IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 February 2025, accepted 18 April 2025, date of publication 22 April 2025, date of current version 30 April 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3563300

RESEARCH ARTICLE

Neural Network-Based Imitation Learning for Approximating Stochastic Battery Management Systems

ANDREA POZZI[1], (Member, IEEE), ALESSANDRO INCREMONA[1], AND DANIELE TOTI[1,2], (Member, IEEE)

[1]Faculty of Mathematical, Physical and Natural Sciences, Catholic University of the Sacred Heart, 25133 Brescia, Italy
[2]Department of Sciences, Roma Tre University, 00146 Rome, Italy

Corresponding author: Daniele Toti (daniele.toti@uniroma3.it)

This work was supported in part by European Union–Next Generation EU, Mission 4, Component 2, under Grant (CUP) n. 2022AEEKXS.

ABSTRACT Lithium-ion batteries play a pivotal role in enabling eco-friendly mobility, particularly in electric vehicles, but optimizing their charging process to improve battery lifespan, safety, and overall efficiency remains a significant challenge. Traditional predictive control methods are limited by their reliance on precise models, which are often hindered by uncertainties in battery parameters due to aging, production variability, and operational conditions. While stochastic predictive control policies can address these uncertainties by incorporating them directly into the optimization process, they typically introduce considerable computational complexity. In response to this challenge, this paper presents a novel approach that adapts imitation learning to efficiently approximate stochastic predictive control strategies, thus significantly reducing the computational burden through offline training. Specifically, the proposed method leverages the Dataset Aggregation algorithm to overcome the issue of distributional shift, a common limitation in imitation learning frameworks. Simulations based on a detailed electrochemical model demonstrate the effectiveness of the method, adhering to probabilistic constraints while offering a scalable and computationally efficient solution for advanced battery management systems.

INDEX TERMS Imitation learning, neural networks, stochastic control, battery management systems.

I. INTRODUCTION

In the landscape of modern industrial innovation, ecological sustainability has emerged as a paramount concern, propelling the advancement of energy storage technologies, particularly lithium-ion batteries, to the forefront of this transformation. This surge in prominence is significantly attributed to their critical role in sustainable mobility, especially with the expanding adoption of electric vehicles. Lithium-ion batteries, heralded for their efficiency and longevity, have become integral in this green transition, marking a significant stride towards eco-friendly transportation solutions. However, the efficient management of these batteries, especially during the charging phase, poses a complex challenge. Improper charging protocols can lead to underutilization, safety risks, and accelerated degradation of the battery [1], thereby necessitating the adoption of sophisticated control strategies for optimization [2], [3], [4] and efficient monitoring systems [5], [6], [7], [8], [9]. Recent developments have seen the integration of advanced Battery Management Systems (BMSs) in industrial applications, which leverage intricate mathematical models to optimize battery charging processes, thus enhancing performance and safety.

One of the most prominent strategies employed in battery management is Model Predictive Control (MPC) [10]. This method, through the formulation of the control task as a constrained optimization, intricately captures the dynamics and operational bounds of battery systems, achieving enhanced performance and safety. The efficacy of MPC in battery management is well-documented, with numerous studies demonstrating its superiority over conventional charging protocols. Unlike standard methods, MPC's model-based approach allows for the incorporation of safety constraints

The associate editor coordinating the review of this manuscript and approving it for publication was Vitor Monteiro.

© 2025 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

without being overly conservative, making it a more effective and efficient solution [11], [12], [13], [14], [15], [16].

Despite the successes of MPC, its practical application is not without challenges. A deterministic model predictive framework requires precise battery models [17] and accurate parameter estimation, requirements that are often difficult to meet due to the inherent uncertainties in battery parameters, especially for cell capacity and resistance [18].

Furthermore, throughout a battery's life, parameters related to its cells are bound to vary. This variation forces models to require a synchronous, online phase for the identification of these parameters. In this regard, controllers that are deterministic, given the fact that they only consider fixed, nominal parameters, show a lackluster performance when faced with realistic scenarios, when the latter diverge from the models in non-negligible ways. Besides, in such cases safety hazards may also occur, basically undermining one of the core principles of a BMS, which is supposed to act as a damage-preventer for batteries and users alike.

In response to these challenges, some studies have explored the potential of stochastic MPC to optimize charging strategies under parameter uncertainty [19], [20], [21], [22]. Stochastic MPC aims to minimize the risk of constraint violation by considering the distribution of uncertain parameters [23]. However, the integration of stochastic MPC, particularly with nonlinear models, introduces significant computational complexity, making real-time application challenging [24]. This complexity stems from the need to either analyze multiple scenarios of parameter distributions or estimate statistical moments of model outputs, both of which are computationally intensive tasks.

In this paper, the computational hurdles of stochastic MPC are addressed by introducing a novel approach that leverages imitation learning techniques. Imitation learning, a subset of the broader reinforcement learning field [25], is fundamentally about emulating the behavior of an expert agent by learning a policy that approximates the expert's decision-making process [26]. While the domain of battery management has seen the application of various machine learning paradigms, including reinforcement learning for optimal charging [27], [28], and supervised learning techniques for state and parameter estimation [29], [30], [31], the exploration of imitation learning specifically for this purpose is a more recent endeavor. Specifically, imitation learning in the form of behavioral cloning has been applied to emulate an expert charging strategy by observing and mimicking its state-action relations [32], [33]. This technique fits a machine learning model to transform states into actions, with the aim of mirroring the experienced agent's strategy in a supervised fashion. However, behavioral cloning faces significant hurdles, including issues with shifts in data distribution. Such phenomenon, known as distributional shift or covariate shift, is caused by the fact that, during testing, the trained agent comes across scenarios that are either absent or scarcely represented in the dataset composed of expert demonstrations [34]. Such encounters can lead to deviations from the desired path, potentially leading to a compounding of errors and rendering the model incapable of recovering from unexpected states [35]. To address this inherent limitation of behavioral cloning, the Dataset Aggregation (DAgger) algorithm emerges as a robust solution [36]. DAgger curtails the compounding of errors attributed to shifts in data distribution by iteratively integrating the learning model's actions with the expert policy. This approach keeps the trained agent closely aligned with the desired path, significantly mitigating inaccuracies stemming from unexplored state spaces.

In the context of smart grid and battery management, imitation learning solutions have been recently applied to approximate charging protocols based on deterministic MPC [37], [38]. The current study distinguishes itself by considering a stochastic predictive control as the expert agent, explicitly integrating information about parameter distributions. By leveraging the strengths of DAgger, this research not only addresses the computational challenges of stochastic MPC but also sets a precedent in the application of these methodologies for optimizing lithium-ion battery charging under uncertainty. The core of the proposed methodology lies in the offline generation of synthetic demonstrations by the stochastic MPC, which serves as a rich dataset for training a neural network. This network, in turn, maps states to optimal actions, effectively approximating the stochastic MPC policy and significantly reducing the computational cost required during online execution, owing to the shift of stochastic optimization's complexity to the offline training phase. The efficacy of this approach is validated through a battery simulator based on an electrochemical model, showcasing the potential of the presented methodology to optimize battery charging while adhering to probabilistic constraints and maintaining manageable computational demands.

II. BATTERY MODEL AND OPTIMAL CHARGING

This section provides an in-depth exploration of the battery charging problem, outlining the framework and methodologies traditionally utilized in addressing this complex issue. Within Subsection II-A, the battery model selected for simulations is detailed, including a description of both the nominal parameters and the probabilistic distributions that represent their inherent variabilities. This is followed by Subsection II-B, which presents the requirements essential for safe and rapid battery charging, highlighting the objectives that optimal charging strategies must fulfill. Finally, Subsection II-C introduces the stochastic predictive control framework, designated as the expert agent in this study. The discussion within this section is pivotal for comprehending the challenges, limitations, and complexities associated with traditional charging protocols and stochastic approaches. It establishes a critical foundation of knowledge necessary for the effective implementation of innovative methodologies, such as the imitation learning-based approach introduced in Section III of this paper.

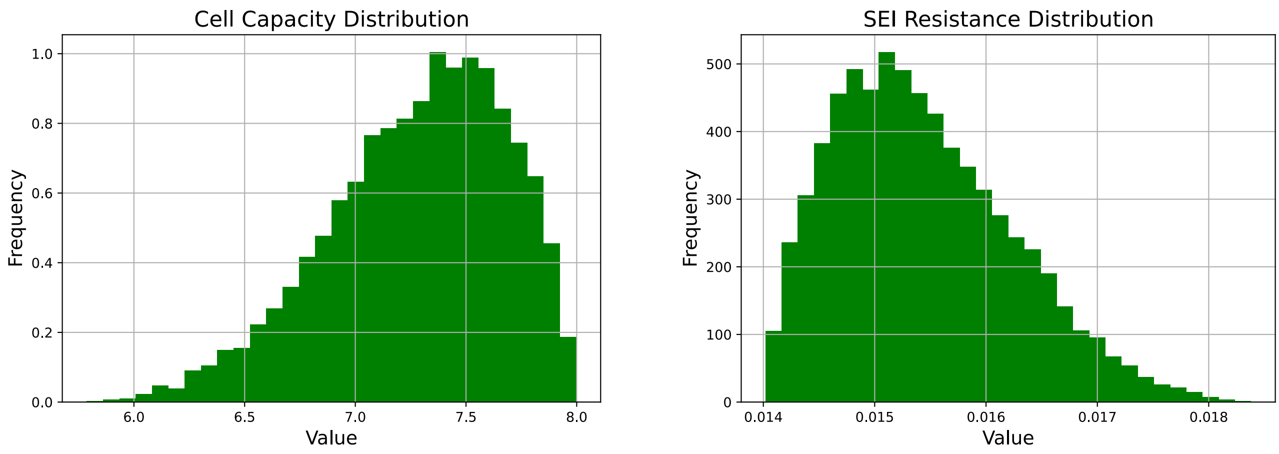


FIGURE 1. This figure illustrates the Beta distributions used to model the variability in cell capacity (left) and SEI resistance (right) around their nominal values ($C^n = 8Ah$ and $R^n_{sei} = 14m\Omega$), considering the variation ranges $\Delta C = 2.5Ah$ and $\Delta R_{sei} = 5m\Omega$, and the Beta distribution parameters $a = 2$ and $b = 5$.

A. ADOPTED BATTERY MODEL

This study employs a widely recognized mathematical representation of batteries, which is rooted in the foundational work by [39], named Single-Particle Model (SPM). This model, which stems from simplifying the more comprehensive one proposed in [40] by conceptualizing the electrodes as spherical entities, strikes a balanced compromise between computational efficiency and model fidelity, as suggested by the studies in [41] and [42]. In this paper, the model is further augmented to include a two-state temperature dynamics as proposed in [43], offering an enriched depiction of the battery's thermal attributes.

In this exposition, the focus is directed on the pivotal variables that significantly influence the battery's behavior. A critical parameter is the State of Charge (SoC), indicated as $z(t)$ and ranging from 0 (complete discharge) to 1 (full charge), whose time dynamics as a function of the applied current $i(t)$ is as follows:

$$\dot{z}(t) = \frac{i(t)}{3600C}, \tag{1}$$

where C represents the battery's capacity measured in ampere-hours. The battery's voltage, denoted by $v(t)$, is formulated as:

$$v(t) = v_{\text{ocp}}(z(t)) + \Delta\eta(t) + R_{sei}i(t). \tag{2}$$

In this equation, the battery's open circuit potential $v_{\text{ocp}}(z(t))$ is a nonlinear function of the state of charge, while $\Delta\eta(t)$ represents the difference between the positive and negative overpotentials, for which more details can be found in [32]. The term $R_{sei}i_{\text{app}}(t)$ corresponds to the voltage loss attributed to the resistance within the Solid Electrolyte Interphase (SEI).

For the thermal model, a two-state approach, accounting for both core ($T_c(t)$) and surface ($T_s(t)$) temperature dynamics, is adopted, as presented in [43]:

$$\dot{T}_c(t) = \frac{P_{\text{th}}(t)}{C_c} - \frac{T_c(t) - T_s(t)}{R_{c,s}C_c}, \tag{3a}$$

$$\dot{T}_s(t) = \frac{T_c(t) - T_s(t)}{R_{c,s}C_s} - \frac{T_s(t) - T_{env}}{R_{s,e}C_s}. \tag{3b}$$

In these equations, $R_{c,s}$ and $R_{s,e}$ signify the thermal impedances linking, respectively, the core to the surface, and the surface to the ambient environment (with temperature T_{env}). The symbols C_c and C_s describe, respectively, the core and surface thermal capacitances, and $P_{\text{th}}(t)$ quantifies the thermal power produced by the internal electrochemical reactions, calculated by:

$$P_{\text{th}}(t) = |i(t)\Delta v(t)|, \tag{4}$$

where $\Delta v(t) = v(t) - v_{\text{ocp}}(z(t))$.

Given the circumstances of inherent variability and the inevitable changes a cell undergoes throughout its lifecycle, especially due to aging, the reliance on nominal parameter values alongside a probabilistic distribution becomes a pragmatic approach. This paper utilizes nominal electrochemical parameters based on the empirical analysis of a commercial cell (notably the Kokam SLPB 75106100), as thoroughly investigated in [44] and [45]. Concurrently, nominal thermal parameters are adopted in accordance with the findings by Perez et al. [43]. Nevertheless, the actual parameters of a battery can significantly diverge from these nominal values due to cell-to-cell variability and age-related alterations. Critical parameters such as the capacity of the battery and the internal resistance are particularly prone to these fluctuations. This paper addresses the variability and uncertainty of these parameters by adopting nominal values $C^n = 8Ah$ and $R^n_{sei} = 14m\Omega$, refined by Beta distributions to represent the inherent uncertainties and variabilities. The Beta distribution is a family of continuous probability distributions defined on the interval $[0, 1]$. It is characterized by two shape parameters, a and b, which determine the distribution's shape. Specifically, $a > b$ results in a distribution skewed toward the upper end of the interval, $a < b$ skews the distribution toward the lower end, and $a = b$ produces a symmetric distribution centered around 0.5. In this study, the parameters

$a = 2$ and $b = 5$ are selected, with the values sampled from the distribution appropriately scaled and shifted to model variability in cell capacity and SEI resistance. This specific choice reflects the intention to account for a population of lithium-ion cells with varying degrees of aging. The majority of the cells are considered to be in a mid-life state, while only a small proportion of cells are highly aged and yet to be replaced. This assumption aligns well with the empirical evidence observed for commercially available cells.

The probability distributions for cell capacity and SEI resistance considered in this study are visualized in Figure 1. Given the approach to model the variability and uncertainty inherent in the battery parameters, the sampling method for the cell capacity and SEI resistance is mathematically represented by the following equations:

$$C \sim C^n - \text{Beta}(a, b) \cdot \Delta C, \tag{5a}$$

$$R_{sei} \sim R_{sei}^n + \text{Beta}(a, b) \cdot \Delta R_{sei}, \tag{5b}$$

where the variations ΔC and ΔR_{sei} define the ranges of fluctuation for these parameters from their nominal values. Such values are chosen as $2.5Ah$ and $5m\Omega$, respectively, which, being nearly 30% of the nominal value, can be considered as the variations associated with the battery end-of-life.

B. FAST AND SAFE BATTERY CHARGING

The following section provides a comprehensive overview of the task associated with achieving safe and expedited battery charging. The objective is multifaceted, focusing not just on optimizing the effort in current applications to attain a desired state of charge in minimal time, but also on complying with critical safety measures. Ensuring safety is paramount during battery charging to prevent potential risks such as overheating or battery damage, which may arise from excessively rapid charging procedures. Hence, maintaining stringent control during the charging phase over variables like voltage and temperature is indispensable.

A possible way to formally approach this task is to formulate a constrained optimization framework, whose solution determines the most efficient schedule for the applied current. In this setup, a discrete-time formulation of the battery dynamics is used by the controller, with control actions (i.e. the applied current) supplied at designated time intervals with a predefined sampling period t_s.

Specifically, the sequence $\mathbf{i}^\star_{[k,\,k+H]}$ of the optimal actions $\left[i^\star(t_k),\ \ldots,\ i^\star(t_{k+H-1})\right]$, computed at a given time-step t_k, over the upcoming H time steps, with H being the prediction horizon, are obtained as the solution of the subsequent optimization:

$$\mathbf{i}^\star_{[k,\,k+H]} \in \underset{\mathbf{i}_{[k,\,k+H]}}{\text{argmin}} \sum_{j=k}^{k+H-1} l(t_j), \tag{6}$$

with the cost $l(t_k)$ defined as:

$$l(t_k) = \alpha_z \cdot (z(t_{k+1}) - z_{\text{ref}})^2 + \alpha_i \cdot i(t_k)^2, \tag{7}$$

subject to the following constraints:

$$\text{model dynamics as defined by (1)–(4),} \tag{8a}$$

$$i_{\min} \leq i(t_j) \leq i_{\max}, \quad j \in \{k,\ \ldots,\ k+H-1\}, \tag{8b}$$

$$z_{\min} \leq z(t_j) \leq z_{\max}, \quad j \in \{k+1,\ \ldots,\ k+H\}, \tag{8c}$$

$$T_c(t_j) \leq T_{c,\max}, \quad j \in \{k+1,\ \ldots,\ k+H\}, \tag{8d}$$

$$T_s(t_j) \leq T_{s,\max}, \quad j \in \{k+1,\ \ldots,\ k+H\}, \tag{8e}$$

$$v(t_j) \leq v_{\max}, \quad j \in \{k+1,\ \ldots,\ k+H\}. \tag{8f}$$

Here, the weights α_z and α_i in the optimization function enable the prioritization of objectives in the control problem, such as the speed of reaching the target state, the minimization of control efforts, or a compromise between the two. The ranges $[i_{\min},\ i_{\max}]$ and $[z_{\min},\ z_{\max}]$ define the permissible zones for variables $i(t_k)$ and $z(t_k)$, respectively. Moreover, $T_{c,\max}$, $T_{s,\max}$ and $v_{\max}$ set upper limits for $T_c(t_k)$, $T_s(t_k)$ and $v(t_k)$, respectively.

In the simulation presented in the following sections, the weights in the objective function are set to $\alpha_z = 1$ and $\alpha_i = 10^{-6}$, while the operating ranges for the state of charge and the applied current are specified by $z_{\min} = 0$, $z_{\max} = 1$, $i_{\max} = -10A$, and $i_{\max} = 10A$. The upper bounds for voltage and temperatures are selected to be $v_{\max} = 4.2V$ and $T_{s,\max} = T_{c,\max} = 313.15K$. It is noteworthy that the target state of charge (z_{ref}) is variable, contingent upon specific user requirements.

C. STOCHASTIC MODEL PREDICTIVE CONTROL

An optimization problem such as the one presented in Subsection II-B can be effectively tackled using model predictive control strategies. Renowned for their proficiency in managing nonlinear processes bound by input and state constraints, MPC methods leverage a receding horizon framework, which has proven to be highly effective [46]. According to the receding horizon principles, the optimal input sequence $\mathbf{i}^\star_{[k,\,k+H]}$ is computed at every discrete instant t_k as the solution of the optimization presented in (6) - (8) over a finite time horizon H. Then, only the first control action $i^\star(t_k)$ is actually supplied to the system, while the remainder of the sequence is discarded. This iterative methodology provides a dynamic and adaptive approach, particularly beneficial for systems characterized by nonlinearity and operational constraints.

Nonetheless, when confronting real-world scenarios where model parameters are inherently uncertain, a deterministic MPC that relies solely on nominal models may lead to inevitable prediction inaccuracies. Such discrepancies can jeopardize the optimality and feasibility of the control actions across different possible scenarios. In these circumstances, model parameters are better represented as random variables, which consequently render the sequence of system states and outputs as stochastic. In light of this, it becomes evident that a stochastic control framework, capable of ensuring constraint satisfaction with a quantifiable level of confidence, is more appropriate. This approach acknowledges the probabilistic nature of system behaviors, thereby offering a more robust and reliable control strategy that aligns with the realities of uncertain system parameters.

The optimization solved at each time step by the stochastic MPC can be expressed as:

$$\mathbf{i}^{\star}_{[k,\,k+H]} \in \underset{\mathbf{i}_{[k,\,k+H]}}{\operatorname{argmin}}\, \mathbb{E}\left[\sum_{j=k}^{k+H-1} l(t_j)\right], \tag{9}$$

where $\mathbb{E}[\cdot]$ is the expectation over the system parameters and with the cost $l(t_k)$ defined as in (7), subject to (8a), (8b), and to the following probabilistic constraints on the generic variable $\psi(t_j)$ with $\psi \in \{z,\, T_c,\, T_s,\, v\}$:

$$\mathrm{P}\left[\psi_{\min} - \psi(t_j) \le 0\right] \ge \beta^{\psi}_{\min}, \tag{10a}$$

$$\mathrm{P}\left[\psi(t_j) - \psi_{\max} \le 0\right] \ge \beta^{\psi}_{\max}, \tag{10b}$$

for $j \in \{k+1,\, \dots,\, k+H\}$, with $\mathrm{P}[\cdot]$ denoting the probability function, $\beta^{\psi}_{\min}$ and $\beta^{\psi}_{\max}$, both within the range of $(0,\, 1)$ represent the minimum desired probabilities for satisfying the bounds for the variables under uncertainty. Notably, $\psi_{\min}$ and $\psi_{\max}$ are the lower and upper limits of the considered variables.

Addressing the complexities of stochastic optimization in an efficient manner often involves converting it into a deterministic optimal control problem. This conversion is achievable by reinterpreting probabilistic constraints, such as those expressed in (10), into distributionally robust probabilistic constraints. As substantiated by Theorem 3.1 in [47], these robust probabilistic constraints are effectively equivalent to specific deterministic second-order cone constraints, delineated as:

$$k^{\psi}_{\min}\, \mathrm{var}[\psi_{\min} - \psi(t_j)] + \mathbb{E}[\psi_{\min} - \psi(t_j)] \le \quad 0, \tag{11a}$$

$$k^{\psi}_{\max}\, \mathrm{var}[\psi(t_j) - \psi_{\max}] + \mathbb{E}[\psi(t_j) - \psi_{\max}] \le \quad 0, \tag{11b}$$

for $j \in \{k+1,\, \dots,\, k+H\}$, where $\mathrm{var}[\cdot]$ represents the variance and the coefficients $k^{\psi}_{\min}$ and $k^{\psi}_{\max}$ are determined based on the guaranteed probability levels, following the relations $k^{\psi}_{\min} = \sqrt{\frac{\beta^{\psi}_{\min}}{1-\beta^{\psi}_{\min}}}$ and $k^{\psi}_{\max} = \sqrt{\frac{\beta^{\psi}_{\max}}{1-\beta^{\psi}_{\max}}}$.

The pivotal step of replacing the probabilistic constraints detailed in (10) with the deterministic second-order cone constraints presented in (11) facilitates the transformation of the stochastic optimization in a deterministic framework. This substitution, however, necessitates the computation of mean and variance of the model variables across the prediction horizon. Achieving this necessitates additional computational efforts to propagate the model uncertainties over the time horizon, a process that can be executed using for instance Monte Carlo simulations, as discussed in references like [22].

While this transformation brings structure and tractability to the problem, it comes at the cost of increased computational complexity. This heightened complexity poses challenges for the online execution of the control methodology, as the computational demands may exceed the timely constraints required for effective system operation. In response to this computational challenge, the remainder of this paper introduces an innovative approach rooted in imitation learning to approximate the solution of the stochastic MPC. In particular, the optimal strategy followed by the stochastic MPC algorithm discussed above can be described through a policy $\pi^{\star}$ mapping system states into optimal actions as follows:

$$\pi^{\star}:\ \mathbf{s}_{t_k} \to i^{\star}(t_k) \tag{12}$$

with $\mathbf{s}_{t_k}$ being the states vector, composed by crucial electrochemical and thermal variables measured at the time t_k, as well as the desired target for the state of charge. The optimal strategy $\pi^{\star}$ can be regarded as the "expert" policy within the context of imitation learning.

III. IMITATION LEARNING APPROXIMATION

While the effectiveness of stochastic predictive control is indeed compelling, its practical deployment, especially in scenarios demanding real-time execution, is fraught with substantial challenges. Foremost among these is the necessity to retrieve the optimal input online by solving a constrained optimization, a requirement that can impose severe computational demands. This computational intensity is notably pronounced in systems characterized by expansive state spaces or intricate, nonlinear dynamics. A particular aspect that exacerbates this complexity is the need to propagate the uncertainty in the parameters of a nonlinear model across the prediction horizon. This process involves forecasting the future states of the system under varying uncertain conditions, a task that becomes increasingly intricate and computationally demanding with the model complexity and the prediction horizon size. The computational load intensifies further in applications that necessitate high-frequency control decisions, such as those commonly encountered in battery charging processes.

Given these constraints, imitation learning is proposed in this paper as a possible solution to reduce the computational load associated with the control algorithm. The operational requisites of this framework are substantially less demanding in real-time scenarios, as it predominantly involves conducting a prediction using a machine learning model, rather than repetitively solving an optimal control problem.

Firstly, in Subsection III-A, behavioral cloning is explored, which is a simple and straightforward implementation of imitation learning based on the principles of supervised learning. Despite its intuitive nature, this approach is not free from challenges. A significant limitation is the issue of distributional shift, a scenario where the states encountered executing the learned strategy diverges from those experienced by the expert, with a consequent decline in performance. To address this challenge, the focus shifts to the DAgger methodology in Subsection III-B. This technique is designed to effectively counter the detrimental impacts of distributional shift, enhancing both the resilience and effectiveness of the imitation learning process.

A. BEHAVIORAL CLONING

Imitation learning aims to emulate the actions of an expert without a deep understanding of the system's underlying mechanics. In essence, a machine learning model is trained to closely replicate expert behavior by learning from the expert's demonstrations. Imitation learning's core goal is to develop a policy, symbolized as $\pi_{\boldsymbol{\theta}}$, that mirrors an expert's policy $\pi^\star$ as closely as possible.

Among the various techniques in imitation learning, behavioral cloning stands out as the simplest, essentially functioning as a supervised learning task. This method is centered around forging a direct map from the state space $\mathcal{S}$ to the action space $\mathcal{A}$, utilizing a dataset $\mathcal{D}$ comprising state-action pairs $(s, a) \in \mathcal{S} \times \mathcal{A}$ encountered when executing the experienced policy in the real world or in a relevant simulator. The objective is to train a policy, represented as $\pi_{\boldsymbol{\theta}} : \mathcal{S} \to \mathcal{A}$ with parameters $\boldsymbol{\theta}$, to echo the expert's actions as faithfully as possible. This involves finding the optimal vector $\hat{\boldsymbol{\theta}}$ to ensure that the resultant policy $\pi_{\hat{\boldsymbol{\theta}}}$ closely aligns with the expert's strategy.

Behavioral cloning is considered a special case of supervised learning since the machine learning model is trained to fit the function that maps states (inputs) into actions (outputs) based on the dataset $\mathcal{D}$. The alignment between the mimicking policy and the experienced one is achieved by minimizing their mean square distance across the dataset $\mathcal{D}$, as represented in the equation:

$$L(\boldsymbol{\theta}) = \frac{1}{N} \sum_{k=1}^{N} \left[\pi_{\boldsymbol{\theta}}(\mathbf{s}_k) - \pi^\star(\mathbf{s}_k)\right]^2 . \tag{13}$$

Despite its intuitive appeal and simplicity, behavioral cloning is not devoid of challenges. A significant drawback is its susceptibility to distributional shift, a phenomenon where the states encountered by the learned policy differ from the ones which are used for training. As the learned policy diverges from the expert's trajectory and explores these new states, its performance can significantly degrade due to a lack of training in these areas. This issue can be further compounded over time as each deviation can lead the policy into increasingly unfamiliar territory, resulting in a cascade of errors.

B. DAGGER

The issue of distributional shift motivated further research to develop innovative solutions, including the so-called Dataset Aggregation algorithm, commonly referred to as DAgger, which stands out as a prominent strategy. Initially introduced by the authors in [36], DAgger is an algorithm specifically crafted to address the distributional shift challenge, which often plagues behavioral cloning. This iterative technique ingeniously amalgamates insights from both the learned and expert policies, thereby progressively enlarging the dataset used for training.

Within this framework, during each iteration $i = 1, \ldots, n_D$, the system interacts with the environment and collects several state trajectories by applying a hybrid policy π_i. Such policy tactfully integrates the most recently learned policy $\pi_{\hat{\boldsymbol{\theta}}_{i-1}}$ with the expert's one $\pi^\star$:

$$\pi_i(\mathbf{s}_{t_k}) = \begin{cases} \pi^\star(\mathbf{s}_{t_k}), & \text{with probability } \beta_i \\ \pi_{\hat{\boldsymbol{\theta}}_{i-1}}(\mathbf{s}_{t_k}), & \text{with probability } 1 - \beta_i \end{cases} \tag{14}$$

with $\beta_i \in [0, 1]$ and $1 - \beta_i$ signifying the probability of applying the expert's action or the learner's one, respectively. Note that $\pi_{\hat{\boldsymbol{\theta}}_0}$ is initialized through behavioral cloning. Moreover, it is important to highlight that over time, the reliance on the expert policy gradually diminishes, permitting the learner agent to increasingly act autonomously.

Following data collection through the interactions with the environment under the blended policy, the subsequent step involves refining the learned policy. Specifically, each state $\mathbf{s}$ experienced during the execution of the hybrid strategy π_i is labeled with the action $\pi^\star(\mathbf{s})$ that the expert would take in that state. An aggregated dataset is therefore built by adding to the training dataset of original demonstrations the labeled pairs $(\mathbf{s}, \pi^\star(\mathbf{s}))$. The improved learner's policy $\pi_{\hat{\boldsymbol{\theta}}_i}$ is then computed by optimizing the loss function in (13) over the cumulatively aggregated dataset. Consequently, the policy of the learner is progressively fine-tuned to accurately mirror the expert's, benefiting from the comprehensive and varied collection of state-action experiences accumulated thus far.

This methodical and incrementally optimized approach provides an active learning environment that slowly synchronizes the state distribution faced by both the learner and the expert. This approach allows the learners to continuously evolve, enhancing their proficiency in emulating the expert and managing an extensive state space. This significant improvement over traditional behavioral cloning underlines DAgger's effectiveness in addressing the complex challenge of distributional shift.

IV. RESULTS

The major results of this paper are presented below. Initially, Subsection IV-A elucidates the performance of the stochastic predictive control strategy, as delineated in Subsection II-C. This analysis is conducted across various scenarios, where battery parameters are randomly derived from predefined distributions. Emphasis is placed on evaluating adherence to safety constraints, to ensure that the stochastic MPC maintains optimal and secure operation under conditions of uncertainty.

Following this, Subsection IV-B provides a comprehensive validation of the DAgger-based methodology. This examination is twofold: it assesses the accuracy with which the DAgger approach approximates the decision-making process of the stochastic predictive controller—the expert agent—and it quantifies the reduction in computational load facilitated by the adoption of the imitation learning strategy. Findings indicate a significant reduction in computational burden, alongside maintaining close performance to that of the expert agent. This dual assessment highlights the effectiveness of the

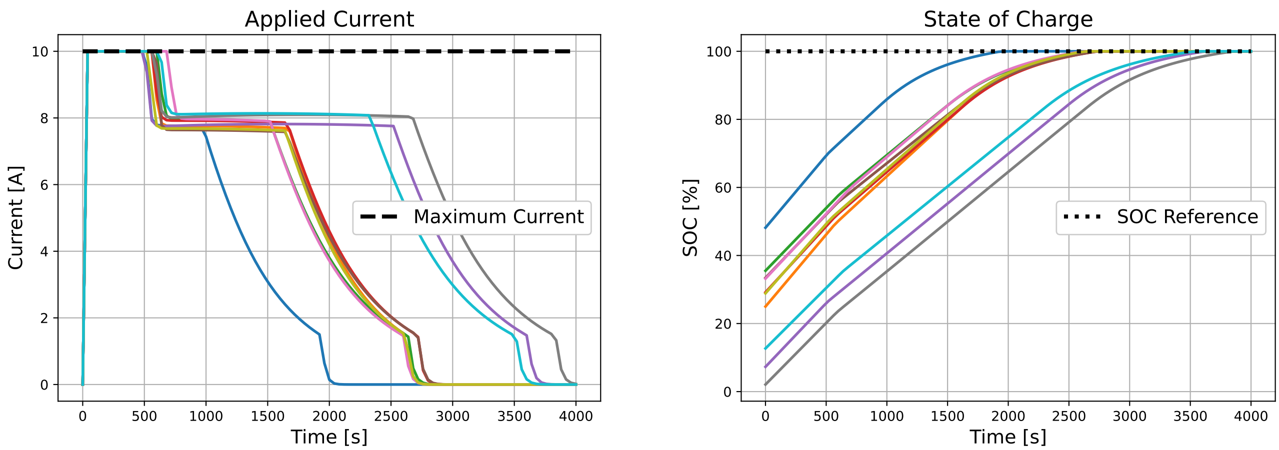


FIGURE 2. This figure presents the dynamic profiles of key variables during the battery charging process across a selection of 10 different scenarios, each characterized by unique values of cell capacity and SEI resistance sampled according to predefined distributions. (Left) Applied current profiles, highlighting the compliance with operational limits. (Right) SOC profiles, demonstrating the controller's ability to achieve a full charge in every scenario.

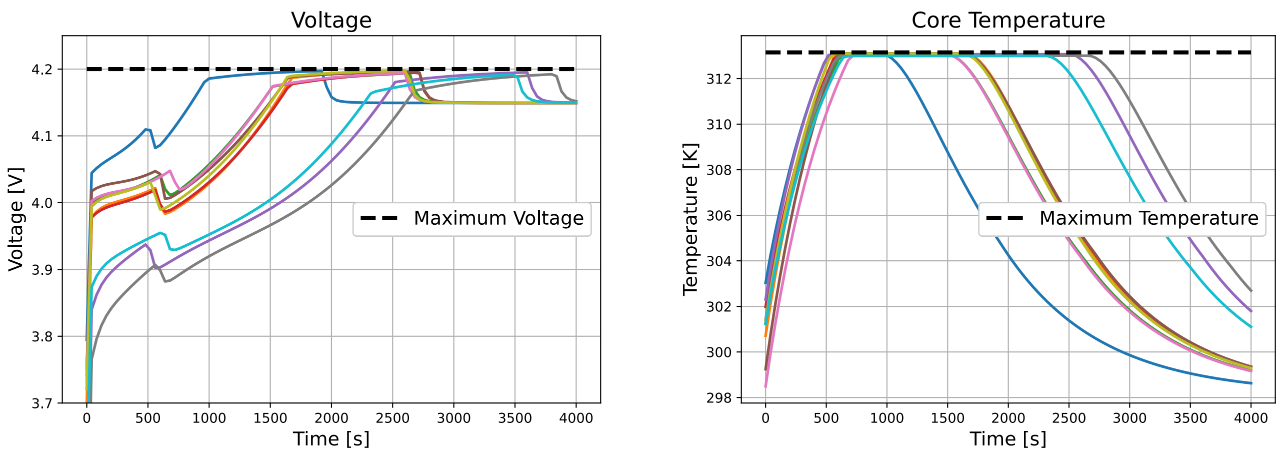


FIGURE 3. This figure showcases the variation in critical variables throughout the battery charging cycle, captured across ten unique scenarios. These scenarios are differentiated by specific cell capacity and SEI resistance values, drawn from designated probability distributions. (Left) Voltage profiles across the charging process, showcasing that the operational boundaries are respected without any violations of safety constraints. (Right) Core temperature profiles, illustrating that despite nearing threshold limits, temperatures remain within safe operational bounds, thus ensuring the integrity and safety of the battery system.

proposed framework in achieving a balance between expert-level decision-making accuracy and a considerable decrease in computational demands.

Finally, the implementation details are briefly described in Subsection IV-C, followed by practical deployment considerations outlined in Subsection IV-D.

A. VALIDATION OF STOCHASTIC MPC

The validation of the stochastic predictive controller outlined in Section II-C, which is considered the expert agent in this study, was conducted through an extensive series of battery charging simulations. These simulations were performed under various samples of battery parameters to confirm the effectiveness of the stochastic MPC as a suitable controller. Specifically, 100 distinct scenarios were examined, each characterized by different values of cell capacity and SEI resistance, as dictated by the distributions described in (5). For all scenarios, a SoC reference of 1, indicative of a fully charged battery, was maintained, while the safety constraint bounds were aligned with those specified in Subsection II-B. The initial battery's SoC was sampled from a uniform distribution within the interval [0, 0.5]. Finally, the horizon of the prediction is set to $H = 8$, while the sampling time is fixed at $t_s = 40s$.

For illustrative clarity and without losing generality, Figure 2 and Figure 3 present the profiles of the key variables from a representative subset of 10 scenarios out of the total 100 conducted simulations. Notably, in every simulation, the SOC successfully reaches the target, aided by an applied current that not only complies with the limits but also progressively diminishes to zero as the battery nears full charge. Furthermore, despite operating close to their respective thresholds, both the voltage and core temperature profiles consistently adhere to the safety limits. This proximity to constraint boundaries is characteristic of constrained optimal control problems, where the optimal trajectory is typically the one that achieves the objective swiftly without breaching any constraints. Interestingly, the observed profiles adhere to the typical pattern of constant current, constant temperature and constant voltage, which has been identified as generally optimal in numerous studies within the literature on battery management (see for instance [48]).

B. IMITATION LEARNING RESULTS

In the following, the capability of a DAgger-based agent to accurately approximate the actions of a stochastic predictive controller is evaluated. Specifically, the expert agent considered is the stochastic MPC designed to solve the optimal charging problem detailed in Subsection II-B and validated in Subsection IV-A.

1) NEURAL NETWORK ARCHITECTURE

The model $\pi_{\boldsymbol{\theta}}$ used to approximate the expert agent consists of a deep neural network. Neural networks have shown exceptional efficacy in imitation learning tasks due to their powerful representational capabilities. Given the measurability of battery states, a feed-forward neural network is employed effectively, negating the need for recurrent structures required in partially observable settings. The chosen architecture comprises 3 fully connected hidden layers with 100, 50, and 10 neurons, respectively. All hidden layers utilize a rectified linear unit activation function, whereas the output layer employs a hyperbolic tangent function to naturally limit the network's output to a continuous action range. The neural network is trained using a stochastic gradient descent algorithm for 50 epochs, incorporating an early stopping criterion based on validation performance, halting after 3 epochs without improvement. Training employs the Adam optimizer, featuring a learning rate of $\alpha = 0.0005$, is used during training.

2) DATASET GENERATION

Generating a suitable training dataset $\mathcal{D}$, consisting of expert demonstrations, is a critical aspect of imitation learning. In a behavioral cloning framework, this dataset could be compiled by simply applying the stochastic MPC algorithm to various initial conditions and parameter configurations of the battery model. However, the imitating agent trained on such a dataset would exhibit the well-known distributional issue, due to the accumulation of prediction errors. For this reason, in this paper the dataset $\mathcal{D}$ is generated according to the DAgger paradigm, which consists of iteratively collecting data applying the imitating agent and asking the expert for corrections.

At the basis of this data collection process lies the concept of episode, which is defined as a single simulation of battery charging using stochastic MPC over 50 time steps with random battery parameters and initial conditions. The initial policy $\pi_{\hat{\boldsymbol{\theta}}_0}$ is derived by supervised training on data collected from 500 initial episodes applying the expert policy (i.e. the stochastic MPC). Then, within the DAgger framework, 10 iterations (i.e. $n_D = 10$), each comprising 100 episodes, are conducted, with $\beta_i = 0$ for $i = 1, \ldots, n_D$. The reduction in the root mean square error of the policy $\pi_{\hat{\boldsymbol{\theta}}_i}$ during training across DAgger iterations is illustrated in Figure 4, indicating significant improvement from an initial error of 0.11.

It is important to underline that, despite the time-intensive nature of collecting expert demonstrations and aggregating the dataset for repeatedly training the neural network, this process incurs only an offline computational cost. This is negligible compared to the substantial online computational savings achieved by inferring optimal actions through the machine learning model instead of solving optimization problems directly.

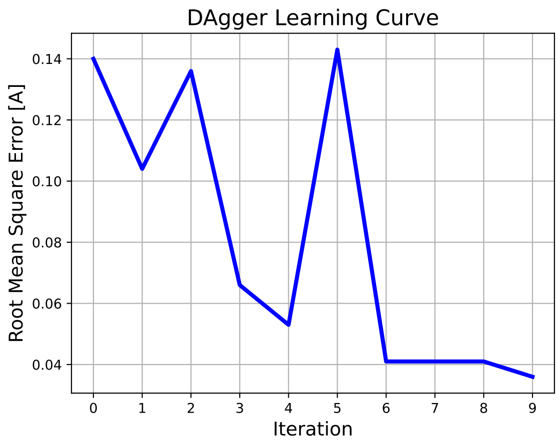


FIGURE 4. This graph depicts the decrease in root mean square error over DAgger iterations, showcasing the progressive improvement in the machine learning model's capability to approximate the stochastic MPC's decisions.

3) APPROXIMATION PERFORMANCE

The efficacy of the DAgger algorithm in mimicking the stochastic MPC is demonstrated through its performance in a single charging episode. The neural network's ability to generate a current profile closely mirroring that of the stochastic MPC is evident, as depicted in the left part of Figure 5. This leads to similar profiles also for the state of charge, thus implying the effectiveness of the DAgger algorithm in successfully stabilizing the battery's SoC at the reference point, as it can be noticed by the right panel in Figure 5. Moreover, as demonstrated in Figure 6, the proposed approach satisfies the constraints for both voltage and temperature, ensuring a safe charging protocol for batteries with parameters in the considered range of variability.

Finally, Figure 7 presents the error distribution generated by the DAgger algorithm when estimating the optimal charging current across 50 episodes. This distribution highlights a mean error of 0.2 mA and a standard deviation of 100 mA, convincingly illustrating the proposed method's accuracy in approximating the optimal charging profile under uncertain parameters across a variety of scenarios.

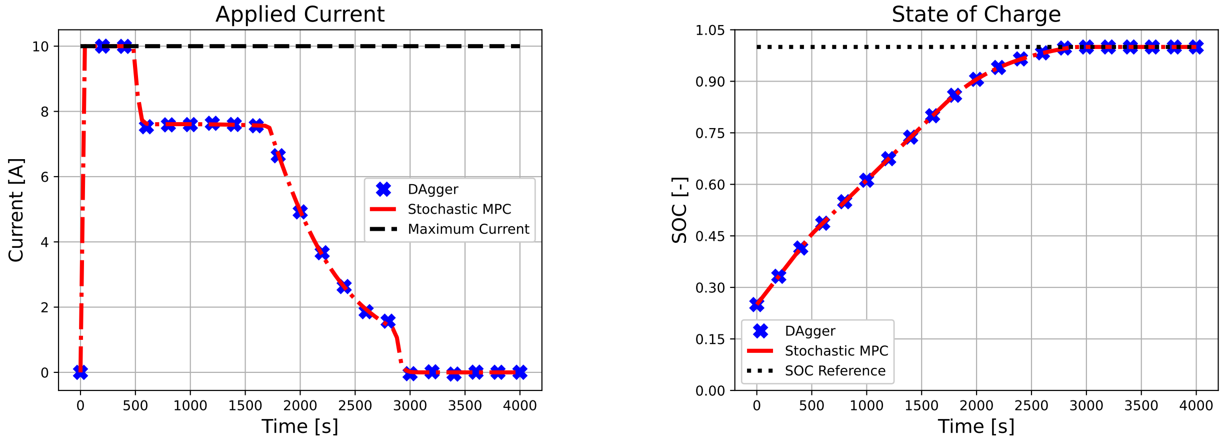


FIGURE 5. This figure displays the variation in crucial variables during a single battery charging episode, comparing the outcomes of the DAgger algorithm with those of the stochastic MPC. (Left) Applied current profiles reveal the DAgger algorithm's proficiency in mirroring the expert's charging strategy, showcasing its accuracy in replicating the stochastic MPC's actions. (Right) State-of-charge profiles highlight the algorithm's effectiveness in achieving the target SOC levels, emphasizing its capability to precisely emulate the desired charging outcomes.

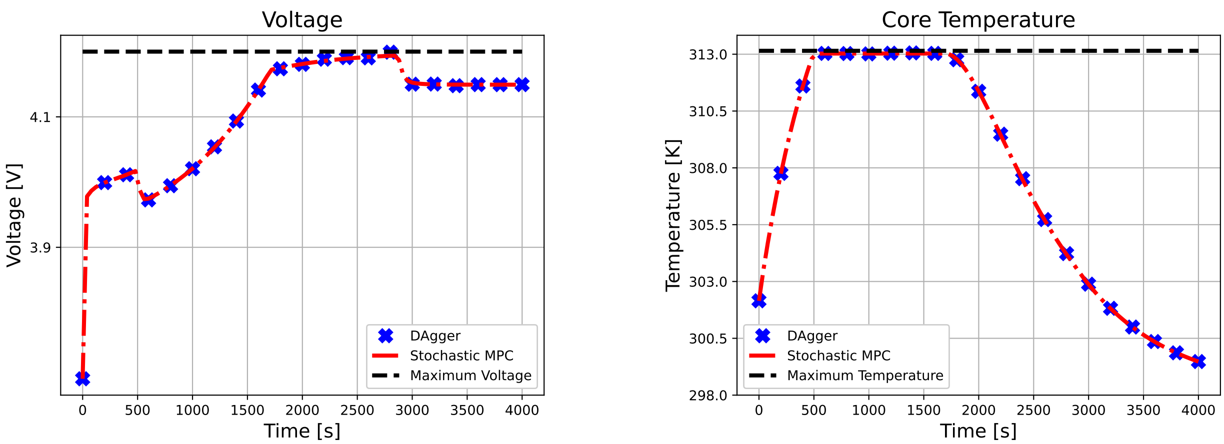


FIGURE 6. This figure illustrates the regulation of voltage and temperature profiles during a charging episode, comparing the DAgger algorithm's performance with that of the stochastic MPC. (Left) Voltage profiles are shown to be maintained within safe operational limits by the DAgger algorithm, reflecting its alignment with the expert agent's control strategy. (Right) Core and surface temperature profiles underscore the algorithm's compliance with thermal constraints, showcasing its ability to prevent overheating by closely adhering to the stochastic MPC's thermal management approach. These comparisons highlight the DAgger algorithm's capability to ensure safety and efficiency in battery charging.

4) COMPUTATIONAL COST ANALYSIS

The online computational demands of both the stochastic MPC and the DAgger algorithm are analyzed across varying prediction horizons $H \in \{2, 6, 10, 14, 18\}$. For each horizon value, the time required to compute the optimal action is recorded over 10 episodes, each comprising 50 steps, providing insights into the computational effort involved. Figure 8 contrasts the computational times, with stochastic MPC displaying a superlinear increase in time with the horizon length, whereas the DAgger algorithm maintains a nearly constant mean time of 0.039 seconds. This disparity highlights the efficiency of DAgger, which relies on a straightforward neural network evaluation for decision-making, as opposed to the computationally intensive optimization required by stochastic MPC at every step. The consistent computational efficiency of the DAgger algorithm can be attributed primarily to its operational mechanism. Specifically, during online execution, the algorithm merely requires the evaluation of a nonlinear function—the neural network—whose parameters have been meticulously trained offline. This evaluation process is considerably faster than solving a constrained optimization problem. As a result, the computational time for the DAgger approach remains nearly constant regardless of the length of the prediction horizon, underscoring its efficiency in real-time applications.

C. IMPLEMENTATION SPECIFICATIONS

This study's simulations were executed on a personal computing system equipped with Windows 11, powered by an i7-1260P CPU, and supplemented with 32 GB of memory. Python 3.7 served as the foundation for the implementation process. The construction and training of the deep learning framework were performed using TensorFlow 2.0. Additionally, CasADi [49] was utilized for integrating the model's equations and addressing the stochastic optimal control problem.

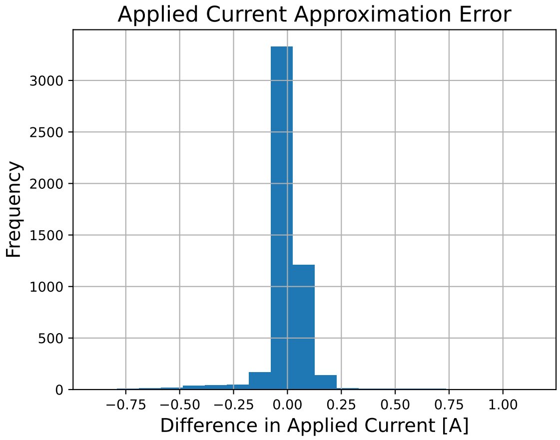


FIGURE 7. The figure provides a distribution analysis of the errors in approximating the optimal charging current across 100 episodes by the DAgger algorithm. It quantifies the algorithm's performance variability, offering insights into the consistency and reliability of the approximation over multiple scenarios.

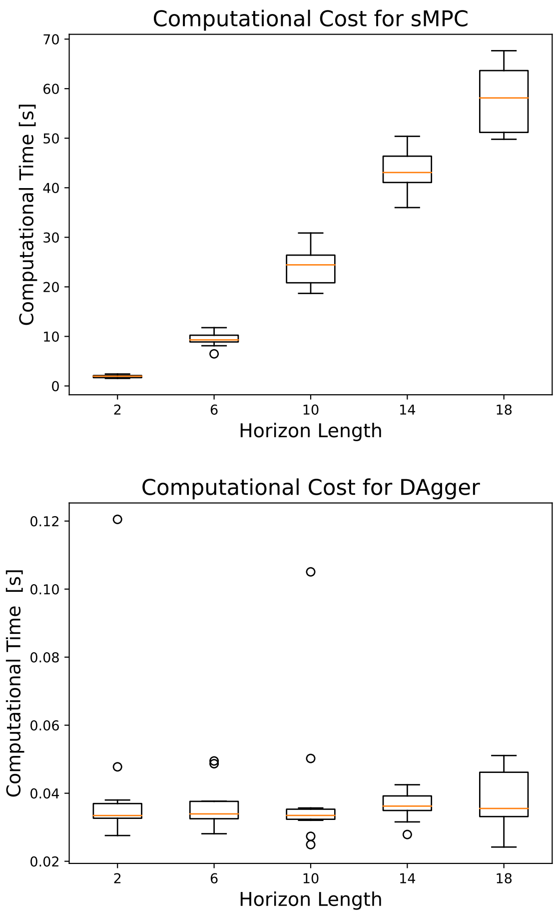


FIGURE 8. This figure contrasts the computational time required by the stochastic MPC and the DAgger algorithm across different prediction horizons. While the stochastic MPC shows a superlinear increase in computational time with horizon length, the DAgger algorithm demonstrates a nearly constant computational effort, highlighting its efficiency and scalability in operational environments.

D. PRACTICAL DEPLOYMENT CONSIDERATIONS

While the presented study primarily focuses on the simulation-based evaluation of the proposed imitation learning framework, practical deployment on physical hardware introduces additional considerations. Specifically, the computational complexity of the trained policy and its feasibility for real-time execution on embedded systems are key factors to assess. The implementation of the neural network-based policy on hardware-constrained environments, such as microcontrollers, requires careful optimization. The inference speed of the model is a crucial aspect, as real-time decision-making is essential in battery management applications. Although deep learning inference can be computationally demanding, techniques such as model quantization, pruning, and hardware acceleration (e.g., using Tensor Processing Units or optimized deep learning libraries) could facilitate deployment on resource-limited devices. Moreover, memory constraints must be considered. The storage requirements of the trained model should be evaluated against the available memory of the target hardware. Lightweight neural architectures or knowledge distillation techniques could be employed to reduce memory overhead while maintaining accuracy.

V. CONCLUSION

This study presented a novel approach leveraging imitation learning to approximate stochastic predictive control strategies for lithium-ion battery charging. The proposed framework, based on the DAgger algorithm, offers a scalable solution to address the computational challenges associated with stochastic MPC, achieving expert-level decision-making accuracy while reducing online computational costs. The findings highlight the potential of advanced imitation learning techniques to improve battery management systems under uncertainty.

A. THEORETICAL AND PRACTICAL IMPLICATIONS

This work provides valuable insights into both theoretical advancements and practical applications. From a theoretical standpoint, it bridges the gap between advanced control strategies and imitation learning, establishing a methodology that retains the robustness of stochastic MPC while improving computational efficiency. On the practical side, the proposed approach enables the real-time application of computationally demanding control strategies, paving the way for smarter, safer, and more efficient battery management systems. Beyond battery applications, this framework could be extended to other domains, such as energy storage systems and robotics, where managing uncertainty and meeting real-time constraints are crucial.

B. MAJOR TAKEAWAYS

Several key insights emerge from this study. Firstly, imitation learning, particularly the DAgger algorithm, has proven to be an effective approach for approximating complex control strategies, significantly reducing online computational effort. Additionally, the study highlights the inherent trade-off between accuracy and computational efficiency, demonstrating that shifting complexity to offline training facilitates real-time deployment. Finally, a neural network-based policy trained through imitation learning has been shown to effectively replicate expert-level performance while adhering to safety and operational constraints in highly uncertain environments.

C. POTENTIAL LIMITATIONS

The proposed methodology has certain limitations that warrant consideration. First, the offline computational cost associated with the DAgger algorithm can grow significantly, particularly for large systems and long prediction horizons. This is due to the repeated need for expert intervention to generate training data during the offline phase. While this upfront expense can be substantial, it is often an acceptable trade-off in practical applications, as the resulting policy eliminates the need for repeated constrained optimization during online deployment. This allows for real-time control in complex and uncertain environments, where computational efficiency is critical. Second, the effectiveness of the overall framework heavily relies on the accuracy of the estimated range of variability in battery parameters and their associated uncertainty distributions. Although this limitation is not directly tied to the DAgger algorithm itself, any inaccuracies in these estimations could impair the performance of the stochastic predictive control strategy, potentially diminishing the efficacy of the entire pipeline due to the poor quality of training data. Addressing this challenge may require further advancements in parameter estimation techniques to improve robustness under real-world conditions.

D. FUTURE RESEARCH DIRECTIONS

Future research could explore extending this framework to incorporate additional constraints, such as degradation-aware control strategies, to improve battery lifespan and reliability. Additionally, applying imitation learning to multi-agent or hierarchical battery systems could help address coordination challenges in large-scale storage applications. Furthermore, the adaptation of the proposed approach for real-world deployment could be explored by testing its performance on embedded hardware and improving the model to balance computational efficiency with accuracy. Finally, extending this methodology to other complex control tasks, including energy management, robotics, and smart grid systems, could further refine the algorithm, enhancing its performance and broadening its applicability to domains where computational efficiency and control accuracy are critical.

REFERENCES

[1] L. Lu, X. Han, J. Li, J. Hua, and M. Ouyang, "A review on the key issues for lithium-ion battery management in electric vehicles," *J. Power Sources*, vol. 226, pp. 272–288, Mar. 2013.

[2] N. A. Chaturvedi, R. Klein, J. Christensen, J. Ahmed, and A. Kojic, "Algorithms for advanced battery-management systems," *IEEE Control Syst. Mag.*, vol. 30, no. 3, pp. 49–68, Jun. 2010.

[3] P. García, J. P. Torreglosa, L. M. Fernández, and F. Jurado, "Control strategies for high-power electric vehicles powered by hydrogen fuel cell, battery and supercapacitor," *Expert Syst. Appl.*, vol. 40, no. 12, pp. 4791–4804, Sep. 2013.

[4] K. Liu, C. Zou, K. Li, and T. Wik, "Charging pattern optimization for lithium-ion batteries with an electrothermal-aging model," *IEEE Trans. Ind. Informat.*, vol. 14, no. 12, pp. 5463–5474, Dec. 2018.

[5] M. Sheikhan, R. Pardis, and D. Gharavian, "State of charge neural computational models for high energy density batteries in electric vehicles," *Neural Comput. Appl.*, vol. 22, no. 6, pp. 1171–1180, May 2013.

[6] Y. Qin, S. Adams, and C. Yuen, "Transfer learning-based state of charge estimation for lithium-ion battery at varying ambient temperatures," *IEEE Trans. Ind. Informat.*, vol. 17, no. 11, pp. 7304–7315, Nov. 2021.

[7] A. Kara, "A data-driven approach based on deep neural networks for lithium-ion battery prognostics," *Neural Comput. Appl.*, vol. 33, no. 20, pp. 13525–13538, Oct. 2021.

[8] Y. Zhang, Y.-F. Li, M. Zhang, and H. Wang, "A novel health indicator by dominant invariant subspace on Grassmann manifold for state of health assessment of lithium-ion battery," *Eng. Appl. Artif. Intell.*, vol. 130, Apr. 2024, Art. no. 107698.

[9] P. D. Nostro, G. Goldbeck, F. Kienberger, M. Moertelmaier, A. Pozzi, N. Al-Zubaidi-R-Smith, and D. Toti, "Battery testing ontology: An EMMO-based semantic framework for representing knowledge in battery testing and battery quality control," *Comput. Ind.*, vol. 164, Jan. 2025, Art. no. 104203.

[10] M. Schwenzer, M. Ay, T. Bergs, and D. Abel, "Review on model predictive control: An engineering perspective," *Int. J. Adv. Manuf. Technol.*, vol. 117, nos. 5–6, pp. 1327–1349, Aug. 2021.

[11] L. McCurlie, M. Preindl, and A. Emadi, "Fast model predictive control for redistributive lithium-ion battery balancing," *IEEE Trans. Ind. Electron.*, vol. 64, no. 2, pp. 1350–1357, Feb. 2017.

[12] G. Li, J. Zhang, and H. He, "Battery SOC constraint comparison for predictive energy management of plug-in hybrid electric bus," *Appl. Energy*, vol. 194, pp. 578–587, May 2017.

[13] C. Zou, C. Manzie, and D. Nešić, "Model predictive control for lithium-ion battery optimal charging," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 2, pp. 947–957, Apr. 2018.

[14] Y. Xie, C. Wang, X. Hu, X. Lin, Y. Zhang, and W. Li, "An MPC-based control strategy for electric vehicle battery cooling considering energy saving and battery lifespan," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14657–14673, Dec. 2020.

[15] A. Pozzi and D. Toti, "Lexicographic model predictive control strategy in ageing-aware optimal charging procedure for lithium-ion batteries," *Comput. Chem. Eng.*, vol. 163, Jul. 2022, Art. no. 107847.

[16] D. Locatelli, D. M. Raimondo, Z. Khalik, H. J. Bergveld, and M. C. F. Donkers, "Closed-loop optimal ageing-aware charging of Li-ion batteries using a surrogate model," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 7140–7146, 2023.

[17] E. Martínez-Rosas, R. Vásquez-Medrano, and A. Flores-Tlacuahuac, "Modeling and simulation of lithium-ion batteries," *Comput. Chem. Eng.*, vol. 35, no. 9, pp. 1937–1948, Jun. 2011.

[18] M. Dubarry, N. Vuillaume, and B. Y. Liaw, "Origins and accommodation of cell variations in Li-ion battery pack modeling," *Int. J. Energy Res.*, vol. 34, no. 2, pp. 216–231, Feb. 2010.

[19] Q. Zhang, W. Deng, and G. Li, "Stochastic control of predictive power management for battery/supercapacitor hybrid energy storage systems of electric vehicles," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3023–3030, Jul. 2018.

[20] 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 Trans. Power Syst.*, vol. 33, no. 4, pp. 4397–4406, Jul. 2018.

[21] A. Groß, C. Wittwer, and M. Diehl, "Stochastic model predictive control of photovoltaic battery systems using a probabilistic forecast model," *Eur. J. Control*, vol. 56, pp. 254–264, Nov. 2020.

[22] A. Pozzi and D. M. Raimondo, "Stochastic model predictive control for optimal charging of electric vehicles battery packs," *J. Energy Storage*, vol. 55, Nov. 2022, Art. no. 105332.

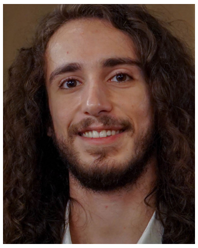
[23] A. Mesbah, "Stochastic model predictive control: An overview and perspectives for future research," *IEEE Control Syst. Mag.*, vol. 36, no. 6, pp. 30–44, Dec. 2016.

[24] M. Mammarella, T. Alamo, F. Dabbene, and M. Lorenzen, "Computationally efficient stochastic MPC: A probabilistic scaling approach," in *Proc. IEEE Conf. Control Technol. Appl. (CCTA)*, Aug. 2020, pp. 25–30.

[25] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed., Cambridge, U.K.: Cambridge Univ. Press, 2018.

[26] A. Hussein, M. M. Gaber, E. Elyan, and C. Jayne, "Imitation learning: A survey of learning methods," *ACM Comput. Surveys*, vol. 50, no. 2, pp. 1–35, 2017.
[27] W. Li, H. Cui, T. Nemeth, J. Jansen, C. Ünlübayir, Z. Wei, L. Zhang, Z. Wang, J. Ruan, H. Dai, X. Wei, and D. U. Sauer, "Deep reinforcement learning-based energy management of hybrid battery systems in electric vehicles," *J. Energy Storage*, vol. 36, Apr. 2021, Art. no. 102355.
[28] M. Trimboli and L. Avila, "Optimal battery charge with safe exploration," *Expert Syst. Appl.*, vol. 237, Mar. 2024, Art. no. 121697.
[29] M. Jiao, D. Wang, Y. Yang, and F. Liu, "More intelligent and robust estimation of battery state-of-charge with an improved regularized extreme learning machine," *Eng. Appl. Artif. Intell.*, vol. 104, Sep. 2021, Art. no. 104407.
[30] T. Alsuwian, S. Ansari, M. A. A. M. Zainuri, A. Ayob, A. Hussain, M. S. H. Lipu, A. R. H. Alhawari, A. H. M. Almawgani, S. Almasabi, and A. T. Hindi, "A review of expert hybrid and co-estimation techniques for SOH and RUL estimation in battery management system with electric vehicle application," *Expert Syst. Appl.*, vol. 246, Jul. 2024, Art. no. 123123.
[31] P. Takyi-Aninakwa, S. Wang, G. Liu, A. N. Bage, F. Masahudu, and J. M. Guerrero, "An enhanced lithium-ion battery state-of-charge estimation method using long short-term memory with an adaptive state update filter incorporating battery parameters," *Eng. Appl. Artif. Intell.*, vol. 132, Jun. 2024, Art. no. 107946.
[32] A. Pozzi, S. Moura, and D. Toti, "A deep learning-based predictive controller for the optimal charging of a lithium-ion cell with non-measurable states," *Comput. Chem. Eng.*, vol. 173, May 2023, Art. no. 108222.
[33] A. Pozzi, E. Barbierato, and D. Toti, "Optimizing battery charging using neural networks in the presence of unknown states and parameters," *Sensors*, vol. 23, no. 9, p. 4404, Apr. 2023.
[34] M. Zare, P. M. Kebria, A. Khosravi, and S. Nahavandi, "A survey of imitation learning: Algorithms, recent developments, and challenges," *IEEE Trans. Cybern.*, vol. 54, no. 12, pp. 7173–7186, Dec. 2024.
[35] S. Ross and D. Bagnell, "Efficient reductions for imitation learning," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, Mar. 2010, pp. 661–668.
[36] S. Ross, G. J. Gordon, and J. A. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, Jan. 2010, pp. 627–635.
[37] A. Pozzi and D. Toti, "Imitation learning for agnostic battery charging: A DAGGER-based approach," *IEEE Access*, vol. 11, pp. 115190–115203, 2023.
[38] S. Matrone, A. Pozzi, E. Ogliari, and S. Leva, "Deep learning-based predictive control for optimal battery management in microgrids," *IEEE Access*, vol. 12, pp. 141580–141593, 2024.
[39] S. Santhanagopalan, Q. Guo, P. Ramadass, and R. E. White, "Review of models for predicting the cycling performance of lithium ion batteries," *J. Power Sources*, vol. 156, no. 2, pp. 620–628, Jun. 2006.
[40] M. Doyle, T. F. Fuller, and J. Newman, "Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell," *J. Electrochem. Soc.*, vol. 140, no. 6, pp. 1526–1533, Jun. 1993.
[41] H. E. Perez, X. Hu, and S. J. Moura, "Optimal charging of batteries via a single particle model with electrolyte and thermal dynamics," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2016, pp. 4000–4005.
[42] S. J. Moura, F. B. Argomedo, R. Klein, A. Mirtabatabaei, and M. Krstic, "Battery state estimation for a single particle model with electrolyte dynamics," *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 2, pp. 453–468, Mar. 2017.
[43] H. E. Perez, S. Dey, X. Hu, and S. J. Moura, "Optimal charging of Li-ion batteries via a single particle model with electrolyte and thermal dynamics," *J. Electrochem. Soc.*, vol. 164, no. 7, pp. A1679–A1687, 2017.
[44] M. Ecker, T. K. D. Tran, P. Dechent, S. Käbitz, A. Warnecke, and D. U. Sauer, "Parameterization of a physico-chemical model of a lithium-ion battery: I. Determination of parameters," *J. Electrochemical Soc.*, vol. 162, no. 9, pp. A1836–A1848, 2015.
[45] M. Ecker, S. Käbitz, I. Laresgoiti, and D. U. Sauer, "Parameterization of a physico-chemical model of a lithium-ion battery: II. Model validation," *J. Electrochem. Soc.*, vol. 162, no. 9, pp. A1849–A1857, 2015.
[46] J. M. Maciejowski and M. Huzmezan, "Predictive control," in *Robust Flight Control: A Design Challenge*. Cham, Switzerland: Springer, 2007, pp. 125–134.
[47] G. C. Calafiore and L. E. Ghaoui, "On distributionally robust chance-constrained linear programs," *J. Optim. Theory Appl.*, vol. 130, no. 1, pp. 1–22, Dec. 2006.
[48] L. Patnaik, A. V. J. S. Praneeth, and S. S. Williamson, "A closed-loop constant-temperature constant-voltage charging technique to reduce charge time of lithium-ion batteries," *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 1059–1067, Feb. 2019.
[49] J. A. E. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl, "CasADi: A software framework for nonlinear optimization and optimal control," *Math. Program. Comput.*, vol. 11, no. 1, pp. 1–36, Mar. 2019.

ANDREA POZZI (Member, IEEE) received the bachelor's degree in industrial engineering, the master's degree in electrical engineering, and the Ph.D. degree in electronics, informatics, and electrical engineering from the University of Pavia, in 2015, 2017, and 2021, respectively. He has held visiting scholar positions at TU Braunschweig, in 2016, and UC Berkeley, in 2019. After serving as a Postdoctoral Researcher with the University of Pavia, he joined the Catholic University of the Sacred Heart, Brescia, Italy, as an Assistant Professor of machine learning with the Faculty of Mathematical, Physical and Natural Sciences, in January 2022. His research interests encompass reinforcement learning, imitation learning, machine learning, approximate dynamic programming, and advanced control theory.



ALESSANDRO INCREMONA received the bachelor's and master's degrees in computer engineering and the Ph.D. degree in electronics, informatics, and electrical engineering from the University of Pavia, in 2015, 2017, and 2021, respectively. In 2019, he held a position as a Research Intern with IBM Research Dublin. He is currently a Postdoctoral Researcher with the Catholic University of the Sacred Heart, Brescia, Italy. His research interests include time series forecasting, regularization, statistical process control, and machine learning.

DANIELE TOTI (Member, IEEE) received the bachelor's and master's degrees in computer engineering and the Ph.D. degree in computer science and engineering from Roma Tre University, in 2005, 2008, and 2012, respectively. His career spans both worlds of academia and industry, with appointments in consultancy companies, research centers, universities, and software houses. He also acted as a scientific consultant for a number of international IT companies and research organizations. Since 2020, he has been a member with the Faculty of Mathematical, Physical and Natural Sciences, Catholic University of the Sacred Heart, Brescia, Italy. He is currently an Associate Professor of computer engineering. He participated in several EU-funded international and national projects, including: ARISTOTELE and LASIE (FP7); eJRM, MODERN, SIRET and Wheesbee (PON); and SURVANT, DANTE, VIMMP, MarketPlace, OntoTrans, OntoCommons, OYSTER, OpenModel, and NanoMECommons (H2020). He is an Oracle and Sun Certified Professional with seven industry-level certifications. His research interests include information extraction, natural language processing, and semantic knowledge discovery and modeling.