Optimal BESS Management for Peak Load Shaving and Battery Health Under Prediction Uncertainty

Lixin Li, *Student Member, IEEE*, Tim Kappler, *Student Member, IEEE*, Bernhard Schwarz, Nina Munzke, Xinliang Dai*, *Member, IEEE*, Veit Hagenmeyer, *Member, IEEE*, and Marc Hiller, *Member, IEEE*

***Abstract*—In modern power grids integrated with renewable energy sources (RESs), deploying battery energy storage systems (BESSs) is increasingly vital for mitigating power fluctuations. However, optimizing BESS operation remains challenging amidst uncertainties in both RES and load forecasting. This paper proposes a novel stochastic model predictive control (SMPC) framework for BESS operation, focusing on peak load shaving and battery health while addressing prediction uncertainties. The proposed framework employs a long-short-term memory (LSTM) neural network for forecasting and integrates a constraint-tightening technique into a stochastic optimization (SO) problem with a receding horizon. Based on the load profile of a company in Germany, the proposed framework achieves an additional reduction of 99 kW (5.8%) in peak grid take-out power compared with the traditional model predictive control (MPC) approach, demonstrating its advantage in addressing uncertainties.**

***Index Terms*—Battery energy storage system, Stochastic model predictive control, Prediction uncertainty, Peak shaving, Battery degradation, Long short-term memory**

I. Introduction

The intermittency of renewable energy sources (RESs) underscores the critical role of energy storage systems (ESSs) in modern power grids, facilitating supply-demand balancing and reliability enhancement. Consequently, there has been growing interest in studying various types of ESSs in recent years [1], [2]. Among these, Lithium-ion battery energy storage systems (BESSs) have emerged as market leaders due to their notable attributes, including high efficiency, high energy density, and low self-discharge rate [3]. Grid-connected BESSs have a wide range of applications, such as peak load shaving, photovoltaic (PV) smoothing, and energy arbitrage [3], [4].

Among the various applications, peak shaving is one of the most popular and profitable [1], [5]. Peak shaving is a load control method employed in the energy industry to stabilize the power drawn from the utility grid [6]. Peak demand in power systems leads to higher investments by grid companies, as the infrastructure of power transmission and distribution is designed based on the system's highest demand. Consequently, large customers incur a "peak demand cost" in countries such as Germany [7], determined by the maximal power demand drawn from the utility grid within a month or a year [8], [9]. In this context, a BESS can help these customers save power costs by storing energy during off-peak hours and releasing it during peak load hours. However, the effectiveness of these actions depends on an appropriate energy management scheme.

Multiple studies [8], [10]–[12] have investigated different operation strategies of BESSs for peak load shaving. In [10], a simple rule-based dispatch strategy was utilized. When the load power exceeds a predefined threshold (also known as the demand limit), the BESS discharges to meet the excess demand. Otherwise, it recharges immediately, maintaining a high state of charge (SoC) level to prepare for potential future discharge requests. However, this traditional approach has several disadvantages. Firstly, determining an appropriate threshold is contentious: an excessively high threshold limits the battery's full utilization, while an excessively low threshold hinders the battery from fully compensating for surpassing loads. Secondly, charging the battery as early as possible significantly increases the dwell time at high SoC, thereby accelerating its degradation [13], [14]. Lastly, keeping the BESS full for peak shaving can prevent the battery from multi-use applications such as storing surplus PV power and frequency regulation [4].

Lithium-ion battery degradation refers to the gradual loss of a battery's capacity and performance over time or with usage. It occurs due to various factors such as chemical reactions within the battery, cycling (charging and discharging), temperature exposure, and other environmental conditions [13], [15], [16]. Key stress factors that accelerate battery degradation include high and low operation temperatures, high and low SoCs, high depth of discharge (DoD), and high current or power. Therefore, optimizing the operation of BESSs is crucial not only for effectively performing tasks such as peak shaving but also for maintaining battery health [14], [17], [18].

To address peak shaving and battery health simultaneously, an optimal operation strategy using model predictive control (MPC) was developed in [11]. The approach is implemented as follows: First, load forecasting is conducted over a specific time horizon. Then, a multi-objective optimization problem is solved based on the forecasts to determine the optimal thresh-

*Corresponding Author: X. Dai (xinliang.dai@kit.edu).

This work was supported in part by the "ESiP" project under Grant 03EI6062B funded by the Federal Ministry for Economic Affairs and Climate Action (BMWK), in part by the ENSURE III Project with grant number 03SFK1F0-3 funded by Federal Ministry of Education and Research (BMBF), and in part by the Helmholtz Association under the program "Energy System Design".

L. Li, T. Kappler, B. Schwarz, N. Munzke, and M. Hiller are with the Institute of Electrical Engineering, Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany (e-mail: lixin.li, tim.kappler, bernhard.schwarz, nina.munzke, marc.hiller@kit.edu).

X. Dai was with the Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany. He is now with the Andlinger Center for Energy and the Environment, Princeton University, Princeton, NJ 08544 USA (xinliang.dai@princeton.edu).

V. Hagenmeyer is with the Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany (e-mail: veit.hagenmeyer@kit.edu).

old for peak shaving, as well as the optimal power distribution between the BESS and the grid within the given horizon. Afterward, only the first set of decisions is implemented. This process is repeated at subsequent time points. However, this framework does not account for forecasting errors.

In practical applications, typical PV prediction errors range from 6% to 12% [19], while load prediction errors can be even larger [20]. To address these uncertainties, two enhanced MPC methods are introduced in [21]: robust model predictive control (RMPC) and stochastic model predictive control (SMPC). RMPC integrates robust optimization with classic MPC to handle prediction uncertainty within known bounds; RMPC-based control strategies for microgrid operations are discussed in [22], [23]. However, RMPC algorithms often exhibit an overly conservative nature. Consequently, SMPC has gained significant attention for its approach to handling uncertainties as random variables with known probability distribution. Additionally, real-world uncertainties are generally probabilistic rather than strictly constrained by simple upper and lower bounds [24].

In SMPC, chance constraints are introduced into the optimization model, requiring the constraints to be satisfied with at least a pre-specified probability level. Thus, SMPC seeks a trade-off between fulfilling control objectives and ensuring probabilistic constraint satisfaction. Numerous SMPC approaches have been summarized by Mesbah [24], among which the scenario-based approach is widely adopted. Zhu et al. [25] and Kumar et al. [26] have implemented SMPC for optimal ESS dispatch by generating scenarios, which increases computational costs. Similarly, Ravichandran et al. [27] introduced scenario-based MPC to address uncertainties in load, generation, and electric vehicles within microgrids. Besides the scenario-based approach, analytical methods offer an alternative by encompassing numerous potential scenarios while reducing computational costs. For instance, Baker et al. [28] proposed a two-stage SMPC method for sizing and scheduling energy storage, employing an analytical reformulation of chance constraints. Additionally, Bazmohammadi et al. [29] presented a stochastic energy management strategy for interconnected microgrids, where chance constraints in the optimization problem were transformed into deterministic counterparts using a technique known as constraints tightening. However, SMPC techniques have not been applied to peak load shaving using BESS while considering battery degradation.

Bridging this gap, this paper proposes a novel energy management scheme based on SMPC for peak load shaving and optimizing battery degradation. The uncertainties addressed in this study stem from imperfect forecasting of RES and load. The primary contributions of this work are summarized as follows:

1) We propose an efficient SMPC framework for peak load shaving that accounts for battery health, incorporates time-varying prediction errors, maintains computational tractability, and ensures constraint satisfaction.
2) A comprehensive case study using real-world data validates the proposed framework. Compared to the classic MPC approach, our SMPC method achieves more effective peak load reduction.
3) We broaden the framework's applicability through extended simulations that analyze the impact of penalty parameters, confidence levels, and the length of look-ahead horizons.

The remainder of the paper is structured as follows. Section II represents the classic MPC framework for peak shaving and battery health. Section III introduces the SMPC formulation integrated with generation/load uncertainties. In Section IV, different schemes are compared through a case study using one-year data. The effect of confidence level on the effectiveness of the SMPC framework is analyzed in Section V. Finally, the paper is concluded in Section VI.

II. MPC-Based Energy Management Framework

In this section, we introduce an energy management framework for a large industrial or commercial consumer connected to the utility grid, incorporating both a PV plant and an on-site BESS, as illustrated in Fig. 1*. The focus is on optimizing power distribution by solving a multi-objective optimization problem over a receding horizon. The framework is structured into three main components: PV/load prediction, optimization, and real-time operation.

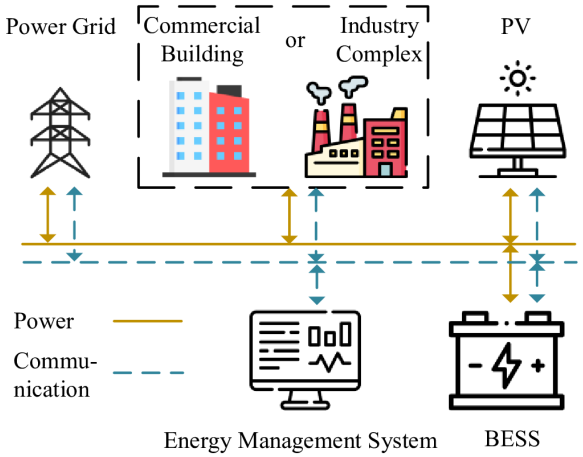

Fig. 1. Overview of large power consumers with on-site BESS, PV generation plant, and energy management system.

A. Forecasting

A long short-term memory (LSTM) network is a type of recurrent neural network designed to capture long-term dependencies in time-series data [30]. It is equipped with memory cells that can store information over extended periods, and is controlled by three key gates: the input gate $\mathbf{i}_t$, the forgetting gate $\mathbf{f}_t$, and the output gate $\mathbf{o}_t$. These gates manage the flow of information, allowing the network to retain relevant data while discarding less useful information. This architecture makes LSTMs particularly effective for tasks such as time-series forecasting, natural language processing, and sequence prediction.

*Flaticon contributors copyright icons in Fig. 1 and are available from https://www.flaticon.com/

For PV power forecasting, inputs to the LSTM network include irradiance, temperature, sun elevation, and the current month and hour. The network's output is given by (1),

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t^l] + \mathbf{b}_o) \tag{1}$$

where $\mathbf{x}_t^l$, $\mathbf{h}_{t-1}$, $\mathbf{W}_o$, and $\mathbf{b}_o$ are input vector, previous hidden state, weight matrix, and bias matrix, respectively.

While solar power prediction benefits from a comprehensive data foundation, achieving a satisfactory load prediction network is challenging due to the lack of long-term load data and sufficient input features. In this context, Bazmohammadi et al. [28] focused solely on errors in RES without considering load errors, whereas Dutta et al. [31] assumed errors follow a standard distribution. Inspired by Gross et al. [32], we use the average weekly load curve, after removing the anomalous weeks, as the prediction curve and cycle through it weekly.

B. Problem Formulation

In the following section, we introduce a MPC framework designed to manage the prediction mismatches during system operation.

1) BESS Modeling

The model of battery energy storage system (BESS) within the power nodes framework [33] in a time horizon $\mathcal{K} = \{0, 1, \cdots, N-1\}$ can be expressed as

$$x_{k+1} = x_k - \frac{\Delta T}{C^{\mathrm{s}}} P_k^{\mathrm{s}},\ \forall k \in \mathcal{K}, \tag{2a}$$

$$P_k^{\mathrm{s}} = \eta^{-1} P_k^{\mathrm{s,dhg}} - \eta P_k^{\mathrm{s,chg}},\ \forall k \in \mathcal{K}, \tag{2b}$$

where the parameters ΔT, C^{s}, and η denote the time interval, the energy capacity, and the efficiency of the BESS, respectively. The variable P_k^{s} represents the power of the BESS at time k, consisting of the discharge power $P_k^{\mathrm{s,dhg}}$ and the charge power $P_k^{\mathrm{s,chg}}$ at the bus side. The variable x_k denotes the battery's energy state at time k, which is approximated by SoC in the following. The energy stored in the BESS must abide by lower and upper storage bounds:

$$\underline{x} \leq x_{k+1} \leq \overline{x},\ \forall k \in \mathcal{K}, \tag{3}$$

where $\bar{\cdot}$ and $\underline{\cdot}$ denote the corresponding upper and lower system limits. Approximating state of energy (SoE) using SoC is a common practice in system-level applications [34]. Still, it is not applicable in high and low SoC regions, as small SoC variations in these regions can induce significant voltage fluctuations, leading to substantial approximation errors and increasing the risk of overcharging or overdischarging. Moreover, extreme SoCs accelerate battery aging. Therefore, proper selection of SoC bounds is crucial to both the SoE estimation validity and battery health. Limits about charging and discharging are formulated as

$$0 \leq P_k^{\mathrm{s,dhg}} \leq \overline{P}^s,\ \forall k \in \mathcal{K}, \tag{4a}$$

$$0 \leq P_k^{\mathrm{s,chg}} \leq \overline{P}^s,\ \forall k \in \mathcal{K}. \tag{4b}$$

2) Grid Connection Modeling

The power balance can be formulated as

$$P_k^{\mathrm{g}} = P_k^{\mathrm{s,chg}} - P_k^{\mathrm{s,dhg}} + P_k^{\mathrm{d}} - P_k^{\mathrm{pv}},\ \forall k \in \mathcal{K}, \tag{5a}$$

$$P_k^{\mathrm{g}} = P_k^{\mathrm{from,g}} - P_k^{\mathrm{to,g}},\ \forall k \in \mathcal{K}, \tag{5b}$$

where P_k^{g}, P_k^{d}, P_k^{pv} denote the grid side power, the demand power, and the PV power at time instant k, respectively. The variable P_k^{g} is classified into extraction power $P_k^{\mathrm{from,g}}$ and feed-in power $P_k^{\mathrm{to,g}}$ due to the price differential between buying and selling electricity. After each optimization step, the results will be checked to ensure that extraction and feed-in power do not coincide.

The extraction and feed-in power are constrained as follows:

$$0 \leq P_k^{\mathrm{from,g}} \leq \theta^{\mathrm{init}} + \gamma,\ \forall k \in \mathcal{K}, \tag{6a}$$

$$0 \leq \gamma, \tag{6b}$$

$$0 \leq P_k^{\mathrm{to,g}},\ \forall k \in \mathcal{K}, \tag{6c}$$

where the pre-set parameter θ^{init} represents the peak shaving threshold, and the slack variable γ is introduced to account for threshold violations.

Remark 1 (**Complementarity constraints in BESS**). *To prevent the BESS from simultaneously charging and discharging, and to avoid concurrent power injection into and extraction from the grid, a standard approach is to introduce complementarity constraints:*

$$0 \leq P_k^{s,dhg} \perp P_k^{s,chg} \geq 0,\ \forall k \in \mathcal{K}, \tag{7a}$$

$$0 \leq P_k^{from,g} \perp P_k^{to,g} \geq 0,\ \forall k \in \mathcal{K}, \tag{7b}$$

where $0 \leq x \perp y \geq 0$ *enforces* $x, y \geq 0$ *and* $x \cdot y = 0$. *However, explicitly enforcing* (7) *introduces nonsmoothness into the system models, increasing the problem complexity dramatically.*

3) Objective

The proposed energy management framework addresses multiple objectives, with the primary focus on reducing electricity costs, which include both energy and peak demand charges. Energy costs are calculated as the cost of purchasing electricity from the grid, minus the revenue earned from feeding PV power into the grid. Additionally, peak demand costs are proportional to the highest grid power during the billing period [8]. The cost function J_{grid} is formulated as

$$J_{\mathrm{grid}} = \sum_{k \in \mathcal{K}} \Delta T (c^{\mathrm{buy}} P_k^{\mathrm{from,g}} - c^{\mathrm{sell}} P_k^{\mathrm{to,g}}) + c^{\mathrm{peak}} (\theta^{\mathrm{init}} + \gamma), \tag{8}$$

where c^{buy}, c^{sell}, c^{peak} denote the electricity buying and selling prices in €/kWh, and the prorated peak demand price in €/kW, respectively.

Note that the complementarity constraints for power exchanges (7b) are inherently satisfied under the cost-minimization objective in (8):

$$c^{\mathrm{buy}} P_k^{\mathrm{from,g}} - c^{\mathrm{sell}} P_k^{\mathrm{to,g}}$$

assuming the buying price strictly exceeds the selling price:

$$c^{\text{buy}} > c^{\text{sell}}.$$

If $c^{\text{buy}} > c^{\text{sell}}$, simultaneous buying and selling would result in a financial loss, e.g., buying at a higher price while selling at a lower price, which violates cost minimization. This logic remains valid for negative electricity prices, including *negative selling prices* ($c^{\text{buy}} > 0 > c^{\text{sell}}$) and *negative buying/selling prices* ($0 \geq c^{\text{buy}} > c^{\text{sell}}$). In both cases, the inequality $c^{\text{buy}} > c^{\text{sell}}$ ensures that simultaneous grid injection and extraction remains suboptimal.

Furthermore, the framework aims to mitigate battery degradation costs. Due to the complexity of the lithium-ion battery degradation process, these costs are typically assessed offline. Studies [15], [16] on the degradation modeling of various lithium-ion batteries suggest that high average SoC, high temperatures, high current rates, and deep cycle depths significantly affect calendar and cycle aging. In [11], [35], quadratic terms related to battery power and SoC were incorporated into the objective function to capture battery degradation costs. Building on this approach, we introduce a penalty function

$$J_{\text{pen}} = \sum_{k\in\mathcal{K}} \rho_1\, x_{k+1}^2 + \rho_2 (P_k^{\text{s,chg}})^2 + \rho_2 (P_k^{\text{s,dhg}})^2, \qquad (9)$$

where ρ_1 and ρ_2 denote the penalty parameters associated with the SoC level and the SoC change (number of cycles) of the BESS, respectively.

The L2 penalty term on the BESS charging and discharging power in (9) serves as a computationally tractable alternative to address the complexity introduced by the complementarity constraints (7a), as discussed in Remark 1. By penalizing the quadratic terms $(P_k^{\text{s,chg}})^2+(P_k^{\text{s,dhg}})^2$, this formulation indirectly discourages nonzero values of the product $P_k^{\text{s,dhg}} \cdot P_k^{\text{s,chg}}$, $\forall k \in \mathcal{K}$ towards zero. Combined with the physical power limits of the BESS (4), this approach ensures smoothness in the optimization problem, enabling computationally tractable solutions while avoiding simultaneous charging and discharging.

Hence, the optimization problem over the horizon $\mathcal{K}$ is formulated as

$$\begin{aligned} \min_{\mathbf{P}^{\text{s}},\mathbf{P}^{\text{g}},\gamma}\ & J_{\text{grid}} + J_{\text{pen}}, \\ \text{s.t. }\ & \text{BESS model: (2), (3), (4),} \\ & \text{Grid connection model: (5), (6).} \end{aligned} \qquad (10)$$

where $\mathbf{P}^{\text{s}}$, $\mathbf{P}^{\text{g}}$ stand for stacked power of the BESS and the grid across the period $\mathcal{K}$. In this multi-objective optimization problem, the slack variable γ in (6a) relaxes the constraint on the maximum grid extraction power. This allows trade-offs between economic performance and feasibility, ensuring the problem remains solvable even under tight physical constraints. Simultaneously, the soft penalty term J_{pen} penalizes high SoC and excessive charging/discharging power, steering the solution toward more desirable operation strategies within the feasible region.

The quadratic programming (QP) problem can be solved using off-the-shelf optimizers such as *Gurobi* [36].

Algorithm 1 MPC Operation Framework

Input: $\overline{P}^{\text{s}}$, $\underline{x}$, $\overline{x}$, θ^{init}
Output: P_t^{s}, x_{t+1}, P_t^{g}, γ_t
Initialisation : $x_0 = x^{\text{init}}$

```
for t = 0 → t_end do
    Update forecasts P^pv_{k|t}, P^d_{k|t} and P^net_{k|t} for all k ∈ K.
    Solve (10) to obtain P^s_{k|t}, P^g_{k|t}, and γ_t for all k ∈ K.
    if P^net_{t-1} ≤ θ^init then
        Implement storage power following mode: P^s_{0|t}.
    else
        Implement peak shaving mode: P^g_{0|t}.
    end if
    Update x_{t+1} of the BESS.
    t ← t + 1.
end for
```

C. Operation In Receding Horizon

At each sampling instant t during real-time operation, the process begins with load and PV predictions over a horizon length N. Next, during the optimization stage, power decisions $(P_{k|t}^{\text{s}}, P_{k|t}^{\text{g}})^T$ are calculated for all $k \in \mathcal{K}$, where the subscript $k|t$ denotes the time interval $[t + k\Delta T, t + (k + 1)\Delta T)$. Only the first decision $(P_{0|t}^{\text{s}}, P_{0|t}^{\text{g}})^T$ could be implemented. During the implementation stage, the system in Fig. 1 operates in two distinct modes. If the netload at the previous time step, $P_{t-1}^{\text{net}} = P_{t-1}^{\text{d}} - P_{t-1}^{\text{pv}}$, is below the initial peak shaving threshold θ^{init}, the system operates in storage power-following mode, implementing $P_{0|t}^{\text{s}}$. Otherwise, the system works in peak-shaving mode, implementing $P_{0|t}^{\text{g}}$. Switching between the two modes, the BESS is either on standby or actively involved in peak shaving. This process is iteratively executed in a receding horizon manner, utilizing the system's end state at time t as the initial condition for the subsequent time point $t+1$. The overall operational strategy is detailed in Algorithm 1 (Alg. 1).

In peak-shaving mode, the BESS manages node power balance by compensating for power deviations caused by prediction uncertainties. However, if the power or energy required for compensation exceeds the physical limits of the BESS, as specified in (3) and (4), the grid-side power P_t^{g} must make up the excess to maintain power balance. In this case, P_t^{g} may exceed the optimally corrected threshold $\theta^{\text{init}} + \gamma_t$, potentially increasing peak demand costs. To mitigate the impact of prediction errors on the success of peak shaving, it is beneficial to quantitatively incorporate these errors into the optimization process. This approach is detailed in the stochastic optimization (SO) problem outlined in Section III.

III. SMPC-Based Energy Management Framework

In this section, we first formulate an SO problem with chance constraints for optimal power distribution. Next, we apply the constraint-tightening technique to address prediction uncertainties and reformulate certain constraints. This reformulated optimization model is then solved and implemented using a receding horizon approach.

A. Forecasting Error

After training the LSTM network for PV generation as described in Section II-A, its performance is evaluated during the validation phase. A histogram of PV prediction errors is constructed, and a Gaussian distribution is fitted to this error distribution. It is assumed that in subsequent simulations, the LSTM's prediction errors will continue to follow this Gaussian distribution. For load prediction errors, a time-varying Gaussian distribution is fitted, which will be detailed in Section IV-A.

B. Stochastic Optimization Formulation

For notational simplification, the netload P_k^{net} is defined as the difference between load and PV generation, i.e., $P_k^{\text{net}} = P_k^{\text{d}} - P_k^{\text{pv}}$. The forecast errors of load and PV generation $\varepsilon_k^{\text{load}}$ and $\varepsilon_k^{\text{pv}}$ are assumed to be independent and normally distributed. Consequently, the netload forecast error $\varepsilon_k^{\text{net}} = \varepsilon_k^{\text{load}} - \varepsilon_k^{\text{pv}}$ also follow a Gaussian distribution, i.e., $\varepsilon_k^{\text{net}} \sim \mathcal{N}(\mu_k^{\text{net}}, (\sigma_k^{\text{net}})^2)$, where μ_k^{net}, σ_k^{net} represent the mean and standard deviation of the netload forecast error at time step k, respectively.

We propose a real-time operation strategy that uses BESS to compensate for the netload mismatch. Given that the charging/discharging conversion losses related to the netload error are small enough to be neglected [28], the actual state of the BESS $\tilde{x}_{k+1}$ evolves according to:

$$\tilde{x}_{k+1} = x_{k+1} - \frac{\Delta T}{C^{\text{s}}}\varepsilon_k^{\text{net}}, \tag{11}$$

where x_{k+1} and $\frac{\Delta T}{C^{\text{s}}}\varepsilon_k^{\text{net}}$ denote the nominal energy state (2a) and the stochastic part that needs to be compensated by the BESS. This decomposition explicitly separates the deterministic nominal trajectory from the stochastic forecast error. This formulation is practical in uncertainty-aware modeling, as it allows the nominal state x_{k+1} to serve as the decision variable in the optimization problem, while the impact of uncertainty is handled by tightening the constraints. This separation further facilitates the reformulation of chance constraints (13) into their deterministic equivalents (16).

Given that the netload forecast error $\varepsilon_k^{\text{net}}$ is normally distributed, $\tilde{x}_{k+1}$ also follows a normal distribution, whose mean and variance can be formulated as

$$\mu_{k+1} = x_{k+1} - \frac{\Delta T}{C^{\text{s}}}\mu_k^{\text{net}}, \tag{12a}$$

$$\sigma_{k+1} = \frac{\Delta T}{C^{\text{s}}}\sigma_k^{\text{net}}. \tag{12b}$$

A chance-constrained optimization approach [28], [29] can be employed to impose probabilistic constraints on the BESS actual energy state:

$$\mathbb{P}\{\tilde{x}_{k+1} \geq \underline{x}\} \geq \beta, \tag{13a}$$

$$\mathbb{P}\{\tilde{x}_{k+1} \leq \overline{x}\} \geq \beta, \tag{13b}$$

which means that the capacity limits should be fulfilled with at least a pre-defined confidence level β. The left side of the chance constraints (13) can be transformed into the cumulative distribution function (CDF) of the corresponding Gaussian distribution evaluated at 0, which can be expressed as

$$\mathbb{P}\{\underline{x} - \tilde{x}_{k+1} \leq 0\} = \Phi\left(\frac{0 - (\underline{x} - \mu_{k+1})}{\sigma_{k+1}}\right) \geq \beta, \tag{14a}$$

$$\mathbb{P}\{\tilde{x}_{k+1} - \overline{x} \leq 0\} = \Phi\left(\frac{0 - (\mu_{k+1} - \overline{x})}{\sigma_{k+1}}\right) \geq \beta, \tag{14b}$$

where $\Phi(\cdot)$ denotes the CDF of the standard normal distribution. Taking the inverse CDF of both sides yields

$$\mu_{k+1} - \underline{x} \geq \sigma_{k+1}\Phi^{-1}(\beta), \tag{15a}$$

$$\overline{x} - \mu_{k+1} \geq \sigma_{k+1}\Phi^{-1}(\beta), \tag{15b}$$

where $\Phi^{-1}(\cdot)$ denotes the inverse CDF of the standard normal distribution. Substituting (12) into (15) yields

$$x_{k+1} \geq \underline{x} + \frac{\Delta T}{C^{\text{s}}}\sigma_k^{\text{net}}\Phi^{-1}(\beta) + \frac{\Delta T}{C^{\text{s}}}\mu_k^{\text{net}}, \tag{16a}$$

$$x_{k+1} \leq \overline{x} - \frac{\Delta T}{C^{\text{s}}}\sigma_k^{\text{net}}\Phi^{-1}(\beta) + \frac{\Delta T}{C^{\text{s}}}\mu_k^{\text{net}}. \tag{16b}$$

Therefore, chance constraints (13) can be replaced by their deterministic counterparts (16) using the constraints-tightening technique. The power constraints can be addressed similarly. However, this part will not be shown since this paper does not apply tightening to the power constraints.

Furthermore, when solving an optimization problem in a time series, errors accumulate over time, representing the propagation of forecast errors. Assuming that the prediction error is independent at each time point, it can simply accumulate over time [28]. Therefore, μ_k^{net} and σ_k^{net} in (16) are replaced with $\mu_k^{\text{net, acc}}$ and $\sigma_k^{\text{net, acc}}$, expressed as

$$\mu_k^{\text{net, acc}} = \sum_{i=0}^{k}(\mu_i^{\text{net}}), \tag{17a}$$

$$\sigma_k^{\text{net, acc}} = \sqrt{\sum_{i=0}^{k}(\sigma_i^{\text{net}})^2}. \tag{17b}$$

In peak shaving applications, BESS plays a crucial role during the discharge process. Therefore, reserving discharge capacity is more practical. The energy state constraint, specifically for peak shaving, can be represented as

$$\underline{x} + F(\beta, k) \leq x_{k+1} \leq \overline{x}, \tag{18a}$$

$$F(\beta, k) = \frac{\Delta T}{C^{\text{s}}}\sigma_k^{\text{net, acc}}\Phi^{-1}(\beta) + \frac{\Delta T}{C^{\text{s}}}\mu_k^{\text{net, acc}}. \tag{18b}$$

Another caution must be taken regarding factors such as long optimization time k and high confidence level β. These factors can lead to over-constriction, potentially causing the lower bound to exceed the upper bound and rendering the problem infeasible. To address this issue, we introduce a fading factor λ into (18), representing the fact that the impact of prediction errors diminishes with time. Thus, the complete

Algorithm 2 SMPC Operation Framework

Input: $\overline{P}^{\mathrm{s}}$, $\underline{x}$, $\overline{x}$, θ^{init}, β, λ
Output: P_t^{s}, x_{t+1}, P_t^{g}, γ_t
Initialisation : $x_0 = x^{\mathrm{init}}$

for $t = 0 \rightarrow t_{\mathrm{end}}$ **do**
  Update forecasts $P_{k|t}^{\mathrm{pv}}$, $P_{k|t}^{\mathrm{d}}$, $P_{k|t}^{\mathrm{net}}$, and forecast error distributions $\mu_{k|t}^{\mathrm{net}}$, $\sigma_{k|t}^{\mathrm{net}}$ for all $k \in \mathcal{K}$.
  Solve (20) to obtain $P_{k|t}^{\mathrm{s}}$, $P_{k|t}^{\mathrm{g}}$, and γ_t for all $k \in \mathcal{K}$.
  if $P_{t-1}^{\mathrm{net}} \leq \theta^{\mathrm{init}}$ **then**
    Implement storage power following mode: $P_{0|t}^{\mathrm{s}}$.
  else
    Implement peak shaving mode: $P_{0|t}^{\mathrm{g}}$.
  end if
  Update x_{t+1} of the BESS.
  $t \leftarrow t + 1$.
end for

energy state constraints, incorporating the fading factor, can be formulated as

$$\underline{x} + F(\beta, k, \lambda) \leq x_{k+1} \leq \overline{x},\ \forall k \in \mathcal{K}, \tag{19a}$$

$$F(\beta, k, \lambda) = \lambda^k \frac{\Delta T}{C^{\mathrm{s}}} (\sigma_k^{\text{net, acc}} \Phi^{-1}(\beta) + \mu_k^{\text{net, acc}}), \tag{19b}$$

$$\mu_k^{\text{net, acc}} = \sum_{i=0}^{k} (\mu_i^{\mathrm{net}}),\ \sigma_k^{\text{net, acc}} = \sqrt{\sum_{i=0}^{k} (\sigma_i^{\mathrm{net}})^2}. \tag{19c}$$

The SO problem with chance constraints is transformed into a deterministic optimization problem and the formulation of the resulting problem is illustrated as

$$\begin{aligned} \min_{\mathbf{P}^{\mathrm{s}}, \mathbf{P}^{\mathrm{g}}, \gamma}\ & J_{\mathrm{grid}} + J_{\mathrm{pen}}, \\ \text{s.t. BESS model: } & (2), (19), (4), \\ \text{Grid connection model: } & (5), (6). \end{aligned} \tag{20}$$

The proposed SMPC-based BESS operation framework is detailed in Algorithm 2 (Alg. 2). By dynamically reserving an amount of energy, the BESS can more effectively compensate for errors. However, as long as the error distribution is unbounded, the confidence level β cannot reach 1, meaning there will always be a residual probability that some errors cannot be fully compensated. Note that the problem (20) is a convex problem with affine constraints, so its computational footprint scales well with horizon length and model size. For more complex or realistic models, fast real-time optimal control techniques can be adopted to preserve tractability without compromising the SMPC design.

IV. Case Study

This section provides a case study using load data from a factory in Germany and PV power data from the campus north of Karlsruhe Institute of Technology (KIT). A comparison between different operation frameworks is conducted, evaluating the advantages and disadvantages of these schemes under prediction errors.

TABLE I
Summary Simulation Setup

Objects	Parameters	Values	Units
Load profile	Yearly peak	1793.5	kW
	Yearly energy consumption	6310	MWh
	Duration factor	3518	Hours
	Electricity buying price	0.15	€/kWh
	Electricity feed-in price	0.06	€/kWh
	Demand unit price[1]	216	€/(kW·a)
BESS	Size	500 / 500	kWh / kW
	Power component price	250	€/kW
	Energy component price	250	€/kWh
	Total costs[2]	250	k€
	SoC limits	[0.1, 0.9]	-
	Roundtrip efficiency	0.8	-
PV	Size	850	kWp
	Geographic location	Karlsruhe, Germany	
	Longitude and latitude	49.1° N, 8.44° E	
	Orientation and Inclination	180° (South), 30°	
Algorithm	Simulation time span	1	Year
	Horizon N	96	-
	Sampling time ΔT	15	Minutes
	Initial threshold θ^{init}	1350	kW
	Initial SoC x^{init}	0.1	-
	Penalty factors ρ_1, ρ_2	5, $\frac{3}{500^2}$	-
	Fading factor λ	$\frac{1}{1.02}$	-

[1] Demand unit price is sourced from [9].
[2] Battery price data and total cost calculations are from [41]–[43]. Total costs = Energy costs + Power costs.

A. Simulation Setup

We select company 46 from [5] as our case study, with data available from [37]. This load profile exhibits an annual energy consumption of $6,310$ MWh with a peak power of $1,793.5$ kW. Peak loads exceeding $1,000$ kW occur regularly during working hours on each working day. Based on the electricity operator's price list [9], we set the demand unit price at $216\,€/(\mathrm{kW} \cdot \mathrm{a})$. Besides, the company is equipped with a 500 kWh lithium-ion BESS. The cost of power components and energy components for the BESS is set at 275 \$/kW and 275 \$/kWh, which translates into 250 €/kW and 250 €/kWh when converted at an exchange rate of $1.11:1$. Furthermore, reflecting recent trends in renewable energy generation in Germany as of 2023, a penetration rate of 50% is considered appropriate, corresponding to an 850 kWp PV installation for the case study. The actual PV values are sourced from the "Solarpark" project [38] at KIT. Consequently, the studied company is equipped with an 850 kWp / 500 kWh PV-BESS system, with additional parameters provided in Table I. It should be noted that this study focuses on developing an optimal energy management framework, so the optimization of system sizing is not discussed further. Regarding the upper and lower SoC limits, prior studies [7], [8], [11], [39], [40] have adopted different SoC limits. In our study, overly restrictive bounds may generate conflicting constraints in (19a), making the optimization problem (20) infeasible. To ensure both model solvability and safe battery operation, we make a trade-off by setting the hard constraints on SoC between 0.1 and 0.9. The parameters ρ_1 and ρ_2 are optimized through a grid search approach, as detailed in Section V-A.

For load prediction, we use the average of actual weekly

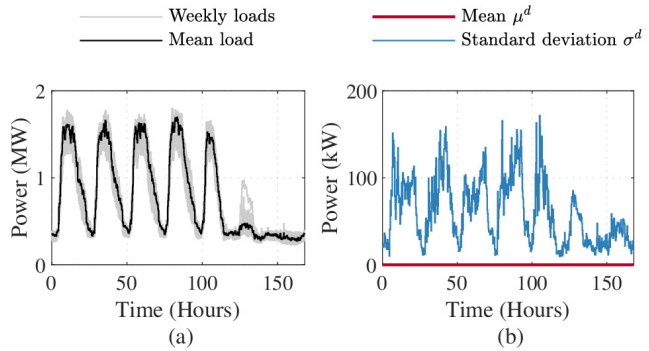

Fig. 2. Assessment of load prediction. (a) Mean of weekly load profiles as a prediction profile. (b) Time-varying Gaussian fit of load prediction errors.

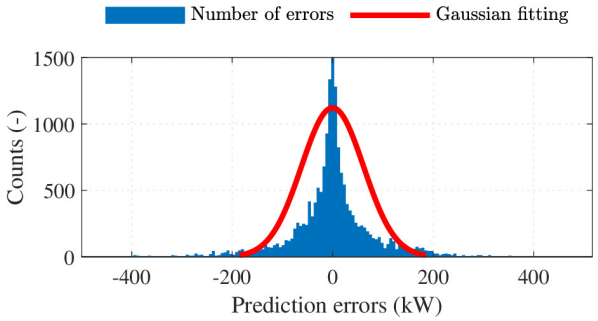

Fig. 3. Histogram of PV prediction errors and the corresponding Gaussian fit during the validation phase.

curves over one year as the weekly load prediction, represented by the solid black line in Fig. 2(a). A time-varying Gaussian distribution is fitted to the load prediction error ε^{d}, as shown in Fig. 2(b). Both load prediction and prediction error exhibit weekly cyclic patterns. The PV prediction is generated using a trained LSTM neural network. Given that the PV prediction error does not exhibit obvious periodicity, a Gaussian distribution is fitted to the entire validation dataset, as depicted in Fig. 3, resulting in $\varepsilon^{\mathrm{pv}} \sim \mathcal{N}(0, 185^2)$. However, there is a noticeable discrepancy between the Gaussian assumption and the actual distribution. This discrepancy is primarily due to the inherent characteristics of the prediction errors, such as the low volatility during nighttime, which leads to a concentration of errors around 0, and the presence of extreme errors caused by rapid weather changes, resulting in a heavier tail in the distribution compared to the Gaussian model. Nevertheless, the Gaussian distribution remains a practical and suitable choice, as it provides a reasonable approximation and facilitates the integration of prediction error analysis into the optimization model.

By combining load and PV power, the netload power error $\varepsilon^{\mathrm{net}}$ follows a new weekly time-varying Gaussian distribution, with a maximum standard deviation of approximately 20% of the peak netload.

B. Performance Metrics

To better evaluate and compare different management schemes, we establish the following performance metrics. The first two metrics assess peak shaving and the last three evaluate battery aging.

1) Peak reduction rate: This metric quantifies the proportion of reduced power demand as a percentage of the original maximum power.
2) Peak shaving success rate: This metric is defined as the ratio of the number of times the peak load is successfully shaved under the threshold to the total number of peak shaving attempts.
3) Battery throughput: This metric refers to the total amount of energy that a battery delivers over a period, characterizing the cyclic aging of the BESS.
4) Average SoC: This metric is the average energy state of the BESS over a period, characterizing calendar aging.
5) Capacity loss: Xu's battery degradation model [16] is used to evaluate the capacity loss of the BESS after operation. A detailed description of the degradation model is provided in the Appendix.

C. Simulation Three Scenarios

Three scenarios in Table II will be simulated. In Scenario 0 (S0), Alg. 1 is implemented under perfect prediction, and the simulation results are used as a benchmark. In Scenario 1 (S1) and Scenario 2 (S2), Alg. 1 and Alg. 2 are implemented under imperfect prediction (Subsection IV-A), respectively, with Alg. 2 operating at a confidence level of 0.99. The simulation spans nearly 1 year (356 days), with a rolling prediction window of 24 hours, denoted by $N = 96$, and a time interval $\Delta T = 15$ minutes. Therefore, the optimization tasks in each scenario entail a total of $34,176$ subproblems.

TABLE II
SETUP OF THE PREDICTION AND MANAGEMENT FRAMEWORK IN THE THREE SCENARIOS

Senario	S0	S1	S2
Prediction	Perfect	Imperfect	Imperfect
Prediction error	None	Unconsidered	Considered
Framework	Alg. 1	Alg. 1	Alg. 2
Confidence level	None	None	0.99

An interesting observation is that on sunny days, the overlap between peak load and peak PV generation reduces the peak netload, thereby decreasing the peak shaving demand. Consequently, for large customers with PV generation, winter, cloudy, and rainy days will present a higher peak shaving demand than sunny days.

For a detailed analysis, we select July 24, 2023, the day with the highest grid-side power occurrence in both S1 and S2 during the simulation. This day's predicted and actual curves are shown in Fig. 4. Three load peaks can be observed, with a maximum of $1,772$ kW at $10:00$. This day is notable because there are showers in the morning, followed by strong sunlight in the afternoon, leading to fluctuations in PV generation that deviate significantly from the predicted PV curve. Combining the load and PV curves, the load peak in the afternoon is suppressed by the PV, resulting in only two peaks in the

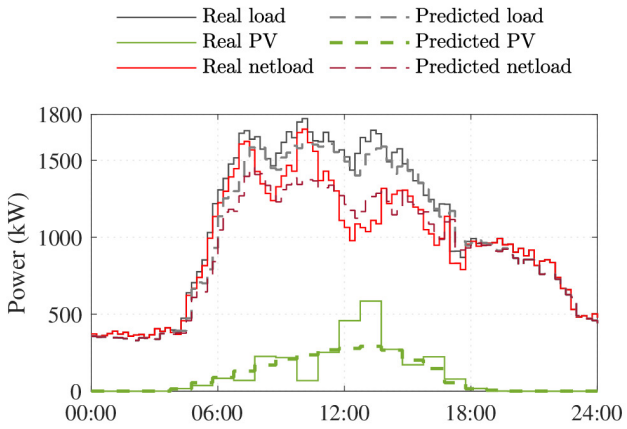

Fig. 4. Real and predicted profiles on the simulation day.

netload curve. The second peak at $10:00$ exhibits a maximum percentage prediction error of 24%.

1) Scenario 0

The power distribution and the BESS operation are illustrated in Figs. 5(a) and 6(a). The predicted netload profile in Fig. 5(a) matches the true netload profile (solid red line in Fig. 4). The operation of the BESS can be delineated into three distinct phases, divided by two dotted lines. Between $00:00$ and $06:30$, Alg. 1 gradually charges the BESS from a SoC of 0.1 to 0.9 in preparation for peak shaving, defining this period as the preparation phase. From $06:30$ to $11:30$, Alg. 1 actively manages the battery's charging and discharging to maintain the grid take-out below the dynamic threshold. By doing so, both spikes in the netload are successfully removed, marking this period as the peak shaving phase. Finally, from $11:30$ to $24:00$, when the netload is below the threshold and the next predicted spikes are distant, Alg. 1 keeps the BESS at a relatively low but non-zero level to avoid deep discharge and reduce long-term battery degradation. This period can be referred to as the standby phase.

2) Scenario 1

Figs. 5(b) and 6(b) illustrate the power distribution and the BESS operation. In the preparation phase, Alg. 1 charges the BESS only to a SoC of 0.16, as illustrated in Fig. 6(b). During the peak shaving phase, the BESS is discharged to its lower limit within the first 15 minutes of the peaks' appearance, losing its capacity for further peak shaving. Hence, the grid take-out power reaches a maximum of $1,703$ kW at $10:15$, as depicted in Fig. 5(b). The standby phase remains the same as in S0. Under Alg. 1, no peaks are suppressed because the imperfect prediction fails to capture the peak characteristics of the real netload, leading to insufficient BESS capacity preparation for peak shaving.

3) Scenario 2

Figs. 5(c) and 6(c) display the simulation results in S2. In the preparation phase, Alg. 2 charges the BESS to a SoC of 0.56 as shown in Fig. 6(c), providing more energy than in S1. During the peak shaving phase, the BESS effectively manages two peaks, keeping the grid take-out power below the dynamic threshold, represented by the red dashed line in Figs. 5(c). In the standby phase, the BESS is gradually recharged to around 0.37. Under imperfect prediction, Alg. 2's success in peak shaving is attributed to its conservative approach. Specifically, Alg. 2 maintains the BESS at an intermediate energy state during both the preparation and standby phases and sets the dynamic threshold higher than that of Alg. 1. These strategies enable the BESS to accommodate prediction errors better.

The simulation results for this day are summarized in Table III. In S0, both peak shaving attempts were successful, resulting in a 18.4% reduction of the maximum grid take-out while maintaining a low average SoC. In S1, both peak shaving attempts failed due to greatly underestimated peaks, leading to minimal manipulation of the BESS by Alg. 1. In S2, both peak shaving attempts succeeded, achieving a peak reduction rate of 5.8%. Due to Alg. 2's conservative nature compared to Alg. 1, the average SoC in S2 was greater than in S1. The comparison of a single day indicates that the SMPC scheme, despite the percentage prediction error varying up to 24%, shaved more peaks but resulted in a higher average SoC compared to the MPC scheme. A more comprehensive comparison should be conducted over a one-year simulation.

TABLE III
ONE-DAY SIMULATION OF THREE SCENARIOS

Metrics	S0	S1	S2
Peak shaving success rate (%)	100	0	100
Peak reduction (kW)	313	0	99
Peak reduction rate[1] (%)	18.4	0	5.8
BESS average SoC (-)	0.3	0.13	0.44

[1] The baseline is the maximum netload for the day, i.e., 1703 kW.

D. Simulation Conclusions

The one-year simulation results are summarized in Table IV and Fig. 7. Since the energy costs are nearly identical across all three scenarios, the comparison will focus exclusively on the demand costs and the battery degradation costs.

Regarding peak shaving, Alg. 1 in S0 reduces the maximum of netload by 14.7% and saves € 54.2k, demonstrating as the benchmark. This is followed by Alg. 2 in S2, which achieves a 5.8% peak power reduction and € 21.4k in savings, while Alg. 1 in S1 performs the worst. Alg. 2's strategy of reserving energy and setting a more conservative threshold against prediction errors proves advantageous compared to Alg. 1. Notably, Alg. 2 exhibits a higher peak shaving success rate than Alg. 1, showing greater robustness. This enhanced robustness of Alg. 2, achieved by reducing predefined targets, is particularly valuable in industrial applications such as peak shaving.

Regarding battery aging, we use the aging model from [16], with the temperature input for the in-house BESS assumed to be constant at 25 °C. To mitigate potential inaccuracies in the aging model at extremely low SoC levels, we have constrained the SoC lower bound to 0.1. We evaluate the battery capacity loss over ten years with a repeating annual SoC curve, with the results shown in Fig. 8, and calculate the average yearly capacity loss. The aging assessment indicates that the capacity loss is smallest under S0 at 1.21%, followed by S1 at 1.43%, with S2 exhibiting the fastest degradation at 1.68%.

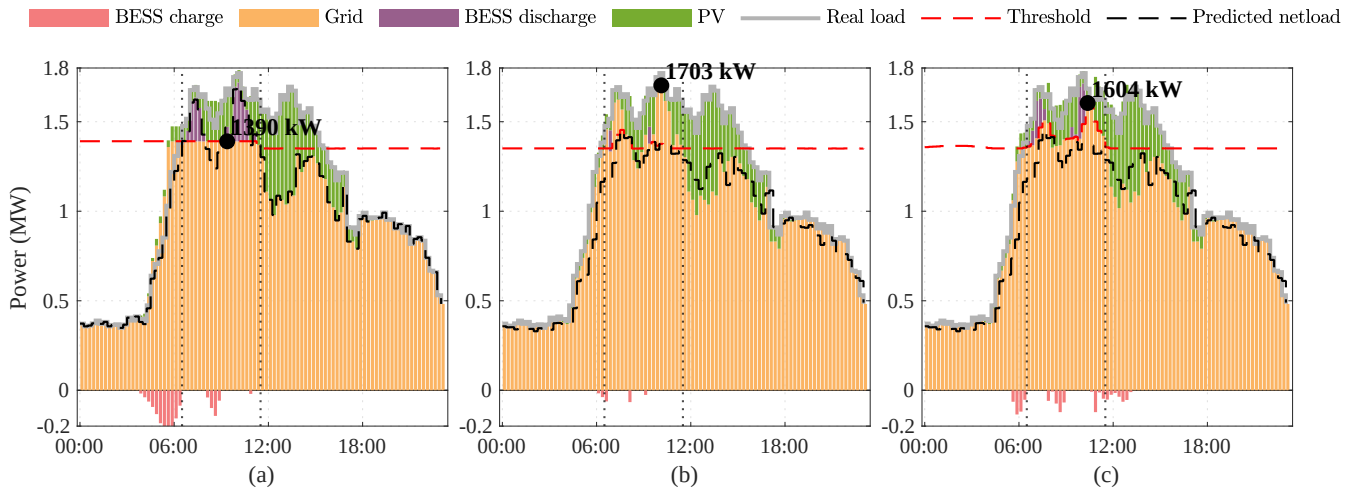

Fig. 5. Power distribution for 3 scenarios on the observation day with maximum netload of 1703 kW. The red dashed line is the optimized threshold $\theta^{\text{init}} + \gamma_t$. (a) Benchmark: 2 successes out of 2 attempts, with the maximum realized grid power reaching 1390 kW between 06 : 30 and 11 : 30. (b) MPC: 0 success out of 2 attempts, with the maximum realized grid power of 1703 kW at 10 : 15. (c) SMPC: 2 successes out of 6 attempts, with the maximum realized grid power reaching 1604 kW at 10 : 30.

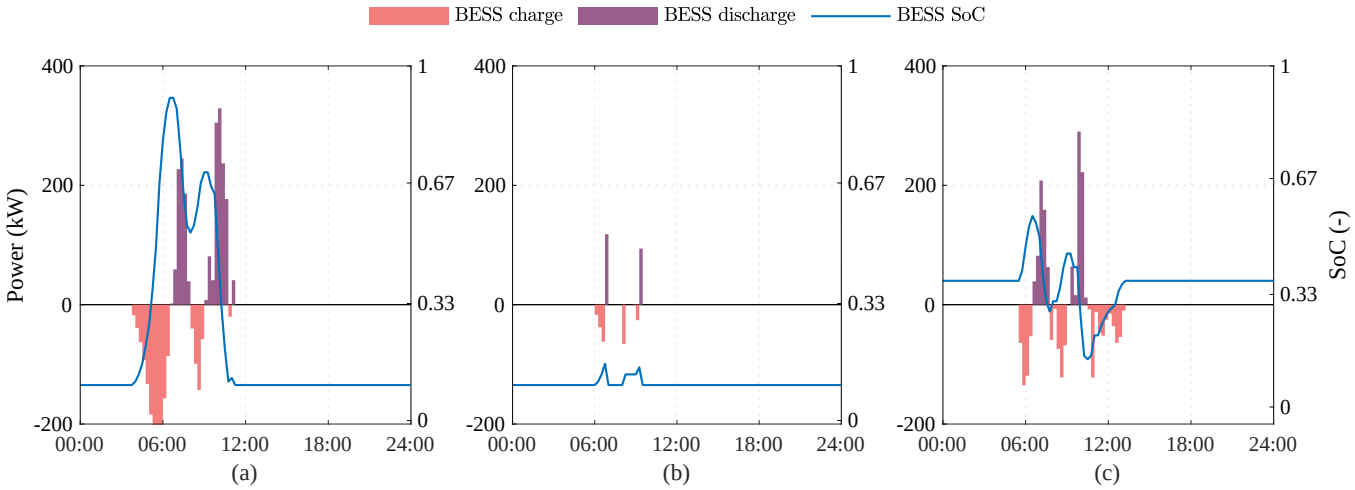

Fig. 6. BESS Profile and SoC for 3 scenarios on the observation day. (a) Benchmark: The maximum, minimum values of SoC are 0.9 and 0.1, respectively. (b) MPC: The maximum, minimum values of SoC are 0.16 and 0.1, respectively. (c) SMPC: The maximum, minimum values of SoC are 0.56 and 0.14, respectively.

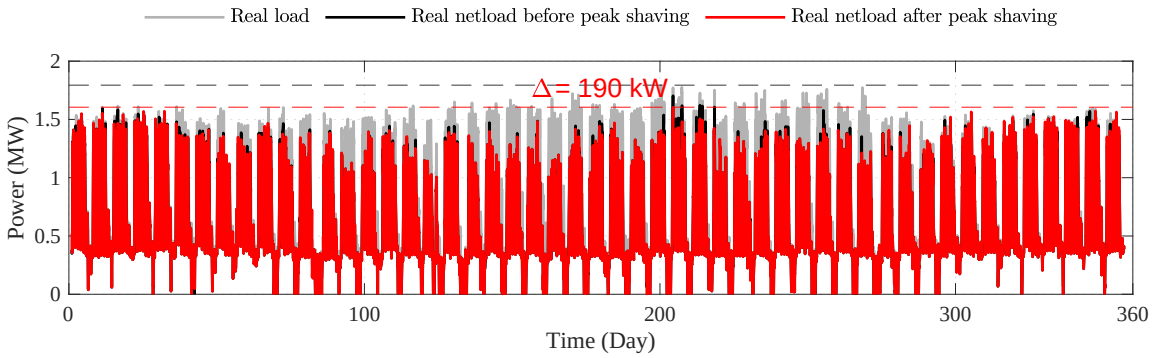

Fig. 7. One-year simulation of peak shaving with SMPC framework. The PV generation reduces the peak load from 1794 kW to 1703 kW (-91 kW), and the SMPC framework reduces the peak netload from 1703 kW to 1604 kW (-99 kW). The total reduction of peak load is 190 kW with the PV-BESS system.

TABLE IV
ONE-YEAR SIMULATION OF THREE SCENARIOS

Metrics	S0	S1	S2
Peak shaving attempts (-)	295	242	168
Peak shaving success rate (%)	100	39	80
Peak reduction (kW)	251	0	99
Peak reduction rate (%)	14.7	0	5.8
Saved demand cost (k€)	54.2	0	21.4
Average SoC (-)	0.12	0.14	0.4
Battery throughput (MWh)	46	96	97
Capacity loss (%)	1.21	1.43	1.68
Degradation cost[1](k€)	15	18	21
Net revenue[2](k€)	39	−18	0.4

[1] Degradation cost=BESS total costs×Capacity loss/0.2.
[2] Net revenue=Saved demand cost−Degradation cost.

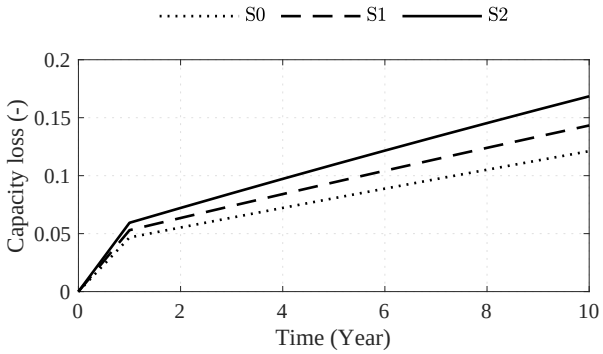

Fig. 8. Battery aging assessment using SoC profiles in 3 scenarios. The formation of the Solid Electrolyte Interphase (SEI) film is only considered in the first year.

In summary, as shown in the last row of Table IV, Alg. 1 achieves a gain of € 39k under perfect prediction but incurs a loss of € 18k under imperfect prediction presented in this study. Conversely, Alg. 2 improves peak shaving in both success rate and peak reduction rate under the same imperfect prediction, at the cost of accelerating the BESS loss by 0.25%. Ultimately, Alg. 2 in S2 achieves profitability (€ 0.4k) without consideration of the maintenance costs. With battery prices continuing to decline and energy prices rising, using Alg. 2 for profitable peak shaving is highly promising.

E. Simulation with Battery Life Assessment

The case study results indicate that the proposed SMPC framework enables the BESS to achieve peak shaving with positive net revenues under the optimized parameter setup. A methodological limitation, however, arises from the aging assessment approach, in which a single year of operation is repeatedly used to extrapolate battery degradation over a ten-year horizon. This simplification may overlook the dynamic interaction between battery capacity fade and operational performance.

To address this limitation and assess the impact of progressive capacity loss on peak shaving effectiveness, we incorporated a battery life assessment mechanism into the simulation process. In this extended simulation, battery capacity is updated every year based on assessment results and subsequently used for the next year. The full analysis spans a ten-year

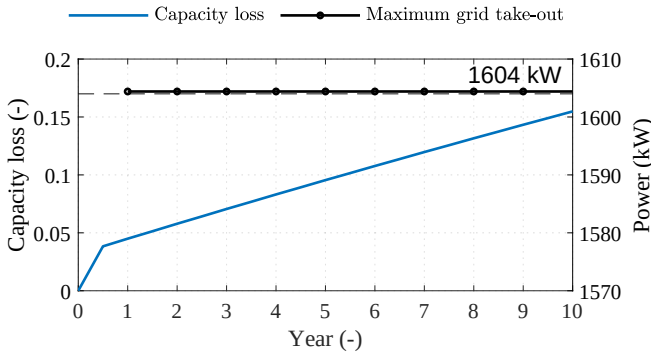

Fig. 9. Ten-year simulation with battery life assessment.

time range, allowing for a more realistic evaluation of aging effects on system performance. The results are shown in Fig. 9. As the BESS degrades over time, it consistently maintains a stable peak reduction in each year, attributed to the risk-averse SMPC management framework. This phenomenon is explained in Fig 6(c), which shows that the BESS operates within a conservative energy band that remains unaffected by battery capacity fade. Given that peak load charges are typically assessed per year, the proposed SMPC framework can maintain effective peak shaving over 10 years, even with a capacity degradation of approximately 16%. These results highlight the robustness of the proposed SMPC management framework for industrial peak load shaving applications.

V. PERFORMANCE ANALYSIS AND DISCUSSION

In this section, we analyze the impact of penalty parameters, confidence levels, and the length of look-ahead horizons through extended simulations, in order to broaden the framework's applicability.

A. Impact of Penalty Parameters

Following [35], we conducted a parameter sweep to examine how the penalty parameters ρ_1 and ρ_2 influence the net economic benefit, defined as peak demand cost savings minus aging costs. In this section, the simulation was conducted at a fixed confidence level of $\beta = 0.99$.

For ρ_1, given the normalized nature of the SoC variable, discrete parameter values (1, 3, 5, and 10) were chosen for systematic evaluation. For ρ_2, the power penalty term was normalized by scaling with $\alpha = \frac{1}{500^2}$ to maintain consistency with the SoC penalty term. Subsequently, appropriate numerators were selected to balance the penalty terms effectively relative to the overall objective. Sixteen simulations were conducted, covering all parameter combinations. Table V summarizes these results, identifying the optimal parameter combination as $(\rho_1, \rho_2) = (5, \frac{3}{500^2})$, as it strikes an optimal balance between achieving peak shaving benefits and minimizing battery degradation costs. The corresponding annual net revenue is highlighted in bold as the highest among all cases. Note that this heuristic approach provides practical guidance for selecting penalty parameters.

TABLE V
ANNUAL NET REVENUE OF S2 WITH DIFFERENT PENALTY PARAMETERS. ALL RESULTS IN THE TABLE ARE IN €. $\alpha = \frac{1}{500^2}$.

ρ_1 \ ρ_2	α	3α	5α	10α
1	-559	-1366	-536	-695
3	-82	225	191	-1480
5	293	**384**	350	316
10	202	202	293	259

B. Impact of Confidence Levels

The confidence level β plays a crucial role in Alg. 2's conservatism, warranting further investigation. In this section, we examine the effects of variations in the internal parameter β on the ultimate operational outcomes.

When β is set to 0.5, $\Phi^{-1}(0.5)$ yields 0. In this context, Alg. 2 completely disregards the presence of prediction errors, exhibiting extreme risk-seeking behavior and essentially emulating the behavior of the MPC-based algorithm (Alg. 1). Conversely, as β increases, the algorithm's error margin will increase, demonstrating risk aversion. However, if β increases too much, the constraints (19a) will become excessively tight, making the optimization problem unfeasible. To prevent this situation, β can be constrained by (21),

$$\underline{x} + F(\beta, k, \lambda) < \overline{x},\ \forall k \in \mathcal{K} \tag{21}$$

indicating that the lower bound of the energy state can not exceed the upper bound.

Therefore, we conducted the one-year simulation using Alg. 2 under different β within the range $[0.5, 0.99]$, with partial results depicted in Table VI. As β increases, Alg. 2 becomes more conservative, as evidenced by fewer peak shaving attempts and an increase in the average SoC of the BESS. The resulting positive effect is a general increase in the power reduction rate and the saved peak demand cost, while the negative effect is that the BESS ages faster and the aging cost increases. In general, however, the increase in revenue from peak shaving by increasing β is faster than the increase in the aging cost of the BESS, increasing overall revenue.

TABLE VI
ONE YEAR SIMULATION OF S2 WITH DIFFERENT CONFIDENCE LEVELS

Objects	Metrics	0.5	0.84	0.95	0.99
Peak shaving	Attempts (-)	242	230	200	168
	Success rate (%)	39	70	76	80
	Reduction rate (%)	0	0	1.7	5.8
	Saved cost (k€)	0	0	6.3	21.4
BESS	Average SoC (-)	0.14	0.22	0.31	0.4
	Throughput (MWh)	96	94	96.7	97
	Capacity loss (%)	1.43	1.49	1.58	1.68
	Deg. cost (k€)	18	18.6	19.8	21
Summary	Revenue (k€)	-18.6	-21.7	-13.5	0.4

C. Impact of Look-Ahead Window Length

To investigate the impact of the look-ahead window length N on the peak shaving performance of the BESS, we conducted experiments during the week containing July 24th, identified as the period with the poorest SMPC performance. The results, detailed in Table VII, indicate that reducing the prediction window from 24 hours to 8 hours resulted in negligible changes to peak reduction, while the single-step computation time decreased from 0.22 seconds to 0.05 seconds. However, further shortening the prediction window from 8 hours to 2 hours resulted in diminished peak-shaving performance and reduced cost savings. These findings suggest that optimizing the prediction window to approximately 8 hours may offer a balanced trade-off between algorithm effectiveness and computational complexity. Extensive and prolonged simulations are necessary to substantiate this conclusion.

TABLE VII
PEAK REDUCTION OF S2 WITH DIFFERENT LOOK-AHEAD WINDOW LENGTHS

Look-ahead window (h)	24	20	16	12	8	4	2
Peak reduction (kW)	99	98	98	98	98	97	92
Demand cost savings (k€)	21	21	21	21	21	20	19

VI. CONCLUSION

This paper proposes a novel energy management framework for large electricity customers, aiming at peak shaving and battery health optimization. It formulates a stochastic optimal power distribution problem with chance constraints and employs a constraint-tightening approach to convert the chance constraints equivalently into their deterministic counterparts, thereby providing conservative but effective solutions that consider prediction errors in both load and PV generation. The practical relevance and applicability of the framework are demonstrated using real load and PV data with varying prediction errors.

For the case study, a risk-averse SMPC management framework with a confidence level of $\beta = 0.99$ achieves a reduction of 99 kW (5.8%) in peak load and maximizes revenue, even when the standard deviation of netload predictions reaches up to 20% of the peak netload.

Future work will focus on improving load forecasting methods, applying alternative probability distribution models to capture prediction uncertainty, refining BESS modeling, and developing multi-use energy management frameworks.

APPENDIX
LITHIUM-ION BATTERY DEGRADATION MODEL

The lithium-ion battery degradation model and battery parameters used in this work are derived from [16]. The model encompasses temperature, SoC, time, and DoD to characterize both calendar and cycle aging. Capacity loss is calculated by applying the battery's SoC and temperature profile. However, it does not account for the increase in impedance and power fading associated with a low SoC, which can contribute to aging. Despite this limitation, the model has demonstrated accurate and reliable aging assessments across LMO, LFP, and NMC batteries.

The battery life L in (22) is modeled as a two-exponential function:

$$L = 1 - \alpha_{\text{sei}} e^{\beta_{\text{sei}} f_{\text{d}}} - (1 - \alpha_{\text{sei}}) e^{-f_{\text{d}}} \tag{22}$$

where α_{sei}, β_{sei} are the coefficients for the SEI model, and f_{d}, $1 - \alpha_{\text{sei}}$ are the degradation function and the relevant

coefficient. $L = 0$ indicates a new battery, $L = 0.2$ means the end of life.

The degradation function f_{d} in (23) is the function of test time t, depth of discharge (DoD) δ, average SoC σ, and average cell temperature T_c, consisting of calendar aging f_{t} and cyclic aging f_{c},

$$f_{\mathrm{d}}(t,\delta,\sigma,T_c) = f_{\mathrm{t}}(t,\sigma,T_c) + \sum_{i}^{N} n_i f_{\mathrm{c}}(\delta_i,\sigma_i,T_{c,i}) \tag{23a}$$

$$f_{\mathrm{t}}(t,\sigma,T_c) = S_t(t)S_\sigma(\sigma)S_T(T_c) \tag{23b}$$

$$f_{\mathrm{c}}(\delta,\sigma,T_c) = S_\delta(\delta)S_\sigma(\sigma)S_T(T_c) \tag{23c}$$

where $S_t(t)$, $S_\delta(\delta)$, $S_T(T_c)$, and $S_\sigma(\sigma)$ are stress models of time, DoD, temperature, and SoC, respectively. The temperature stress model $S_T(T) = e^{k_T(T-T_{\mathrm{ref}})\frac{T_{\mathrm{ref}}}{T}}$ captures the exponential sensitivity of aging to cell temperature, where k_T and T_{ref} are the temperature stress coefficient and the reference temperature in Kelvin. The SoC stress model is expressed as $S_\sigma(\sigma) = e^{k_\sigma(\sigma-\sigma_{\mathrm{ref}})}$, where k_σ and σ_{ref} are the SoC stress coefficient and the reference level. The effect of low SoC levels on battery aging is not modeled here. Time-induced degradation is represented by a linear stress model $S_t(t) = k_t t$, where k_t and t are the time stress coefficient and the test duration. Finally, the DoD stress model is formulated as $S_\delta(\delta) = (k_{\delta 1}\delta^{k_{\delta 2}} + k_{\delta 3})^{-1}$, where $k_{\delta 1}$, $k_{\delta 2}$ and $k_{\delta 3}$ are empirical coefficients.

For used batteries, whose life is near or below 0.9, the SEI film formation is finished. The battery life L can be simplified in (24),

$$L = 1 - (1 - L')e^{-f_{\mathrm{d}}} \tag{24}$$

where L' is the current battery life.

References

[1] A. A. Kebede, T. Kalogiannis, J. Van Mierlo, and M. Berecibar, "A comprehensive review of stationary energy storage devices for large scale renewable energy sources grid integration," *Renewable and Sustainable Energy Reviews*, vol. 159, p. 112213, May 2022.

[2] K. Mongird, V. V. Viswanathan, P. J. Balducci, M. J. E. Alam, V. Fotedar, V. Koritarov, and B. Hadjerioua, "Energy storage technology and cost characterization report," Pacific Northwest National Laboratory (PNNL), Richland, WA (United States), Tech. Rep., 2019.

[3] M. Rouholamini, C. Wang, H. Nehrir, X. Hu, Z. Hu, H. Aki, B. Zhao, Z. Miao, and K. Strunz, "A Review of Modeling, Management, and Applications of Grid-Connected Li-Ion Battery Storage Systems," *IEEE Transactions on Smart Grid*, vol. 13, no. 6, pp. 4505–4524, Nov. 2022.

[4] H. Hesse, M. Schimpe, D. Kucevic, and A. Jossen, "Lithium-Ion Battery Storage for the Grid—A Review of Stationary Battery Storage System Design Tailored for Applications in Modern Power Grids," *Energies*, vol. 10, no. 12, p. 2107, Dec. 2017.

[5] F. Braeuer, J. Rominger, R. McKenna, and W. Fichtner, "Battery storage systems: An economic model-based analysis of parallel revenue streams and general implications for industry," *Applied Energy*, vol. 239, pp. 1424–1440, Apr. 2019.

[6] M. Uddin, M. F. Romlie, M. F. Abdullah, S. Abd Halim, A. H. Abu Bakar, and T. Chia Kwang, "A review on peak load shaving strategies," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 3323–3332, Feb. 2018.

[7] S. Henni, J. Becker, P. Staudt, F. Vom Scheidt, and C. Weinhardt, "Industrial peak shaving with battery storage using a probabilistic forecasting approach: Economic evaluation of risk attitude," *Applied Energy*, vol. 327, p. 120088, Dec. 2022.

[8] Y. Shi, B. Xu, D. Wang, and B. Zhang, "Using Battery Storage for Peak Shaving and Frequency Regulation: Joint Optimization for Superlinear Gains," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 2882–2894, May 2018.

[9] N. B. GmbH, "Electricity price from Netze BW GmbH," Jan. 2024. [Online]. Available: https://www.netze-bw.de/unternehmen/veroeffentlichungen

[10] M. Zheng, C. J. Meinrenken, and K. S. Lackner, "Smart households: Dispatch strategies and economic analysis of distributed energy storage for residential peak shaving," *Applied Energy*, vol. 147, pp. 246–257, Jun. 2015.

[11] M. Koller, T. Borsche, A. Ulbig, and G. Andersson, "Defining a degradation cost function for optimal control of a battery energy storage system," in *2013 IEEE Grenoble Conference*. Grenoble, France: IEEE, Jun. 2013, pp. 1–6.

[12] J. Engels, B. Claessens, and G. Deconinck, "Optimal Combination of Frequency Control and Peak Shaving With Battery Storage Systems," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3270–3279, Jul. 2020.

[13] X. Han, L. Lu, Y. Zheng, X. Feng, Z. Li, J. Li, and M. Ouyang, "A review on the key issues of the lithium ion battery degradation among the whole life cycle," *eTransportation*, vol. 1, p. 100005, Aug. 2019.

[14] N. Munzke, F. Büchle, A. Smith, and M. Hiller, "Influence of efficiency, aging and charging strategy on the economic viability and dimensioning of photovoltaic home storage systems," *Energies*, vol. 14, no. 22, p. 7673, 2021.

[15] J. Schmalstieg, S. Käbitz, M. Ecker, and D. U. Sauer, "A holistic aging model for Li(NiMnCo)O2 based 18650 lithium-ion batteries," *Journal of Power Sources*, vol. 257, pp. 325–334, Jul. 2014.

[16] B. Xu, A. Oudalov, A. Ulbig, G. Andersson, and D. S. Kirschen, "Modeling of Lithium-Ion Battery Degradation for Cell Life Assessment," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1131–1140, Mar. 2018.

[17] J. Li and M. A. Danzer, "Optimal charge control strategies for stationary photovoltaic battery systems," *Journal of Power Sources*, vol. 258, pp. 365–373, Jul. 2014.

[18] G. Angenendt, S. Zurmühlen, H. Axelsen, and D. U. Sauer, "Comparison of different operation strategies for PV battery home storage systems including forecast-based operation strategies," *Applied Energy*, vol. 229, pp. 884–899, Nov. 2018.

[19] K. J. Iheanetu, "Solar Photovoltaic Power Forecasting: A Review," *Sustainability*, vol. 14, no. 24, p. 17005, Dec. 2022.

[20] K. Amasyali and N. M. El-Gohary, "A review of data-driven building energy consumption prediction studies," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1192–1205, 2018.

[21] B. Kouvaritakis and M. Cannon, *Model Predictive Control*, ser. Advanced Textbooks in Control and Signal Processing. Cham: Springer International Publishing, 2016.

[22] L. G. Marín, M. Sumner, D. Muñoz-Carpintero, D. Köbrich, S. Pholboon, D. Sáez, and A. Núñez, "Hierarchical Energy Management System for Microgrid Operation Based on Robust Model Predictive Control," *Energies*, vol. 12, no. 23, p. 4453, Nov. 2019.

[23] P. Malysz, S. Sirouspour, and A. Emadi, "An Optimal Energy Storage Control Strategy for Grid-connected Microgrids," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1785–1796, Jul. 2014.

[24] A. Mesbah, "Stochastic model predictive control: An overview and perspectives for future research," *IEEE Control Systems Magazine*, vol. 36, no. 6, pp. 30–44, 2016.

[25] D. Zhu and G. Hug, "Decomposed Stochastic Model Predictive Control for Optimal Dispatch of Storage and Generation," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 2044–2053, Jul. 2014.

[26] R. Kumar, M. J. Wenzel, M. J. Ellis, M. N. ElBsat, K. H. Drees, and V. M. Zavala, "A Stochastic Model Predictive Control Framework for Stationary Battery Systems," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4397–4406, Jul. 2018.

[27] A. Ravichandran, S. Sirouspour, P. Malysz, and A. Emadi, "A Chance-Constraints-Based Control Strategy for Microgrids With Energy Storage and Integrated Electric Vehicles," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 346–359, Jan. 2018.

[28] K. Baker, G. Hug, and X. Li, "Energy Storage Sizing Taking Into Account Forecast Uncertainties and Receding Horizon Operation," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 1, pp. 331–340, Jan. 2017.

[29] N. Bazmohammadi, A. Tahsiri, A. Anvari-Moghaddam, and J. M. Guerrero, "A hierarchical energy management strategy for interconnected microgrids considering uncertainty," *International Journal of Electrical Power & Energy Systems*, vol. 109, pp. 597–608, Jul. 2019.

[30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[31] S. Dutta and R. Sharma, "Optimal storage sizing for integrating wind and load forecast uncertainties," in *2012 IEEE PES Innovative Smart*

Grid Technologies (ISGT). Washington, DC, USA: IEEE, Jan. 2012, pp. 1–7.
[32] A. Groß, L. Kreilgaard, B. Köpfer, and M. Kühnbach, "Peak-shaving with battery systems in commerce and industry," in *International Renewable Energy Storage Conference*. Atlantis Press, 2023, pp. 193–205.
[33] K. Heussen, S. Koch, A. Ulbig, and G. Andersson, "Unified System-Level Modeling of Intermittent Renewable Energy Sources and Energy Storage for Power System Operation," *IEEE Systems Journal*, vol. 6, no. 1, pp. 140–151, Mar. 2012.
[34] C. Zhao, P. B. Andersen, C. Træholt, and S. Hashemi, "Grid-connected battery energy storage system: a review on application and integration," *Renewable and Sustainable Energy Reviews*, vol. 182, p. 113400, 2023.
[35] U. R. Nair, M. Sandelic, A. Sangwongwanich, T. Dragicevic, R. Costa-Castello, and F. Blaabjerg, "An Analysis of Multi Objective Energy Scheduling in PV-BESS System Under Prediction Uncertainty," *IEEE Transactions on Energy Conversion*, vol. 36, no. 3, pp. 2276–2286, Sep. 2021.
[36] Gurobi Optimization, LLC, "Gurobi Optimizer Reference Manual," 2024. [Online]. Available: https://www.gurobi.com
[37] F. Braeuer, "Load profile data of 50 industrial plants in Germany for one year," Jun. 2020. [Online]. Available: https://zenodo.org/records/3899018
[38] T. Kappler, A. S. Starosta, B. Schwarz, N. Munzke, and M. Hiller, "Inclusion of shading and soiling with physical and data-driven algorithms for solar power forecasting," in *PV-Symposium Proceedings*, vol. 1, 2024.
[39] D. Dongol, T. Feldmann, M. Schmidt, and E. Bollin, "A model predictive control based peak shaving application of battery for a household with photovoltaic system in a rural distribution grid," *Sustainable Energy, Grids and Networks*, vol. 16, pp. 1–13, Dec. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2352467717302916
[40] B. Xu, Y. Shi, D. S. Kirschen, and B. Zhang, "Optimal battery participation in frequency regulation markets," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6715–6725, 2018.
[41] W. J. Cole and A. Frazier, "Cost projections for utility-scale battery storage," National Renewable Energy Laboratory (NREL), Golden, CO, USA, Technical Report NREL/TP-6A20-73222, 06 2019. [Online]. Available: https://www.nrel.gov/docs/fy19osti/73222.pdf
[42] L. Li, A. S. Starosta, B. Schwarz, N. Munzke, H.-M. Strehle, M. Richter, and M. Hiller, "Optimal design of energy storage system for peak-shaving in industrial production," *NEIS 2023 - Conference on Sustainable Energy Supply and Energy Storage Systems*, p. 79 – 85, 2023.
[43] "Lithium-ion Battery price database," publisher: Pacific Northwest National Laboratory. [Online]. Available: https://www.pnnl.gov/projects/esgc-cost-performance/lithium-ion-battery

Lixin Li (Student member, IEEE) received the M.Sc. degree in Electrical Engineering and Information Technology from Karlsruhe Institute of Technology, Karlsruhe, Germany, in 2022. He is currently a research associate and Ph.D. candidate at the Institute of Electrical Engineering (ETI), Karlsruhe Institute of Technology, Karlsruhe, Germany. His research interests include modeling, optimization, and techno-economic analysis of energy storage systems for grid and industrial applications.

Tim Kappler (Student member, IEEE) received the M.Sc. degree in Electrical Engineering and Information Technology from Karlsruhe Institute of Technology, Karlsruhe, Germany, in 2022. He is currently a research associate and Ph.D. candidate at the Institute of Electrical Engineering (ETI), Karlsruhe Institute of Technology, Karlsruhe, Germany. His research interests include machine-learning based solar power and load forecasting.

Bernhard Schwarz received the M.Sc. degree in Electrical and Information Engineering with a specialization in Renewable Energies from the Karlsruhe Institute of Technology (KIT), Germany, in 2015. He then joined KIT as a research associate specializing in stationary energy storage systems. Since 2022, he has been a group leader at the ETI, KIT. His research areas include residential and large-scale energy storage systems, system characterization and modeling, state-of-energy estimation, and energy management.

Nina Munzke received her M.Sc. in Renewable Raw Materials and Regenerative Energies from Brandenburg University of Technology in 2012. She works at the ETI of the Karlsruhe Institute of Technology, leading the Systems Control and Analysis group in the field of energy storage systems since 2013. She has extensive experience in dimensioning and simulation, and in developing intelligent system controls for stationary storage systems. She is also an expert in evaluating the performance of stationary storage systems.

Xinliang Dai (Member, IEEE) received the M.Sc. and Ph.D. degrees in Electrical Engineering and Information Technology from Karlsruhe Institute of Technology, Karlsruhe, Germany, in 2021 and 2025.

He is a postdoctoral research associate in the Zero-carbon Energy systems Research and Optimization Laboratory (ZERO Lab), at Princeton University. His research interests include graph-based distributed optimization, flexibility aggregation, and nonlinear model predictive control.

Veit Hagenmeyer (Member, IEEE) received the Ph.D. degree from Université Paris XI, Paris, France, in 2002. He is currently the Professor of Energy Informatics with the Faculty of Informatics, and the Director of the Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany. His research interests include modeling, optimization and control of sector-integrated energy systems, machine-learning based forecasting of uncertain demand and production in energy systems mainly driven by renewables, and integrated cyber-security of such systems.

Marc Hiller (Member, IEEE) received the diploma degree from the Technische Universität Darmstadt, Darmstadt, Germany, in 1993, and the Ph.D. degree from the Universität der Bundeswehr München, Neubiberg, Germany, in 2008, both in electrical engineering. Since 2005, he has been with the Drive Technology Division, Sector Industry, Siemens AG, Nuremberg, Germany, where he was the Project-Manager in the development of industrial medium-voltage drives. His research interests include modern power-hardwarein-the-loop converters, modeling of electrical drives and grids, and grid connected converter, energy storage integration, and the control of power semiconductors. Dr. Hiller has been the Chair of Power Electronic Systems with the Karlsruhe Institute of Technology, Karlsruhe, Germany, since 2015.