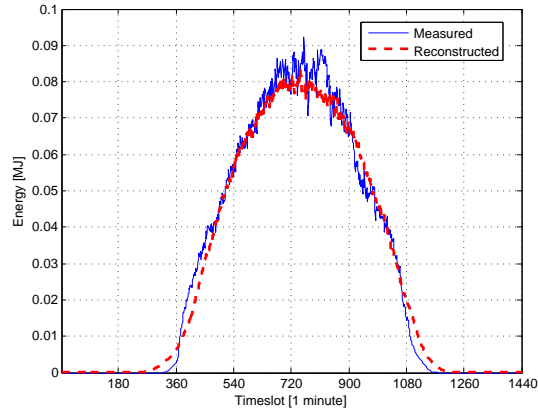


Figure 8: Comparison between empirical mean of the PV energy production profile per minute obtained based on the 28 measured realizations (blue line) and on the 220 reconstructed realizations (red line).

## 5 Performance assessment of the proposed energy management strategy

We shall next illustrate the performance that one can achieve with the proposed approach, based on the available dataset.

Recall that the sampling time is set equal to  $\Delta = 1$  minute. The storage system is initially charged at 50% SOC, i.e.,  $x_0 = 12.390$  MJ.

$\Gamma$  and  $\Theta$  in (8) are parameterized as follows

$$\Gamma = \gamma \bar{d} \quad \Theta = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \theta_1 & 0 & \ddots & \vdots \\ \theta_2 & \theta_1 & 0 & \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \theta_2 & \theta_1 & 0 \end{bmatrix}$$

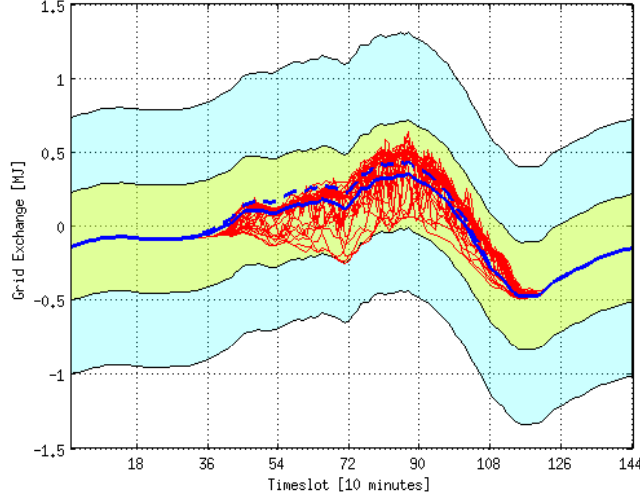


Figure 9:  $\epsilon$ -reachability region obtained with the designed energy management strategy using the storage (green inner tube centered on the solid blue line) together with the energy exchange profiles with the main grid corresponding to the 28 measured PV production realizations (red lines). The  $\epsilon$ -reachability region obtained without storage is the wider light blue outer tube centered on the dashed blue line.

in order to reduce the number of optimization variables in the scenario optimization program (16) to  $n = 5$ . The weighting parameters in (16) are set equal to  $\rho_g = \rho_u = 10^{-4}$ . The time window length is  $w = 10$ . Confidence, violation, and empirical violation parameters are given by  $\beta = 10^{-3}$ ,  $\epsilon = 0.15$ , and  $\eta = 0.035$ . Correspondingly, the bound in Proposition 1 is satisfied with  $N = 220$  reconstructed scenarios of PV energy production. The number of constraints to be removed is then 7. The expected value of the PV energy production  $\bar{\mathbf{d}}$  was estimated as the empirical mean of the 220 reconstructed production profiles (see Figure 8).

The obtained solution of the scenario optimization program (16) is:  $\gamma^* = 0.1290$ ,  $\theta_1^* = 0.3495$ ,  $\theta_2^* = 0.1720$ ,  $h_g^* = 0.3324$ ,  $h_u^* = 0.0455$ .

The green inner tube centered on the nominal profile  $\bar{\mathbf{g}}^* = \gamma^* \bar{\mathbf{d}} + \ell - \bar{\mathbf{d}}$  (blue solid line) with width  $2h_g^*$  in Figure 9 represents the  $\epsilon$ -reachability region containing all profiles of energy exchange with the main grid integrated over time windows of length  $w\Delta = 10$  minutes except for a set of probability smaller than  $\epsilon = 0.15$ , with high confidence ( $\geq 1 - \beta = 1 - 10^{-3}$ ). The 28 realizations of energy exchange with the main grid corresponding to the measured PV energy production profiles are plotted in the same figure, for validation purposes. For 4 of them, the constraints in the optimization program are violated. This fraction is actually smaller than  $\epsilon = 0.15$ . In Figure 9, we also plot the  $\epsilon$ -reachability region obtained by solving (16) with  $\gamma = \theta_1 = \theta_2 = 0$ , i.e, removing the storage system. In this case we obtain  $h_g^* = 0.87026$ , which maps into the light blue outer tube centered on the dashed blue line representing the reference nominal profile  $\bar{\mathbf{g}} = \ell - \bar{\mathbf{d}}$ .

In Figure 10, we plot the SOC evolution obtained with the optimal energy management strategy evaluated on the 28 validation PV energy production profiles (blue lines). The red thick line is the nominal charging profile  $\Gamma^* = \gamma^* \bar{\mathbf{d}}$ . The green dot-dashed lines represent the conservative bounds adopted in (14). They are violated for 4 of the profiles. The actual bounds

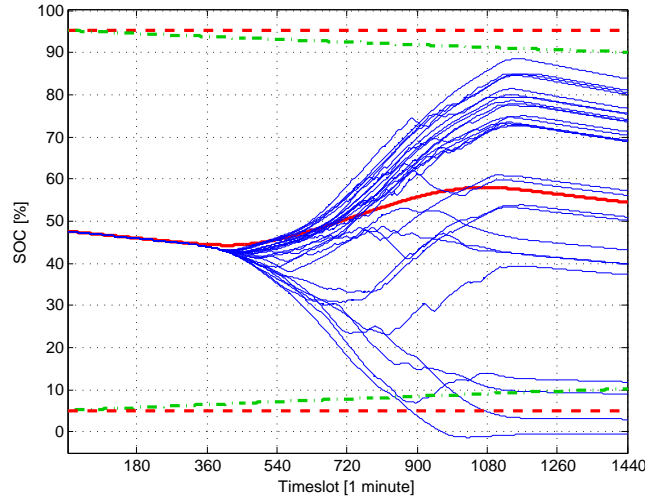


Figure 10: Energy exchange with the storage for the measured PV energy production profiles obtained with the designed energy management strategy (blue lines). The red solid line is its nominal component. The red dashed and green dot-dashed lines represent, respectively, the actual storage capacity constraints and their conservative approximation adopted in the optimization program.

(red dashed lines) are violated only for 2 of them. Note that the SOC evolutions reported in Figure 10 include also profiles where the battery capacity constraints are violated. This is possible because we are simulating the battery behavior to the purpose of verifying that the  $\epsilon$  violation probability is not exceeded. When implementing the proposed strategy on the real battery, the control input will be set to zero (no energy exchange with the battery) when the actual bounds on its capacity are violated.

Alternative strategies can be compared in terms of width of the  $\epsilon$ -reachability region for a given time window length  $w$ . Computation time is not an issue when the design is performed off-line as in our approach. It becomes a relevant criterion when adopting on-line design strategies like in model predictive control, [14], where constrained optimization is repeated at each time step, based on the current measurements, and only the first control action is applied at each step (receding horizon implementation).

## 6 Concluding remarks

In this paper, we propose a benchmark problem on energy management of a smart grid, where a storage system is used to compensate the fluctuations of solar PV energy production so as to provide a certified profile of the energy exchange with a main grid together with tolerance bounds. The main challenge in addressing the problem is that only a limited number of data is available for the energy management strategy data-driven design. As discussed in [2], this highly constraints the complexity of the energy management strategy, which in our design was given by a disturbance compensator parameterized by only three scalar optimization variables.



A possibility to enable the design of a more complex and better performing energy management strategy is to build a stochastic model of the PV energy production from the available data, and then use this model to extract further realizations. This opens up a new topic on the prediction of solar PV power production and requires further investigation.

Also, different control parameterizations and control strategies can be conceived. If a receding horizon implementation is adopted, then, improved predictions of the actual PV energy production profile could be obtained via filtering techniques and used for the on-line design of the control input, as suggested in [3].

## 7 Acknowledgments

This work is partially supported by the European Commission under the project UnCoVerCPS, grant number 643921.

## References

- [1] Teodoro Alamo, Roberto Tempo, Amalia Luque, and Daniel R. Ramirez. Randomized methods for design of uncertain systems: Sample complexity and sequential algorithms. *Automatica*, 52:160–172, 2015.
- [2] S. Bittanti, M. C. Campi, and M. Prandini. Adaptation and the effort needed to adapt. In *Proceedings of the 48th IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pages 5733–5737, Dec 2009.
- [3] C. Brocchini, A. Falsone, G. Manganini, O. Holub, and M. Prandini. A chance-constrained approach to the quantized control of a heat ventilation and air conditioning system with prioritized constraints. In *Proceedings of the 22nd International Symposium on Mathematical Theory of Networks and Systems (MTNS 2016), Minneapolis, Minnesota, USA*, pages 137–144, July 2016.
- [4] G. Calafiore and M.C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.
- [5] G. Calafiore and M.C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.
- [6] M.C. Campi and S. Garatti. A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality. *Journal of Optimization Theory and Applications*, 148(2):257–280, 2011.
- [7] M.C. Campi, S. Garatti, and M. Prandini. The scenario approach for systems and control design. *Annual Reviews in Control*, 33(2):149–157, 2009.
- [8] Haisheng Chen, Thang Ngoc Cong, Wei Yang, Chunqing Tan, Yongliang Li, and Yulong Ding. Progress in electrical energy storage system: A critical review. *Progress in Natural Science*, 19(3):291–312, 2009.
- [9] J.K. Copper, A.B. Sproul, and S. Jarnason. Photovoltaic (PV) performance modelling in the absence of onsite measured plane of array irradiance (POA) and module temperature. *Renewable Energy*, 86:760–769, 2016.
- [10] Luca Deori, Simone Garatti, and Maria Prandini. A relaxation technique to ensure feasibility in stochastic control with input and state constraints. *Systems & Control Letters*, 2016. submitted, available in arXiv:1610.06315 [math.OC].
- [11] Swapnil Dubey, Jatin Narotam Sarvaiya, and Bharath Seshadri. Temperature Dependent Photovoltaic (PV) Efficiency and Its Effect on PV Production in the World – A Review. *Energy Procedia*, 33:311–321, 2013.

- [12] Ludwig Joerissen, Juergen Garcke, Ch. Fabjan, and G. Tomazic. Possible use of vanadium redox-flow batteries for energy storage in small grids and stand-alone photovoltaic systems. *Journal of Power Sources*, 127(1):98–104, 2004.
- [13] S. Krauter and A. Preiss. Comparison of module temperature measurement methods. In *2009 34th IEEE Photovoltaic Specialists Conference (PVSC)*, pages 000333–000338, June 2009.
- [14] J.M. Maciejowski. *Predictive Control with Constraints*. Prentice Hall, 2002.
- [15] J.D. Maclay, J. Brouwer, and G.S. Samuelsen. Dynamic modeling of hybrid energy storage systems coupled to photovoltaic generation in residential applications. *Power Sources*, 163:916–925, 2007.
- [16] Maria Prandini, Simone Garatti, and Riccardo Vignali. Performance assessment and design of abstracted models for stochastic hybrid systems through a randomized approach. *Automatica*, 50(11):2852–2860, 2014.
- [17] A. Prékopa. *Stochastic Programming*. Kluwer, Boston, MA, 1995.
- [18] A. Prékopa. Probabilistic programming. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, London, UK, 2003. Elsevier.
- [19] E. Skoplaki and J.A. Palyvos. On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations. *Solar Energy*, 83(5):614–624, 2009.
- [20] J. K. Tonui and Y. Tripanagnostopoulos. Air-cooled PV/T solar collectors with low cost performance improvements. *Solar Energy*, 81:498–511, April 2007.