2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Optimal Dispatch of Hybrid Renewable–Battery Storage Resources: A Stochastic Control Approach

Thiha Aung and Mike Ludkovski
Department of Statistics & Applied Probability, UC Santa Barbara
taung@ucsb.edu, ludkovski@pstat.ucsb.edu

Abstract— **We study the daily operation of hybrid energy resources that couple a renewable generator with a battery energy storage system (BESS). We propose a dynamic stochastic control formulation for optimal dispatch of BESS to maximize the reliability of the hybrid asset relative to a given day-ahead dispatch target or forecast. We develop a machine-learning algorithm based on Gaussian Process regression to efficiently find the dynamic feedback control map. Several numerical case studies highlight the flexibility and extensibility of our methodology, including the ability to consider alternative objectives, such as peak shaving. We also provide a sensitivity analysis with respect to the energy capacity and power rating of the BESS.**

I. Introduction

The stochastic and intermittent nature of wind and solar energy resources creates a mismatch between their projected generation at the time of day-ahead (DA) unit commitment and their actual power production. Renewable over-generation relative to the DA dispatch target can lead to curtailment and revenue loss. Under-generation requires the use of grid reserves and may trigger financial penalties. To remedy this reliability problem, grid operators are encouraging the deployment of hybrid assets [1] that couple a renewable resource with a battery energy storage system (BESS), participating in the daily power market as a single entity. Hybrid assets are being deployed at an exponential rate: according to the 2022 EIA report [2], more than 10GW of hybrids will come online in U.S. in 2023 and 2024. This corresponds to approximately one-third of all new solar energy projects and more than half of new storage proposals [3]. Independent system operators like ERCOT and CAISO have accordingly been creating new rules for day-to-day operation of hybrid resources.

Contributions: In this article, we investigate operation of integrated hybrid resources aiming to optimally firm their output via BESS-provided real-time control. Thus, the battery is dynamically used to counter deviations of realized renewable production relative to a given dispatch target, optimally taking into account the intertemporal constraints of the BESS capacity and power rating limits. We formulate a time-dependent stochastic control problem and propose a direct implementation of the dynamic programming equations via a machine learning approach, namely by building two Gaussian Process emulators for the continuation value and the optimal control maps. Methodologically, our primary contribution is a general-purpose algorithm that tractably determines BESS adaptive dispatch trajectories through minimizing a proxy step-wise loss functional. Our approach directly works with probabilistic generation scenarios and continuous input and action spaces, offering an experimental investigation of BESS dispatch in hybrid assets.

Literature Review: With the explosive growth in BESS deployment, a rapidly expanding literature considers both standalone BESS operations, as well as the coupling of BESS with renewable resources [4], [5]. Thanks to the flexibility of BESS, there is a slew of potential objectives, such as energy arbitrage (shifting energy production in time) [6], [7], [8], [9], ancillary service participation [10], [11], [12], peak shaving [13], [14], and transmission congestion mitigation. Unlike our approach, most extant works (such as [15]) consider only a limited number of pre-determined scenarios to capture the uncertainty in renewable production.

To solve the stochastic control problem, we utilize Regression Monte Carlo (RMC), which is a simulation-based policy/value function iteration scheme. RMC falls within the realm of *Approximate Dynamic Programming* [16] which exploits the classical Bellman's recursion. RMC has been extensively used for valuing natural gas storage [17], [18], [19], as well as for microgrid operations in [20], [21], [22] where the BESS assists a backup diesel generator. To our knowledge, we are the first to apply RMC to the task of firming hybrids' output.

In this article, we rely on Gaussian Processes (GP) for emulation of conditional expectation. An alternative to using GPs are feedforward neural networks, which have been utilized for related problems in [14], [23]. Unlike the above approaches that directly parametrize the policy map, in RMC we first compute pointwise optimal controls and then learn the interpolating statistical surrogate. Neural networks require large datasets and numerous hyperparameters to tune, whereas GPs work efficiently with small datasets, with the kernel being the sole hyperparameter to consider.

The rest of the paper is organized as follows. In Section II we present the system dynamics, firming objective, and the dynamic programming approach. Section III describes the solution scheme based on GP surrogates and pointwise optimization. Section IV presents 3 numerical case studies, as well as a partial sensitivity analysis; Section V concludes.

Work partially supported by ARPA-E PERFORM grant DE-AR0001289.

II. Problem Formulation

A. Renewable Generation and BESS Dynamics

We consider a control horizon T (typically 24 hours) with a decision time grid $\mathcal{T} = \{0 = t_0, t_1, \dots, t_K = T\}$ where $t_k = k\Delta t$, typically 15 minutes. To fix ideas, we consider a hybrid asset composed of a wind farm and BESS. Following [24], [25] who used mean-reverting Stochastic Differential Equations (SDEs) for probabilistic wind power forecasting, we model wind power generation in MW, $(X_k)_{k\in\mathcal{T}}$, using the discrete-time equivalent of the time-dependent Ornstein-Uhlenbeck SDE:

$$X_{k+1} = X_k + \alpha(m_k - X_k)\Delta t + \sigma \cdot Z_k, \tag{1}$$

where m_k is the mean reversion level (in MW), σ is the volatility (in MW$\Delta t^{-0.5}$), and $\alpha \geq 0$ is the unit-less mean reversion coefficient. Finally, $Z_k \sim \mathcal{N}(0, \Delta t)$ are the exogenous, i.i.d. stochastic shocks driving $(X_k)_{k\in\mathcal{T}}$.

Denote by I_k the state of charge (SoC) of BESS in MWh and B_k the controlled charge/discharge rate in MW at time t_k. Then I_k evolves according to

$$I_{k+1} = I_k + \big(\eta B_k \mathbb{1}_{\{B_k\geq 0\}} + \frac{1}{\eta} B_k \mathbb{1}_{\{B_k<0\}}\big)\Delta t \tag{2}$$

where $\eta \leq 1$ represents the charging efficiency. The BESS SoC bounds are given by the constraints:

$$\mathrm{SoC}_{\min} \cdot I_{cap} \leq I_k \leq \mathrm{SoC}_{\max} \cdot I_{cap} \tag{3}$$

where I_{cap} is the rated capacity in MWh and $\mathrm{SoC}_{\min}, \mathrm{SoC}_{\max}$ are SoC percentage limits.

The power ratings of the BESS are given by $B_{\max} > 0$ and $B_{\min} < 0$. Furthermore, B_k must satisfy the SoC-dependent capacity constraints of (3):

$$\begin{gathered} \eta \cdot \frac{SoC_{\min} I_{cap} - I_k}{\Delta t} \leq B_k \leq \frac{SoC_{\max} I_{cap} - I_k}{\eta \Delta t} \\ B_{\min} \leq B_k \leq B_{\max}. \end{gathered} \tag{4}$$

Given the Markovian structure of (X_k) in (1), we focus on closed-loop feedback-form strategies so that B_k is a function of X_k, I_k. Note that due to the form of (2), this leads to all three processes: the generation $(X_k)_{k\in\mathcal{T}}$, the SoC process $(I_k)_{k\in\mathcal{T}}$, and the controlled $B(k, X_k, I_k)$ being adapted to the information filtration $(\mathcal{F}_k)_{k\geqslant 0}$ generated by the external shocks $\{Z_k\}_{k\geq 0}$. We denote by $\mathcal{A}$ the set of admissible feedback controls $B := (B_0, B_1, \dots, B_k, \dots, B_{K-1})$ such that B_k is adapted to filtration $\mathcal{F}_k$ and satisfies constraints in (4).

B. Firming objective

The hybrid asset firms its renewable generation by aiming for output $O_k := X_k - B_k$ to be close to M_k, representing the dispatch target at time t_k. Thus, the controller decides whether to charge/discharge BESS and how much as a function of SoC and renewable generation, see Figure 1.

We formulate the above objective as minimizing the expected cost functional given by the stochastic control value function:

$$V(0, X_0, I_0) := \inf_{B\in\mathcal{A}} \mathbb{E}\left[\sum_{k=0}^{K-1} f(X_k, M_k, B_k) + g(I_T)\right] \tag{5}$$

with the state dynamics in (1) and (2).

The key idea of (5) is to convert a global objective, such as minimizing the L_∞ norm of the vector $\|O_k - M_k\|_\infty$, into a stepwise criterion that can be recursively optimized via dynamic programming. Hence, the running cost $f(x, m, b)$ is an auxiliary loss function acting as proxy for a desired global criterion. The terminal cost g accounts for constraints on the BESS's SoC at T so as to avoid complete depletion of the battery at the control horizon.

Below we consider the following stepwise losses:

- $f_1(X_k, M_k, B_k) := |X_k - B_k - M_k| \equiv |O_k - M_k|$ corresponding to the L^1 loss;
- $f_2(X_k, M_k, B_k) := (X_k - B_k - M_k)^2 \equiv (O_K - M_k)^2$ corresponding to the L^2 loss.

In Section IV-C we introduce a third choice f_{sh}.

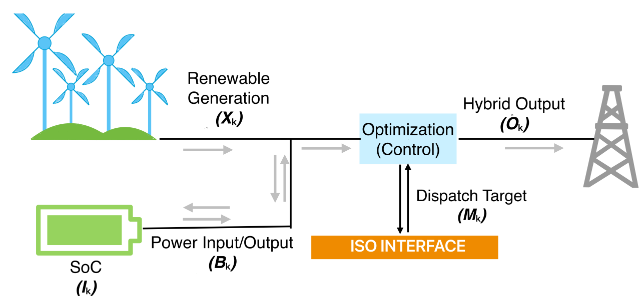

Fig. 1: A schematic description of optimizing the output of a hybrid renewable–BESS resource.

C. Dynamic Programming

To solve the stochastic control problem (5), we rely on the *Dynamic Programming Principle (DPP)* [26] which decouples our multi-stage control problem into intermediate sub-problems via Bellman's recursion:

$$\begin{aligned} V(t_k, X_k, I_k) = \inf_{B_k\in\mathcal{A}_k} \Big\{ & f(X_k, M_k, B_k) \\ & + \mathbb{E}[V(t_{k+1}, X_{k+1}, I_{k+1})|X_k, I_k] \Big\} \end{aligned} \tag{6}$$

where $\mathcal{A}_k$ is the feasible set for B_k and the expectation is taken over the random variable X_{k+1}, conditioned on current generation X_k. In line with the DPP approach, we numerically optimize the control at time t_k by

$$\arg\inf_{B_k\in\mathcal{A}_k} f(X_k, M_k, B_k) + Q(t_k, X_k, I_{k+1}), \tag{7}$$

where $Q(t_k, X_k, I_{k+1}) := \mathbb{E}[V(t_{k+1}, X_{k+1}, I_{k+1}(B_k)|X_k]$ is the continuation q-value. To solve for the optimal control at time t_k according to (7), we need to approximate $Q(t_k, \cdot)$. To do so, we proceed within the realm of RMC algorithms.

III. Implementation

A. RMC Algorithm

We follow the Dynamic Emulation Algorithm (DEA) framework [18]. As in standard DPP, we proceed backward in time, starting with the known terminal condition $V(T, X_T, I_T) = g(I_T)$. Then for $k = K-1, \ldots, 0$, we repeat the following 3 sub-steps:

1) Evaluate the pointwise optimal control B^* in (7) for every input in the simulation design;
2) Evaluate the pointwise value function in (6) for every input using the above optimal control B^*;
3) Construct the continuation value emulator for

$$\widehat{Q}_k : (X, I) \mapsto \mathbb{E}\big[\widehat{V}(t_{k+1}, X_{k+1}, I) \mid X_k = X\big]. \tag{8}$$

In (8), the q-value $\widehat{Q}_k$ is a function of the *current* wind power generation X_k at step t_k and the *lookahead* SoC I_{k+1} at step t_{k+1}. Accordingly, the simulation design is $\mathcal{D}_k = (X_k^n, I_{k+1}^n)_{n=1}^N$.

The first two sub-steps of the DEA loop entail obtaining the training output $v_{k+1}^{1:N}$ for each input (X_k^n, I_{k+1}^n). To do so, we first sample a one-step forward simulation $X_k^n \to X_{k+1}^n$. We then perform the numerical optimization (7) to obtain the corresponding lookahead optimal control $B_{k+1}^{*,n}$. Finally, we compute the resulting pathwise value

$$\begin{aligned} v_{k+1}^n = f(X_{k+1}^n, M_{k+1}, B_{k+1}^{*,n}) + \\ \widehat{Q}_{k+1}(X_{k+1}^n, I_{k+2}^n), \qquad n = 1, \ldots, N. \end{aligned} \tag{9}$$

For sub-step 3), we learn the mapping $\widehat{Q}_k(\cdot)$ by regressing $v_{k+1}^{1:N}$ against the design $\mathcal{D}_k$, i.e. an empirical L^2-projection into a given function space $\mathcal{H}^q$,

$$\widehat{Q}_k := \underset{q_k \in \mathcal{H}^q}{\arg\inf} \sum_{n=1}^N (q_k(X_k^n, I_{k+1}^n) - v_{k+1}^n)^2. \tag{10}$$

B. Simulation Design

We provide details on the simulation design $\mathcal{D}_k$ introduced in the prior section. We use a space filling design over $[X_{\min}^k, X_{\max}^k] \times [0, I_{cap}]$, where the choice of $X_{\min}^k$ and $X_{\max}^k$ is based on the range of X_k. We opt for Latin Hypercube Sampling (LHS), which is a variance-reduced version of random uniform sampling in each coordinate.

We also apply **replication**, dividing our training design into N_{loc} distinct sites, with each distinct input repeated N_{rep} times (for the remainder of the subsection, $\mathbf{x} \equiv (X, I)$ is a generic training input):

$$\mathcal{D}_k = \{\underbrace{\mathbf{x}^1, \ldots, \mathbf{x}^1}_{N_{\text{rep}} \text{ times}}, \underbrace{\mathbf{x}^2, \ldots, \mathbf{x}^2}_{N_{\text{rep}} \text{ times}}, \ldots, \mathbf{x}^{N_{\text{loc}}}\}. \tag{11}$$

The total simulation budget at each step t_k is $N = N_{loc} \times N_{rep}$. Subsequently, one-step forward simulations and optimizations are performed to acquire the respective responses $y^{1,1}, y^{1,2}, \ldots, y^{i,j}, \ldots, y^{N_{\text{loc}}, N_{\text{rep}}}$. After pre-averaging the replicates $\bar{y}^i := \frac{1}{N_{\text{rep}}} \sum_{j=1}^{N_{\text{rep}}} y^{i,j}$, the regression model for the continuation value emulator $\widehat{Q}_k(\cdot)$ is applied to the reduced design $\bar{\mathcal{D}}_k := (\mathbf{x}^{1:N_{loc}}, \bar{y}^{1:N_{loc}})$. The replicated design lowers training errors thanks to the decreased variability in $\bar{y}^i$'s, raising the signal-to-noise ratio.

C. Gaussian Process Emulator for Q

In order to enhance the numerical optimization of B_k in (7), we opt for an emulator that has an analytical gradient. To achieve this, we make use of Gaussian Process regression (GPR). GPR models $Q(\cdot)$ as a Gaussian Process (GP), specified by a mean function $m(\mathbf{x})$ (taken to be zero after standardizing the outputs) and positive definite covariance function $c(\mathbf{x}, \mathbf{x}')$ [27]. The covariance $c(\cdot, \cdot)$ specifies the smoothness of $\widehat{Q}$. Given a training design $(\mathbf{x}^{1:N}, y^{1:N})$ and an input $\mathbf{x}^*$, the continuation value, $\widehat{Q}(\mathbf{x}^*)$, is the posterior mean of the GP given by

$$\widehat{Q}(\mathbf{x}_*) = \mathcal{C}_*^\top (\mathbf{C} + \sigma_\epsilon^2 \mathbf{I})^{-1} \mathbf{y} \tag{12}$$

where $\mathbf{I}$ is $N \times N$ identity matrix, $\mathbf{y} = \left[y^1, \ldots, y^N\right]^\top$,

$$\mathcal{C}_*^\top = \left[c\left(\mathbf{x}_*, \mathbf{x}^1; \vartheta\right), \ldots, c\left(\mathbf{x}_*, \mathbf{x}^N; \vartheta\right)\right]. \tag{13}$$

Finally, σ_ϵ^2 represents observation noise and $\mathbf{C}$ is the $N \times N$ covariance matrix with $\mathbf{C}_{k,l} = c(\mathbf{x}_k, \mathbf{x}_l; \vartheta)$ where ϑ is the parameter vector for the covariance function.

We choose the anisotropic Matérn-5/2 kernel

$$\begin{aligned} c_{M52}(\mathbf{x}, \mathbf{x}'; \vartheta) := \sigma_p^2 \prod_{j=1}^2 \Big(1 + \frac{\sqrt{5}}{\ell_j}|x_j - x_j'| + \\ \frac{5}{3\ell_j^2}(x_j - x_j')^2\Big) \cdot \exp\big(-\frac{\sqrt{5}}{\ell_j}|x_j - x_j'|\big), \end{aligned} \tag{14}$$

where $\vartheta = (\sigma_p^2, \ell_1, \ell_2)$: σ_p^2 indicates the magnitude of the response, and ℓ_1 and ℓ_2 determine how the response fluctuates with respect to wind power generation (MW) and SoC (MWh), which are expressed in different scales and units and hence have different lengthscales. The parameters ϑ and σ_ϵ^2 are optimized using the maximum likelihood estimation (MLE). To accelerate the optimization of $\widehat{Q}_k(\cdot)$, we warm-start with the hyperparameters $\vartheta^{(k+1)}$ obtained from the trained GP $\widehat{Q}_{k+1}(\cdot)$ at time t_{k+1}.

Remark 1: Using replicates not only dramatically speeds up GPR training which is cubic in N_{loc} but moreover offers more stable MLE results thanks to lower observation noise.

Estimating the value function: After training, the RMC algorithm produces continuation value emulators for each time step $\{\widehat{Q}_k(\cdot)\}_{k=0}^{K-1}$. To evaluate the resulting hybrid resource output trajectory (O_k) and the respective value function, we utilize Monte Carlo simulation. Given an initial state (X_0, I_0), we generate M out-of-sample paths $(X_{0:K}^m, I_{0:K}^{*,m})$, $m = 1, \ldots, M$, where the optimized SoC I^* is based on $B^*(X_k^m, I_k^{*,m})$ according to (7). This gives cumulative pathwise realized cost

$$v_{0:K}^m = \sum_{k=0}^{K-1} f(X_k^m, M_k, B_k^{*,m}) + g(I_T^*), \tag{15}$$

and the resulting Monte Carlo estimate of the value function:

$$\widehat{V}(0, X_0, I_0) = \frac{1}{M} \sum_{m=1}^{M} v_{0:K}^m. \tag{16}$$

D. Policy Map

The constrained optimization problem in (7) is given implicitly in terms of the emulator $\widehat{Q}_k(\cdot)$, with the respective first-order condition tied to $\partial \widehat{Q}_k / \partial I$. GPR allows the use of faster gradient-based optimizers thanks to its analytical gradients. Differentiating a GP $\widehat{Q}(\cdot)$ in x_j gives another GP with posterior mean at input $\mathbf{x}_*$ given by

$$\frac{\partial \widehat{Q}}{\partial x_j}(\mathbf{x}_*) = \sum_{n=1}^{N} \alpha_n \frac{\partial c_{M52}}{\partial x_j}(\mathbf{x}_*, \mathbf{x}^n; \vartheta) \tag{17}$$

where α_n is the n-th component of $\left(\mathbf{C} + \sigma_\epsilon^2 \mathbf{I}\right)^{-1} \mathbf{y}$; see [28] for the gradient of the Matérn-5/2 kernel. In our case we differentiate with respect to SoC I, $j = 2$.

The presence of the black-box $\widehat{Q}$ makes the optimization problem non-convex. We employ the unconstrained, gradient-based **L-BFGS** solver from the **SciPy** library and then directly enforce the constraints that define the feasible set $\mathcal{A}_k$ to obtain the optimal pointwise control B^* in (7).

Emulating the feedback control: Both the training stage and the out-of-sample evaluation stage in (15) require repeated calls of **L-BFGS** subroutine tens of thousands of times to determine $B^*(X_k^n, I_k^n)$. To accelerate this aspect, we build an emulator for the map $\widehat{B}_k : (X, I) \mapsto \mathcal{A}_k \subset \mathcal{R}$ for each time step t_k. To do so, we train $\widehat{B}_k$ by regressing $(B_k^{*,n})_{n=1}^{N_b}$ against $(X_k^n, I_k^n)_{n=1}^{N_b}$ using a subset of size $N_b \ll N$ (in the examples below we take $N_b = N_{loc}$) and an approximation space $\mathcal{H}^b$,

$$\widehat{B}_k(\cdot) = \arg\inf_{h_k \in \mathcal{H}^b} \sum_{n=1}^{N_b} \left(h_k\left(X_k^n, I_k^n\right) - B_k^{*,n}\right)^2. \tag{18}$$

To fit $\widehat{B}_k$ we rely on GPR with a Matérn-3/2 kernel. Once trained, we use the control emulators $\{\widehat{B}_k(\cdot)\}_{k=0}^{K-1}$ to evaluate the pathwise v_{k+1}^n's and the final value function in (16), avoiding the need for further numerical optimization. The full procedure is detailed in Algorithm 1.

The stability of Algorithm 1 depends on how well the GPR emulates the optimal control. Figure 2 compares the control map surface from the GP $\widehat{B}_k$ and the direct **L-BFGS** optimizer B^* under the L^2 criterion $f_2(\cdot)$. We observe that the latter can be quite non-smooth across (X, I), witness the "cracks" in the bottom-left and top-right (presumably due to some high-order instability in $\widehat{Q}_k$ that leads to local optima). The GPR helps to smooth out and remove such numerical artifacts.

IV. Numerical Experiments

A. Toy stationary example

As our first demonstration, we consider a time-stationary generation profile with a matching constant dispatch target, $m_k = M_k \equiv 5$ MW for $k = 0, \ldots, 95$ and terminal condition on (I_k) at $k = 96$, representing 24 hours at 15 min frequency. The other parameters are in Table I. The BESS starts with 10% SoC and has the matching terminal condition $g(I_T) = \lambda \cdot (0.1 I_{cap} - I_T)_+$.

Algorithm 1 RMC for dispatching hybrid renewable-BESS resources

Input: K steps, N_{loc} sites, N_{rep} replications
Set $\widehat{Q}_{K-1}(X_{K-1}, I_K) = g(I_T)$ (No emulation)
Generate Design $\mathcal{D}_{K-1} = (X_{K-2}^{i,j}, I_{K-1}^{i,j})$ for $i = 1, 2, \ldots, N_{loc}$ and $j = 1, 2, \ldots, N_{rep}$
Generate one-step paths: $X_{K-2}^{i,j} \to X_{K-1}^{i,j}$ for $i = 1, 2, \ldots, N_{loc}$ and $j = 1, 2, \ldots, N_{rep}$
Optimize $B_{K-1}^{*,i,1}$ in (7) for $(X_{K-1}^{i,1}, I_{K-1}^{j,1})$ respectively, $i = 1, 2, \ldots, N_{loc}$.
Fit control GP $\widehat{B}_{K-1}(\cdot)$ by regressing $B_{K-1}^{*,i,1}$ against $(X_{K-1}^{i,1}, I_{K-1}^{j,1})$.
for k = $K-1$ to 1 **do**
  Evaluate $B_k^{*,i,j}$ using control GP $\widehat{B}_k(\cdot)$ for $i = 1, 2, \ldots, N_{loc}$ and $j = 1, 2, \ldots, N_{rep}$
  Evaluate $v_k^{i,j}$ in (9) for $i = 1, 2, \ldots, N_{loc}$ and $j = 1, 2, \ldots, N_{rep}$
  Average over replicates: $\bar{v}_k^i = \frac{1}{N_{rep}} \sum_{l=1}^{N_{rep}} v_k^{i,l}$ for $i = 1, 2, \ldots, N_{loc}$
  Fit continuation value GP $\widehat{Q}_{k-1}(\cdot)$ by regressing $\bar{v}_k^i$ against $(X_{k-1}^{i,1}, I_k^{i,1})$ for $i = 1, 2, \ldots, N_{loc}$.
  Generate design $\mathcal{D}_{k-1} = (X_{k-2}^{i,j}, I_{k-1}^{i,j})$ for $i = 1, 2, \ldots, N_{loc}$ and $j = 1, 2, \ldots, N_{rep}$
  Generate one-step paths: $X_{k-2}^{i,j} \to X_{k-1}^{i,j}$ for $i = 1, 2, \ldots, N_{loc}$ and $j = 1, 2, \ldots, N_{rep}$
  Optimize $B_{k-1}^{*,i,1}$ in (7) for each sample $(X_{k-1}^{i,1}, I_{k-1}^{i,1})$ for $i = 1, 2, \ldots, N_{loc}$.
  Fit control GP $\widehat{B}_{k-1}(\cdot)$ by regressing $B_{k-1}^{*,i,1}$ against $(X_{k-1}^{i,1}, I_{k-1}^{i,1})$ for $i = 1, 2, \ldots, N_{loc}$.
end for
Output: $\{\widehat{B}_k(\cdot), \widehat{Q}_k(\cdot)\}_{k=0}^{K-1}$.

To train the GP emulators $\widehat{Q}$ we use a training design of size $N = N_{loc} \times N_{rep} = 3 \cdot 10^4$ at each time step t_k, with $N_{loc} = 600$ unique sites and batch size $N_{rep} = 50$. The implementation via the Python `scikit` library is run on a laptop; training all the GPs takes under 18 minutes.

Figure 3 visualizes the resulting policy $\widehat{B}_k$ based on the quadratic f_2 criterion. Observe that when SoC is far from being empty or full, $\widehat{B}_k(X, I) \simeq X - M_k$ is almost linear in the middle of the policy surface. However, when both the SoC I and renewable generation are high, the optimal controller decreases the charging rate $\widehat{B}_k(X, I) < X - M_k$ to maintain some SoC headroom. Similarly, the controller throttles discharging when the SoC and renewable generation are low. As a result, I_k^* tends to stay in a "safe zone" and away from the SoC limits, demonstrating the precautionary risk-mitigating behavior emerging from DPP.

In contrast, the L^1 criterion f_1 leads to undesirable *greedy* behavior: the BESS charges and discharges to the fullest

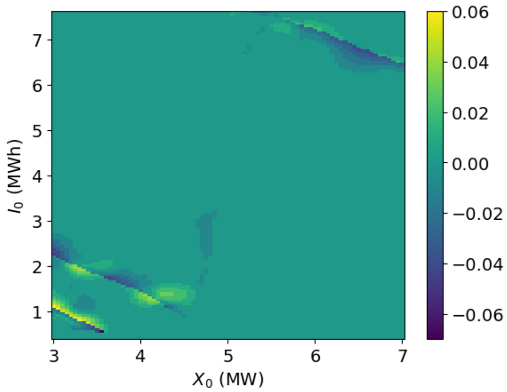

Fig. 2: Difference between the fitted GP control map $\widehat{B}_k$ and original **L-BFGS** optimizers $B_k^{*,n}$ at $k = 0$.

Parameter	Section $IV - A$	Section $IV - B$
α	1	5
σ	1	1
I_{cap}	8 MWh	30 MWh
I_0	$0.1 I_{cap}$	$0.1 I_{cap}$
$B_{\max}$	2 MW	10 MW
$B_{\min}$	-2 MW	-10 MW
λ	\$50/MWh	\$50/MWh
η	0.90	0.90
$\mathrm{SoC}_{\min}$	5%	5%
$\mathrm{SoC}_{\max}$	95%	95%
Δt	0.25 h	0.25 h
T	24 h	24 h

TABLE I: Parameters of the case studies in Section IV.

extent possible to firm the current renewable generation X, $\widehat{B}_k(X, I) = X - M_k$ up to the physical BESS constraints, ignoring SoC. This is caused by the bang-bang feature of the f_1 criterion vis-a-vis the convex f_2 criterion.

Figure 4 displays a 24-hour trajectory of wind power generation (X_k) in the top, the corresponding SoC (I_k^*) in the middle and the resulting hybrid output (O_k) in the bottom panel. Comparing the L^1 and L^2 controllers, (O_k) is much smoother when working with f_2. Of note, the greedy behavior of the f_1 criterion leads to empty SoC $I_k^* = 0$ for several hours which greatly increases the overall target mismatch.

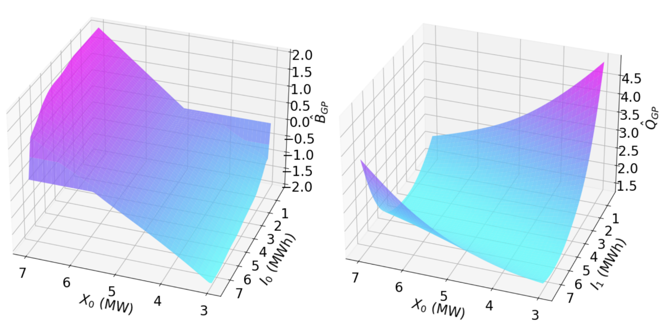

Fig. 3: *Left:* Control emulator $\widehat{B}_k$ as a function of generation X and SoC I. *Right:* Continuation value emulator $(X, I) \mapsto \widehat{Q}_k(X, I)$. Both plots are for the f_2 criterion at initial step $t_k = 0$.

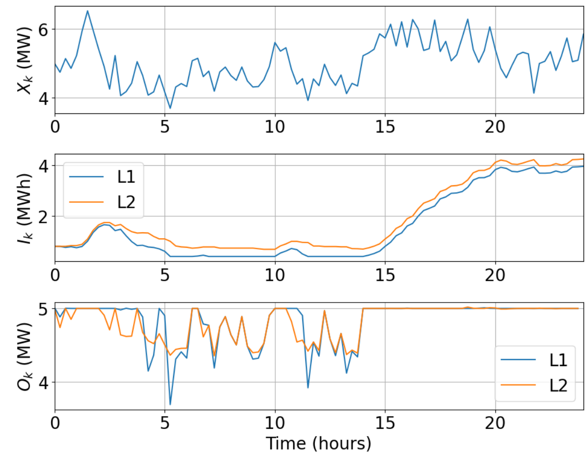

Fig. 4: *Top panel:* A simulation of (X_k) following (1) with constant mean $\mathbb{E}[X_k] = 5$. *Middle*: Corresponding SoC trajectories (I_k^*) following the L^1 and L^2 controls. *Bottom*: Firmed hybrid outputs (O_k).

B. Non-stationary renewable generation

Our second case-study is calibrated to realistic, time-dependent wind generation. Specifically, we use NREL-provided forecasts for renewable resources in ERCOT [29], illustrating with the respective re-analyzed DA forecasts for the Amazon Wind Farm in 2018. This forecast data is hourly and we utilize cubic splines to upscale to intervals of Δt = 15 minutes yielding the smoothed mean-reversion profile m_k in (1). Unlike our first example, it is possible for renewable generation to fall to 0. To make sure that (X_k) stays non-negative, we consider the discrete counterpart of the time-dependent square-root process:

$$X_{k+1} = |X_k + \alpha(m_k - X_k)\Delta t + \sigma\sqrt{X_k} \cdot Z_k|. \quad (19)$$

Due to time-dependent (m_k) and state-dependent volatility, the variance of (X_k) now depends on k. Accordingly we construct adaptive simulation designs $\mathcal{D}_k$ for each k, covering the interval $[\mathbb{E}[X_k] - 2StDev[X_k], \mathbb{E}[X_k] + 2StDev[X_k]]$ (obtained through pilot simulations of (X_k)). We use the same simulation budget as our first example; the other parameters are in Table I.

Figure 5 visualizes the resulting hybrid BESS behavior for several representative days. The left panels consider dispatch targets $M_k = m_k = \mathbb{E}[X_k]$, while the right panels use the original hourly forecasts with a piecewise linear interpolation. The latter context highlights the common possibility that M_k does not match the average X_k. We observe that the BESS provides a significant firming, with the variance of (O_k) an order of magnitude smaller than that of the uncontrolled (X_k). We also observe that the variance of (O_k) is time-dependent: it is higher in the morning due to starting with a near-empty SoC $I_0 = 0.1 I_{cap}$ which limits discharging ability in the morning. $StDev(O_k)$ is also higher when $m_k = \mathbb{E}[X_k]$ rapidly changes which increases

$StDev(X_k)$, and when M_k is far from m_k, see e.g. the September panel.

Figure 5 also reports the improvement of the optimized L^2 policy against the greedy L^1 policy. We report gains from 30 to 90 percent.

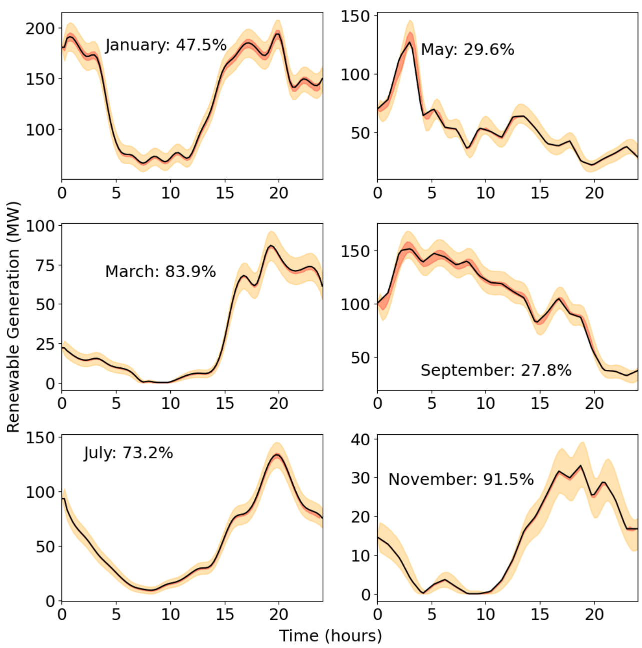

Fig. 5: Firming for several representative days across the listed months of the year. Outer yellow band is 95% CI of renewable generation (X_k). Inner red band is 95% CI of optimized hybrid output (O_k) and the black curve are the dispatch target (M_k). *Left panels:* $M_k = \mathbb{E}[X_k]$. *Right panels:* piecewise linear M_k based on original DA forecast. The printed values denote percent improvement against greedy policy.

$B_{\max}$	I_{cap}	Opt L^2 Loss	Greedy Loss	% Impr
Impact of battery power rating $B_{\max}$				
2.5	30	139.99	183.39	23.7%
5	30	58.18	99.39	41.5%
10	30	21.36	61.75	65.4%
15	30	18.24	58.57	68.9%
Impact of battery capacity I_{cap}				
10	20	22.92	53.39	51.6%
10	30	21.36	61.75	65.4%
10	40	20.48	72.23	71.6%
10	50	19.98	85.15	76.5%

TABLE II: Sensitivity of value functions for the case study in Section IV-B to the BESS parameters $I_{cap}, B_{\max}$ for a representative July day in ERCOT.

Sensitivity Analysis: Table II shows the impact of battery power rating $B_{\max}$ and capacity I_{cap} on the firming objective. In general, larger $B_{\max}$ gives the BESS more headroom to counteract deviations of (X_k) from the dispatch target, while larger I_{cap} allows to sustain longer charging/discharging. When $B_{\max}$ is small relative to the fluctuations in (X_k), the constraint in (4) is frequently binding and is the primary driver of accrued firming losses. As a result, in Table II, optimal loss is roughly inversely proportional to $B_{\max}$ (e.g. doubling $B_{\max}$ halves $V(X_0, I_0)$). However, once $B_{\max} \gg \max_k StDev(X_k)$, the power rating is big enough to handle the vast majority of deviations from the forecast and $B_{\max}$ makes little impact, compare $B_{\max} = 10$ MW vs $B_{\max} = 15$ MW. Comparing to the greedy strategy, if the power rating binds often, there is relatively less gain from dynamic control, while improvements of up to 70% are possible when the primary constraint is the BESS capacity.

To quantify the impact of capacity I_{cap} we consider BESS with durations of 2-hours ($I_{cap} = 2B_{\max}$), 3-, 4- and 5-hours. We observe diminishing returns from adding capacity, with little improvement beyond a 3-hour duration. Note that the greedy strategy performs worse for larger I_{cap} due to incurring larger terminal penalties.

C. Congestion Mitigation

Another important motivation for hybrid resources is mitigating curtailment which leads to revenue loss. Curtailment may arise during periods of over-generation when there are limited regulation-down reserves. Alternatively, curtailment may be triggered by congestion in the transmission grid. As an initial illustration of modeling such objectives, we introduce the loss function f_{sh}, which involves both upper and lower bound generation thresholds. Denote the upper threshold as $\overline{M}_k$ and the lower threshold as $\underline{M}_k$, and set

$$f_{sh}(X_k, M_k, B_k) := (O_k - M_k)^2 + P_1 \cdot (O_k - \overline{M}_k)_+^2 + P_2 \cdot (\underline{M}_k - O_k)_+^2. \quad (20)$$

The penalties P_1, P_2 control the relative importance of the thresholds and can be used to create an asymmetric criterion.

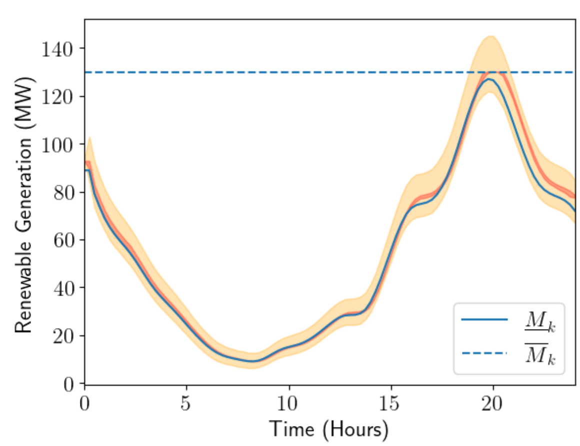

Fig. 6: Firming on a July day with dispatch thresholds $\underline{M}_k = 0.95 M_k$ and $\overline{M}_k = 130$ MW. Outer yellow band is 95% CI of renewable generation (X_k). Inner red band is 95% CI of hybrid output (O_k) after optimal dynamic firming.

In Figure 6 we consider a case study with a fixed upper threshold $\overline{M}_k = 130$ MW (with penalty $P_1 = 100$ per MW) that represents a limited transmission line capacity, and a variable lower threshold $\underline{M}_k = 0.95M_k$ (with penalty $P_2 = 50$) i.e. avoiding under-generation of more than 5% below the DA dispatch target. The underlying dispatch profile matches the July panel of Figure 5 and uses the same parameters as in Section IV-B with power rating $B_{max} = 15$. The resulting output (O_k) can be seen to stay completely below the curtailment threshold $\overline{M}_k$. Also we observe that the variance of O_k is minimal around noon due to the binding lower threshold $\underline{M}_k$, while it is relatively larger after 8pm when neither threshold binds.

V. Conclusion and Future Work

We developed an algorithm for dynamic real-time dispatch of hybrid resources firming a given DA target profile. As demonstrated, our algorithm can handle multiple performance criteria, including L_2-penalization, peak shaving, and two-sided dispatch thresholds. Furthermore, it is agnostic to the underlying dynamics, and can be adjusted to work with nonlinear SDEs, non-Gaussian innovations, or non-parametric scenario simulators.

Among avenues for further work, we highlight the task of incorporating energy prices, both for the purpose of augmenting firming with energy arbitrage, as well as to have a stochastic cost of deviating from the target. This would require to add a third state variable (P_k) and to capture the joint behavior of (P_k, X_k). A related task is to minimize curtailment viewed as a stochastic process that can be zero (no curtailment) or in the range $[0, X_k]$. By adding additional state variables, our method could be extended to analyze hybrid resource participation in ancillary service provision. From a control standpoint, a significant extension would be analysis of several distributed energy resources equipped with storage, necessitating coordination of multiple BESS controls.

References

[1] M. Ahlstrom, J. Mays, E. Gimon, A. Gelston, C. Murphy, P. Denholm, and G. Nemet, "Hybrid resources: Challenges, implications, opportunities, and innovation," *IEEE Power and Energy Magazine*, vol. 19, no. 6, pp. 37–44, 2021.

[2] US Energy Information Administration, "Today in energy: U.S. battery storage capacity expected to nearly double in 2024," 1 2024.

[3] Hybrid Resources Task Force, "Unlocking the flexibility of hybrid resources," tech. rep., Reston, VA: Energy Systems Integration Group, 2022.

[4] H. Ding, Z. Hu, and Y. Song, "Stochastic optimization of the daily operation of wind farm and pumped-hydro-storage plant," *Renewable Energy*, vol. 48, pp. 571–578, 2012.

[5] G. Karmiris and T. Tengnér, "Control method evaluation for battery energy storage system utilized in renewable smoothing," in *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, pp. 1566–1570, 2013.

[6] L. Weber, A. Bušić, and J. Zhu, "Reinforcement learning based demand charge minimization using energy storage," in *62nd IEEE Conference on Decision and Control (CDC)*, pp. 4351–4357, 2023.

[7] G. He, Q. Chen, C. Kang, P. Pinson, and Q. Xia, "Optimal bidding strategy of battery storage in power markets considering performance-based regulation and battery cycle life," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2359–2367, 2016.

[8] D. Krishnamurthy, C. Uckun, Z. Zhou, P. R. Thimmapuram, and A. Botterud, "Energy storage arbitrage under day-ahead and real-time price uncertainty," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 84–93, 2018.

[9] N. Zheng, X. Liu, B. Xu, and Y. Shi, "Energy storage price arbitrage via opportunity value function prediction," in *2023 IEEE Power Energy Society General Meeting (PESGM)*, pp. 1–5, 2023.

[10] G. He, Q. Chen, C. Kang, Q. Xia, and K. Poolla, "Cooperation of wind power and battery storage to provide frequency regulation in power markets," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3559–3568, 2017.

[11] S. Hanif, M. J. E. Alam, K. Roshan, B. A. Bhatti, and J. C. Bedoya, "Multi-service battery energy storage system optimization and control," *Applied Energy*, vol. 311, p. 118614, 2022.

[12] B. Cheng and W. B. Powell, "Co-optimizing battery storage for the frequency regulation and energy arbitrage using multi-scale dynamic programming," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1997–2005, 2016.

[13] 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, 2018.

[14] J. Li, C. Wang, and H. Wang, "Optimal energy storage scheduling for wind curtailment reduction and energy arbitrage: A deep reinforcement learning approach," *arxiv: 2304.02239*, 2023.

[15] A. Belloni, L. Piroddi, and M. Prandini, "A stochastic optimal control solution to the energy management of a microgrid with storage and renewables," in *2016 American Control Conference (ACC)*, pp. 2340–2345, 2016.

[16] W. B. Powell, *Approximate Dynamic Programming: Solving the curses of dimensionality*, vol. 703. John Wiley & Sons, 2007.

[17] R. Carmona and M. Ludkovski, "Valuation of energy storage: an optimal switching approach," *Quantitative Finance*, vol. 10, no. 4, pp. 359–374, 2010.

[18] M. Ludkovski and A. Maheshwari, "Simulation methods for stochastic storage problems: A statistical learning perspective," *Energy Systems*, vol. 11, pp. 377–415, 05 2020.

[19] S. Nadarajah, F. Margot, and N. Secomandi, "Comparison of least squares Monte Carlo methods with applications to energy real options," *European Journal of Operational Research*, vol. 256, no. 1, pp. 196–204, 2017.

[20] C. Alasseur, A. Balata, S. Aziza, A. Maheshwari, P. Tankov, and X. Warin, "Regression Monte Carlo for microgrid management," *ESAIM: Proceedings and Surveys*, vol. 65, 02 2018.

[21] A. Balata and J. Palczewski, "Regress-later Monte Carlo for optimal inventory control with applications in energy," *arXiv: 1703.06461*, 2018.

[22] A. Balata, M. Ludkovski, A. Maheshwari, and J. Palczewski, "Statistical learning for probability-constrained stochastic optimal control," *European Journal of Operational Research*, vol. 290, no. 2, pp. 640–656, 2021.

[23] A. Bachouch, C. Huré, N. Langrené, and H. Pham, "Deep neural networks algorithms for stochastic control problems on finite horizon: Numerical applications," *Methodology and Computing in Applied Probability*, vol. 24, p. 143–178, Jan. 2021.

[24] J. K. Møller, M. Zugno, and H. Madsen, "Probabilistic forecasts of wind power generation by stochastic differential equation models," *Journal of Forecasting*, vol. 35, no. 3, pp. 189–205, 2016.

[25] R. Caballero, A. Kebaier, M. Scavino, and R. Tempone, "Quantifying uncertainty with a derivative tracking SDE model and application to wind power forecast data," *Statistics and Computing*, vol. 31, no. 5, p. 64, 2021.

[26] R. Bellman and S. Dreyfus, *Dynamic Programming*, vol. 33. Princeton University Press, 2010.

[27] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. MIT Press, 2006.

[28] M. Ludkovski and Y. Saporito, "Krighedge: Gaussian process surrogates for delta hedging," *Applied Mathematical Finance*, vol. 28, no. 4, pp. 330–360, 2021.

[29] T. Xu, A. B. Birchfield, K. S. Shetye, and T. J. Overbye, "Creation of synthetic electric grid models for transient stability studies," in *The 10th Bulk Power Systems Dynamics and Control Symposium (IREP 2017)*, pp. 1–6, 2017.