Global-Decision-Focused Neural ODEs for Proactive Generator Deployment and Power Grid Resilience

Shuyi Chen
shuyic@alumni.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Ferdinando Fioretto
fioretto@virginia.edu
University of Virginia
Charlottesville, Florida, USA

Feng Qiu
fqiu@anl.gov
Argonne National Laboratory
Lemont, Illinois, USA

Shixiang Zhu
shixiangzhu@cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

ABSTRACT

Extreme weather events such as wildfires and hurricanes increasingly disrupt power infrastructure, leading to widespread outages and straining supply chains for emergency response. Traditional predict-then-optimize (PTO) frameworks sequentially forecast disruptions and then use these predictions to guide logistics and resource allocation. However, such two-stage approaches often suffer from misaligned objectives, resulting in suboptimal or delayed supply chain interventions. We propose a unified, decision-aware framework—Global-Decision-Focused (GDF) Neural ODEs—that integrates outage forecasting with proactive grid resilience planning. By modeling the spatiotemporal dynamics of power outages using neural ordinary differential equations, our approach embeds optimization objectives directly into the learning process, enabling strategic deployment of mobile generators. This predict-all-then-optimize-globally (PATOG) paradigm ensures system-wide consistency in both prediction and decision-making. Through experiments on real-world outage data and synthetic hazard scenarios, GDF demonstrates significant improvements in forecast quality, decision robustness, and recovery efficiency. Our results underscore the promise of integrated AI methods for resilient supply chain operations in power systems.

KEYWORDS

Outage forecasting, decision-focused learning, neural ODEs, generator deployment optimization, disaster recovery logistics

1 INTRODUCTION

Extreme weather events—such as wildfires, hurricanes, and winter storms—are increasingly disrupting electric power systems, causing prolonged outages and straining supply chains for emergency response [20, 22]. For instance, the January 2025 wildfires in Los Angeles County destroyed major power infrastructure and left over 400,000 customers without electricity [39]. Just months earlier, Hurricane Milton in October 2024 knocked out power for more than 3 million homes and businesses across Florida, underscoring the compounding threat of extreme weather events on grid reliability and recovery logistics [40].

In such crises, timely and effective generator deployment plays a critical role in supply chain resilience—powering hospitals, emergency shelters, and communication hubs while long-term repairs are underway. However, planning these deployments poses a significant challenge: operators must anticipate outages across a large region, coordinate scarce mobile resources, and respond under uncertainty. The ability to align outage predictions with effective deployment decisions is essential for minimizing service disruptions and accelerating recovery.

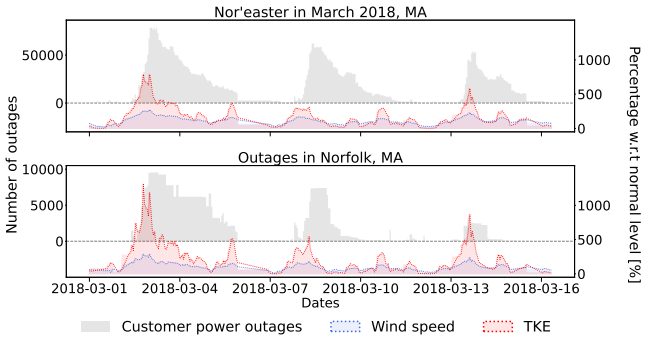

Figure 1: Number of outaged customers during the 2018 Nor'easter in Massachusetts, alongside key meteorological factors such as wind speed and turbulent kinetic energy.

A common approach in disaster logistics is the predict-then-optimize (PTO) framework [13], where forecasts of outages or infrastructure damage are used as inputs to a downstream optimization problem. Yet, in practice, this separation between prediction and decision-making introduces a misalignment. Forecasting models are often trained to minimize prediction error, not to optimize decision outcomes—meaning even small inaccuracies can translate into significant inefficiencies in resource allocation [16, 28]. For instance, overestimating outages in one region while underestimating them in another can cause critical supply misallocations—sending generators to areas that do not need them, while neglecting regions in crisis. These mismatches have real-world consequences, particularly in high-stakes infrastructure planning and humanitarian logistics [19, 30].

This challenge is exacerbated in system-wide recovery efforts, where independent regional forecasts must be aggregated to make globally coordinated supply chain decisions. Ensuring consistency across forecasts is difficult, and local prediction errors can propagate and undermine the overall decision quality. In the context of power

systems, this means reactive strategies may fail to deliver timely support to vulnerable areas, leading to higher restoration costs and prolonged service disruptions.

To overcome this limitation, we propose a new paradigm for disaster logistics, which we call *predict-all-then-optimize-globally (PATOG)*. Rather than separating forecasting and optimization, PATOG integrates them into a single learning objective. At the heart of this approach is our proposed Global-Decision-Focused (GDF) model, which combines neural ordinary differential equations (ODEs) with decision-focused learning to simultaneously simulate outage dynamics and optimize generator deployment policies. Drawing inspiration from compartmental models used in epidemiology, our method models power outage evolution within each region as a continuous-time dynamical system, influenced by weather, grid topology, and socioeconomic vulnerability.

Unlike traditional models that treat regions independently, GDF ensures that predictions are spatially and temporally aligned with a global decision objective—resulting in coherent, system-wide deployment plans. This integration directly addresses a core supply chain challenge: transforming complex, noisy forecasts into reliable operational strategies under uncertainty.

To evaluate our approach, we conduct experiments on both real-world and synthetic datasets. The real dataset captures outage events from a 2018 Nor'easter in Massachusetts, combining county-level outage records [1] with meteorological variables from NOAA's HRRR model [32] and socioeconomic indicators from the U.S. Census Bureau [42]. Figure 1 visualizes the outage counts and corresponding weather patterns during this event. Additionally, a synthetic dataset simulates outage progression using a simplified SIR-based model, enabling controlled experiments across different hazard intensities and grid configurations.

Our results demonstrate that GDF outperforms baseline methods in both predictive accuracy and downstream decision quality, leading to better generator deployment outcomes and improved grid resilience. By aligning learning with operational goals, our method offers a principled and scalable framework for managing supply chain decisions under climate-induced uncertainty.

We summarize our main contributions as follows:

- We introduce the predict-all-then-optimize-globally (PATOG) paradigm for decision-aware forecasting in grid infrastructure and recovery logistics;
- We propose a global-decision-focused (GDF) neural ODE model that aligns the goal of outage prediction with globally optimized generator deployment strategies;
- We validate our approach on both real and synthetic datasets, demonstrating superior performance over conventional predict-then-optimize methods in resilience and efficiency, uncovering key insights that inform more effective supply chain strategies for grid resilience strategies.

1.1 Related Work

This section reviews key advancements in outage modeling, decision-focused learning (DFL), and grid operation optimization, with an emphasis on their practical applications in power system resilience. Despite significant progress, existing methods separate forecasting from optimization, causing inefficiencies in decision-making. This review underscores the need for a unified framework that aligns predictive models with resilience objectives to enhance grid reliability in extreme natural hazards.

Grid Operation Optimization. Grid optimization has been widely studied to improve power system resilience, covering areas such as distributed generator (DG) placement [5, 12, 35], infrastructure reinforcement [36], and dynamic power scheduling [7, 9]. However, traditional approaches often follow a two-stage predict-and-mitigate paradigm—first forecasting system conditions and then optimizing responses [17]. This disconnect between prediction and optimization results in suboptimal grid operations, particularly under high uncertainty, where even small forecasting errors can cause significant deviations from the optimal response [19, 30].

To overcome these limitations, we propose integrating decision-focused learning (DFL) with predictive modeling. By embedding decision objectives directly into the learning process, our approach aligns predictions with optimization goals, enabling more adaptive and proactive resource allocation and grid reinforcement. This integration enhances grid resilience amid escalating natural hazard risks.

Power Outage Modeling. Accurate power outage forecasting is crucial for enhancing grid reliability and resilience [14, 27]. Various machine learning and statistical methods have been employed to predict outages under different conditions. These approaches incorporate neural networks enriched with environmental factors and semantic analysis of field reports, providing real-time updates and enhancing predictive performance through text analysis.

Additionally, ordinary differential equations have been widely used to model dynamic systems, such as outage propagation, capturing evolving disruptions under various conditions [10, 50]. For instance, adaptations of the Susceptible-Infected-Recovered (SIR) model from epidemiology have been applied to simulate outage propagation, drawing parallels between power failures and disease spread [34].

While these models provide valuable insights, they often lack the granularity needed for city- or county-level decision-making, limiting their practical application to localized resilience planning. By integrating local weather forecasts and socio-economic data into compartmental neural ODE models, our approach offers forecasting of local outage dynamics, enabling more targeted and effective interventions.

Decision-Focused Learning. Decision-focused learning (DFL) integrates predictive machine learning models with optimization, aligning training objectives with decision-making rather than purely maximizing predictive accuracy. Unlike traditional two-stage approaches, where predictions are first generated and then used as inputs for optimization, DFL enables end-to-end learning by back-propagating gradients through the optimization process. This is achieved via implicit differentiation of optimality conditions such as KKT constraints [6, 18] or fixed-point methods [24, 48]. For nondifferentiable optimizations, approximation techniques such as surrogate loss functions [13], finite differences [44], and noise perturbations [8] have been developed. Recent work has also explored integrating differential equation constraints directly into optimization models, enabling end-to-end gradient-based learning while ensuring compliance with system dynamic constraints [21, 43].

A well-studied class of DFL problems involves linear programs (LPs), where the Smart Predict-and-Optimize (SPO) framework [13] introduced a convex upper bound for gradient estimation, enabling cost-sensitive learning for optimization. Subsequent work has extended DFL to combinatorial settings, including mixed-integer programs (MIPs), using LP relaxations [29, 47]. Recent advances, such as decision-focused generative learning (Gen-DFL) [45], tackle the challenge of applying DFL in high-dimensional setting by using generative models to adaptively model uncertainty.

Differentiable Optimization. A key enabler of DFL is differentiable optimization (DO), which facilitates gradient propagation through differentiable optimization problems, aligning predictive models with decision-making objectives [4]. Recent advances extend DO to distributionally robust optimization (DRO) for handling uncertainty in worst-case scenarios, improving decision quality under data scarcity [11, 51]. Beyond predictive modeling, DO has advanced combinatorial and nonlinear optimization through implicit differentiation of KKT conditions [6], fixed-point methods [24], and gradient approximations via noise perturbation [8] and smoothing [44]. These techniques bridge forecasting with optimization, ensuring decision-aware learning. In this work, DO enables the backpropagation of resilience strategy losses, aligning the spatio-temporal outage prediction model with grid optimization objectives.

2 PREDICT ALL THEN OPTIMIZE GLOBALLY

In this section, we first provide an overview of decision-focused learning (DFL) and its application to solving predict-then-optimize (PTO) problems. We then introduce a new class of decision-making problems termed *predict-all-then-optimize-globally* (PATOG). Unlike traditional PTO approaches, which generate instantaneous or overly aggregated forecasts and optimize decisions independently, PATOG explicitly accounts for how predictions evolve over time and space, integrating them into a single, system-wide optimization framework.

PATOG is particularly useful for grid resilience management, where decisions must consider complex interactions across all service units. A key example is the mobile generator distribution problem. In a conventional PTO setting, potential damage from an extreme weather event is first forecasted for each unit independently. Decisions, such as scheduling power generator deployments, are then made in isolation, without considering the evolving conditions of other units. This localized approach often leads to resource misallocation and suboptimal resilience outcomes. In contrast, PATOG embeds these interdependencies into a global optimization problem, enabling system-wide decision-making that improves predictive models by incorporating cross-unit interactions. This results in more robust and effective resilience planning.

2.1 Preliminaries: Decision-Focused Learning

The predict-then-optimize (PTO) has been extensively studied across a wide range of applications [17, 28]. It follows a two-step process: First, predicting the unknown parameters $\mathbf{c}$ using a model f_θ based on the input $\boldsymbol{z}$, denoted as $\hat{\mathbf{c}} \coloneqq f_\theta(\boldsymbol{z})$. Second, solving an optimization problem:

$$\boldsymbol{x}^*(\hat{\mathbf{c}}) = \arg\min_{\boldsymbol{x}} g(\boldsymbol{x}, \hat{\mathbf{c}}), \tag{1}$$

where g is the objective function and $\boldsymbol{x}^*(\hat{\mathbf{c}})$ represents the optimal decision given the predicted parameters. This framework has numerous practical applications in power grid operations. For example, PTO is used in power grid operations to predict system stress using synchrophasor data, optimize outage management by forecasting disruptions and proactively dispatching restoration crews, and enhance renewable integration by predicting fluctuations in wind and solar generation to improve scheduling and grid balancing [17]. These predictive insights enable system operators to make informed strategic decisions, enhancing grid reliability and resilience.

However, the conventional two-stage approach does not always lead to high-quality decisions. In this approach, the model parameter θ is first trained to minimize a predictive loss, such as mean squared error. Then the predicted parameters $\hat{\mathbf{c}}$ are used to solve the downstream optimization. This separation between prediction and optimization can result in suboptimal decisions, as the prediction model is not directly optimized for decision quality [25].

To address this limitation, the decision-focused learning (DFL) integrates prediction with the downstream optimization process [28]. Instead of optimizing for predictive accuracy, DFL trains the model parameter θ by directly minimizing the *decision regret* [28]:

$$\theta^* = \arg\min_{\theta} \mathbb{E}\left[g(\boldsymbol{x}^*(f_\theta(\boldsymbol{z})), \mathbf{c}) - g(\boldsymbol{x}^*(\mathbf{c}), \mathbf{c})\right].$$

This approach ensures that the model is learned with the ultimate goal of improving decision quality, making it particularly effective for PTO problems. Note that we assume that the constraints on decision variable $\boldsymbol{x}$ are fully known and do not depend on the uncertain parameters $\mathbf{c}$ in this study. This assumption simplifies the problem by ensuring that all feasible solutions $\boldsymbol{x}$ remain valid regardless of the parameter estimates.

2.2 Proposed PATOG Framework

The objective of PATOG in this work is to develop proactive global recourse actions that enable system operators to better prepare for natural hazards. These actions may include preemptive dispatch of mobile generators, strategic load shedding, grid reconfiguration, or reinforcement of critical infrastructure. The PATOG consists of two steps: (*i*) Predicting the temporal evolution of unit functionality across all the service units in the network throughout the duration of a hazard event. (*ii*) Deriving system-wide strategies that minimize overall loss based on all the predictions, enabling optimized resource allocation by anticipating critical failures before they occur.

Consider a power network consisting of K geographical units, where each unit k serves N_k customers. We define the global recourse actions as $\boldsymbol{x} \coloneqq \{x_k\}_{k=1}^K$, where x_k represents the action taken for unit k. A key challenge in designing effective actions is understanding how the system will respond to an impending hazard event. To this end, we use the number of customer power outages, which is publicly accessible via utility websites, as a measure of system functionality [17].

To model the outage dynamics, we represent the outage state of each unit k using a dynamical system over the time horizon $[0, T]$ during a hazard event. During the event, the state of each unit k at time t is represented by three quantities:

- $Y_k(t) \in \mathbb{Z}_*$: the number of customers experiencing outages;
- $U_k(t) \in \mathbb{Z}_*$: the number of unaffected customers;

- $R_k(t) \in \mathbb{Z}_*$: the number of recovered customers.

The total number of customers, N_k, in the unit k remains constant throughout the event, satisfying the following constraint:

$$Y_k(t) + U_k(t) + R_k(t) = N_k, \ \forall t \in [0, T].$$

For compact representation, we define the state vector for unit k as $\mathrm{S}_k(t) \coloneqq [U_k(t), R_k(t), Y_k(t)]^\top$. To simplify notation, we collectively represent the outage states across all units over the time horizon as $\mathrm{S} \coloneqq \{\mathrm{S}_k(t) \mid t \in [0, T]\}_{k=1}^K$.

Formally, we are tasked to search for the optimal action:

$$\boldsymbol{x}^*(\mathrm{S}) = \arg\min_{\boldsymbol{x}} \ g(\boldsymbol{x}, \mathrm{S}) \tag{2}$$

$$\text{s.t.} \quad \frac{\mathrm{d}\mathrm{S}_k(t)}{\mathrm{d}t} = f_\theta(\mathrm{S}_k(t), \boldsymbol{z}_k), \ \forall k, \tag{3}$$

$$\mathrm{S}_k(0) = [N_k, 0, 0]^\top, \ \forall k, \tag{4}$$

where g quantifies the decision loss of the action $\boldsymbol{x}$ based on the predicted future evolution of outage states S. The transition function f_θ models the progression of unaffected, recovered, and outaged customers in each unit over time, influenced by both local weather conditions and socioeconomic factors, jointly represented as covariates $\boldsymbol{z}_k \in \mathbb{R}^p$. We note that most power outages during extreme weather stem from localized transmission line damage, with cascading failures being rare [37, 52]. Thus, we assume each unit evolves independently under its local conditions.

We emphasize that (2) extends the traditional PTO framework by integrating predictions across multiple units over an extended future horizon to derive a single, globally optimized solution. Unlike PTO, which focuses only on instantaneous or localized dynamics, PATOG captures both temporal and spatial outage states while modeling how decisions influence outage dynamics across the entire system. This comprehensive approach enables proactive, system-wide high-resolution resilience strategies that adapt to evolving conditions.

3 GLOBAL-DECISION-FOCUSED NEURAL ODES

This section presents a novel decision-focused neural ordinary differential equations (ODE) model tailored for solving PATOG problems in grid resilience management. The proposed neural ODE model predicts outage progression at the unit level while simultaneously optimizing global operational decisions by learning model parameters in a decision-aware manner. We refer to this approach as global-decision-focused (GDF) Neural ODEs. Fig. 2 provides an overview of the proposed framework.

3.1 Neural ODEs for Power Outages

Assume that we have observations of I natural hazards (*e.g.*, hurricanes, and winter storms) in the history. For the i-th event, the observations in unit k are represented by a data tuple, denoted by $(\boldsymbol{z}_k^i, \{y_k^i(t)\})$, where $\boldsymbol{z}_k^i$ represents the covariates for unit k during the i-th event and $y_k^i(t)$ is the number of customers experiencing power outages at time t. The outage trajectory $\{y_k^i(t) \mid t \in [0, T^i]\}$ is recorded at 15-minute intervals. A significant challenge in modeling outage dynamics is the lack of detailed observations for the underlying failure and restoration processes. Specifically, while

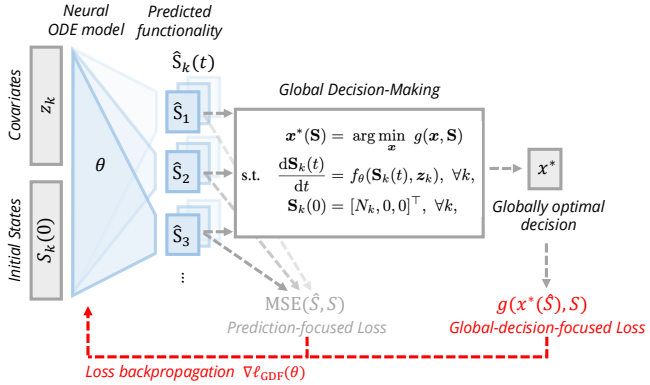

Figure 2: Overview of the proposed GDF framework. Given covariates $\boldsymbol{z}_k$ and initial states $\mathrm{S}_k(0)$ for all K service units, a model parameterized by θ predicts the system states $\hat{\mathrm{S}}_k$ for all units. These predictions inform global decision-making, where optimal actions $\boldsymbol{x}^*(\hat{\mathrm{S}})$ minimize the global decision loss $g(\boldsymbol{x}, \hat{\mathrm{S}})$. The framework optimizes θ by minimizing a global-decision-focused loss, regularized by a prediction-focused loss (*e.g.*, MSE loss) to enhance predictive interpretability. Red arrows denote the backpropagation through $\nabla \ell_{\mathsf{GDF}}(\theta)$, ensuring that the model learns both system-level decision quality and region-specific prediction accuracy.

$y_k^i(t)$ provides the number of customers experiencing outages at time t, we do not directly observe the failure and restoration states, $U_k(t)$ and $R_k(t)$, respectively.

Drawing inspiration from the Susceptible Infectious Recovered (SIR) models commonly used in epidemic modeling [23], we conceptualize power outages within a unit as the spread of a "virus". In this analogy, outages propagate among customers due to local transmission line failures, while restorations provide lasting resistance to subsequent outages. We formalize this analogy with the following three assumptions: (*i*) The number of unaffected customers $U_k(t)$ decreases over time as some customers transition from being unaffected to experiencing outages. This transition is governed by the *failure transmission rate*, denoted by $\phi(\boldsymbol{z}_k; \theta_U)$. Inspired by epidemiological transmission rates, this rate quantifies how local conditions – such as weather patterns and other regional factors encapsulated by covariates $\boldsymbol{z}_k$ – influence the rate at which outages spread within the grid. (*ii*) Conversely, the number of customers with restored power ($R_k(t)$) gradually increases as the system operator repairs transmission lines and restores service. This process is captured by the *restoration rate*, denoted $\phi(\boldsymbol{z}_k; \theta_R)$. Both transmission and restoration rates, $\phi(\boldsymbol{z}_k; \theta_U)$ and $\phi(\boldsymbol{z}_k; \theta_R)$, are functions of the local covariates $\boldsymbol{z}_k$, and are modeled using deep neural networks. (*iii*) The total number of customers within each unit remains constant throughout the studied period. Based on these assumptions, we model the outage state transition for each unit k in (3), *i.e.*, $\mathrm{d}\mathrm{S}_k(t)/\mathrm{d}t$, as follows:

$$\begin{cases} \mathrm{d}U_k(t)/\mathrm{d}t = -\phi(\boldsymbol{z}_k; \theta_U) Y_k(t) U_k(t), \\ \mathrm{d}R_k(t)/\mathrm{d}t = \phi(\boldsymbol{z}_k; \theta_R) Y_k(t), \\ \mathrm{d}Y_k(t)/\mathrm{d}t = -\mathrm{d}U_k(t)/\mathrm{d}t - \mathrm{d}R_k(t)/\mathrm{d}t. \end{cases} \tag{5}$$

These ODEs capture the dynamic evolution of power outage states within each unit k. For notational simplicity, we use $\theta := \{\theta_U, \theta_R\}$ to denote their parameters jointly.

In practice, we work with discrete observations $\{t_j\}_{j=0}^{T}$ and adopt the Euler method to approximate the solution to the ODE model [10], *i.e.*,

$$\mathrm{S}_k(t_{j+1}) = \mathrm{S}_k(t_j) + f_\theta(\mathrm{S}_k(t_j), z_k)\Delta t_j, \quad \forall k,$$

where $\mathrm{S}_k(t_{j+1})$ and $\mathrm{S}_k(t_j)$ are the history states at times t_{j+1} and t_j, respectively, and $\Delta t_j := t_{j+1} - t_j$ is time interval.

3.2 Global-Decision-Focused Learning

The training loss for GDF Neural ODEs is formulated as the aggregated regret across all unit k, with an additional regularization term to ensure stability in prediction:

$$\ell_{\mathsf{GDF}}(\theta) := \frac{1}{I}\sum_i \left[g\Big(\boldsymbol{x}^*(\hat{\mathrm{S}}^i), \mathrm{S}^i\Big) - g\Big(\boldsymbol{x}^*(\mathrm{S}^i), \mathrm{S}^i\Big)\right] + \lambda \cdot \frac{1}{IKT}\sum_{i,k,j}\left[y_k^i(t_j) - \hat{Y}_k^i(t_j)\right]^2, \tag{6}$$

where $\hat{\mathrm{S}}^i$ is the predicted outages states across all units for event i.

The objective function (6) consists of two key components: (*i*) Global-decision-focused loss (first term): This term evaluates the quality of the optimal action based on predictions, capturing the impact of prediction errors on operational decisions. The loss and its gradient are computed over all geographical units affected by the event, ensuring that learning is guided by system-wide decision quality. However, this loss alone does not provide direct insights into the structure of outage trajectories, which may limit the interpretability of the learned model. (*ii*) Prediction-focused loss (second term): To address this limitation, a prediction-based penalty is introduced to minimize discrepancies between observed outage trajectories $y_k^i(t)$ and predicted values $\hat{Y}_k^i(t)$. This term refines the model's ability to capture outage dynamics without explicitly observing the failure and restoration processes. A user-specified hyperparameter λ governs the trade-off between prediction accuracy and the regret associated with suboptimal operational decisions.

The most salient feature of the proposed GDF neural ODEs method is its incorporation of both global and local perspectives. The global-decision-focused loss aggregates errors across all geographical regions and time steps, directly linking prediction quality to system-wide resilience measures and operational strategies (*e.g.*, resource dispatch, outage management, and service restoration). Meanwhile, the prediction-focused component refines local accuracy by penalizing deviations at each service unit. By incorporating predictive regularization, the model empirically improves generalization to new events with unknown distributional shifts, mitigating overfitting, particularly given the limited availability of extreme event data.

3.3 Model Estimation

The learning of GDF Neural ODEs is carried out through stochastic gradient descent, where the gradient is calculated using a novel algorithm based on differentiable optimization techniques [4, 6, 51]. To enable differentiation through the arg min operator in (2) embedded in the global-decision-focused loss, we relax the decision variable $\boldsymbol{x}$ from a potentially discrete space to a continuous space. For combinatorial optimization problems, the problem is reformulated as a differentiable quadratic program, and a small quadratic regularization term is added to ensure continuity and strong convexity [47]. Formally, we replace the original objective in (2) with the following:

$$\min_{\boldsymbol{x}} g(\boldsymbol{x}, \mathrm{S}) + \rho\|\boldsymbol{x}\|_2^2, \tag{7}$$

where $\rho > 0$ ensures differentiability. The full training procedure is summarized in Algorithm 1. More implementation details can be found in the supplementary material.

Algorithm 1 Learning of GDF Neural ODEs

Input: Data $\mathcal{D} = \{(\boldsymbol{z}_k^i, \{y_k^i(t)\})\}$, initial parameter θ_0, initial states $\mathrm{S}^i(0)$, learning rate η, trade-off parameter λ, epochs N.
Output: θ^*

for epoch $n = 1, \ldots, N$ **do**
  for $i = 1$ to I **do**
    Initialize: $\hat{\mathrm{S}}^i(0) \leftarrow \mathrm{S}^i(0)$.
    for $j = 1$ to T_i **do**
      $\hat{\mathrm{S}}^i(t_j) \leftarrow \hat{\mathrm{S}}^i(t_{j-1}) + f_\theta(\hat{\mathrm{S}}^i(t_{j-1}), \boldsymbol{z}_k^i)\Delta t_{j-1}$
    end for
    $\ell_{\mathsf{GDF}}^i(\theta) \leftarrow g\big(\boldsymbol{x}^*(\hat{\mathrm{S}}^i), \mathrm{S}^i\big) - g\big(\boldsymbol{x}^*(\mathrm{S}^i), \mathrm{S}^i\big)$
    $\theta \leftarrow \theta - \eta\,\nabla_\theta \ell_{\mathsf{GDF}}^i(\theta)$.
  end for
  for each mini-batch $\mathcal{B} \subset \mathcal{D}$ **do**
    $\ell_{\mathrm{Pred}}(\theta) \leftarrow \frac{1}{|\mathcal{B}|KT}\sum_{(i,k,t)\in\mathcal{B}}\left[y_k^i(t) - \hat{Y}_k^i(t)\right]^2$
    $\theta \leftarrow \theta - \eta\,\lambda\,\nabla_\theta \ell_{\mathrm{Pred}}(\theta)$.
  end for
end for
return θ

4 APPLICATION: MOBILE GENERATOR DEPLOYMENT

Our method is broadly applicable to various proactive decision-making problems in grid operations, particularly in response to potential hazard events. This section highlights the adaptability of GDF by applying it to a mobile generator deployment task [5, 12, 35], a representative PATOG problem.

The objective of the mobile generator deployment problem is to strategically deploy mobile (*e.g.*, diesel) generators across a network of potentially affected sites before a large-scale power outage to minimize associated costs. Due to the time-sensitive nature of power restoration, it is crucial to anticipate the spatiotemporal dynamics of outages while accounting for operational constraints such as generator capacity, fuel availability, and transportation logistics. GDF is well-suited for this task as it jointly learns outage patterns across cities over the planning horizon while optimizing deployment decisions, enabling proactive decision-making that adapts to evolving conditions.

Formally, let $Q_w \in \mathbb{Z}_+$ denote the initial inventory of generators at warehouse w. For simplicity, we assume that warehouses also function as staging areas, where generators are initially stored and returned after deployment. We assume that each generator has a fixed maximum capacity, capable of supplying electricity

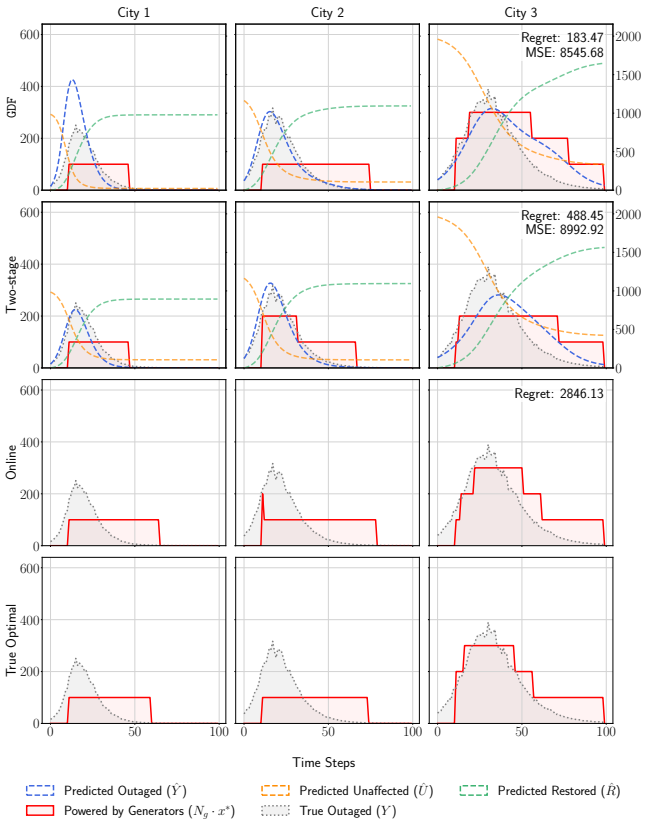

Figure 3: A synthetic example of the mobile generator deployment problem for a system with three cities and five generators ($Q_w = 5$). The y-axis represents the number of outaged households. In this example, the uniform travel time $\delta_t = 10$, transportation cost is set to $c = 400$, the customer interruption cost is $\tau = 1$, and the operational cost is $\gamma = 2$. Four methods are compared on out-of-sample data: the proposed GDF framework (regret = 183.47), a two-stage approach (regret = 488.45), an online baseline with observation lag of 5 (regret = 2846.13), and the optimal solution with the groundtruth. Details of the synthetic data setup are provided in Section 5.1.

to N_g customers, and that a single extreme event is anticipated. The planning horizon is discretized into uniform time periods $\mathcal{T} = \{1, 2, \ldots, T\}$, where T represents the expected duration of the outage. Let $\mathcal{K} = \{1, 2, \ldots, K\}$ be the set of K service units (*e.g.*, cities or counties) where generators can be deployed, and $\mathcal{W} = \{1, \ldots, W\}$ be the set of warehouses where all generators are initially stored. We denote the stock of generators in unit k at time t as q_{tk}.

The mobile generator deployment problem is therefore defined on a directed graph denoted by $(\mathcal{V}, \mathcal{E})$, where $\mathcal{V} := \mathcal{K} \cup \mathcal{W}$ is the vertex set, and $\mathcal{E}$ represents the edge set defining feasible transportation routes. Each edge $(k, k') \in \mathcal{E}$ incurs a transportation cost $c_{kk'}$, and an associated travel time $\delta_{kk'} \in \mathbb{Z}_+$, representing the number of time periods required for generators to be transported from location k to location k'. The decision variables, $\boldsymbol{x} = \{x_{tkk'}\}$, $(k, k') \in \mathcal{E}$, $t \in \mathcal{T}$, specifies the transportation schedule, where $x_{tkk'} \in \mathbb{Z}_+$ represents the number of generators transported from location k to k' at time t. For simplicity, we assume negligible deployment time, allowing generators to become operational immediately upon arrival to supply power to N_g customers.

The cost function of generator deployment problem is:

$$g(\boldsymbol{x}, \mathbf{S}) := \underbrace{\sum_{(k,k')\in\mathcal{E}} \sum_{t=1}^{T} c_{kk'} x_{tkk'}}_{\text{Transportation Cost}} + \underbrace{\gamma \sum_{k\in\mathcal{K}} \sum_{t=1}^{T} q_{tk}}_{\text{Operation Cost}} + \underbrace{h(\boldsymbol{x}, \mathbf{S})}_{\text{Outage Cost}}, \tag{8}$$

which consists of three components: (*i*) Transportation Cost: The cost of delivering generators from warehouses to service units. (*ii*) Operation Cost: The fixed operational cost per unit time (γ) for deployed generators. (*iii*) Outage Cost: Economic loss due to power outages, determined by the number of customers experiencing outages given the generator deployment decision $\boldsymbol{x}$. For simplicity, we define the outage cost as:

$$h(\boldsymbol{x}, \mathbf{S}) = \tau \sum_{k\in\mathcal{K}, t\in\mathcal{T}} \max\left\{Y_k(t) - q_{tk} N_g, 0\right\},$$

where τ denotes unit customer interruption cost, an adapted version of the Value of Lost Load (VoLL) [38], representing the economic impact per household outage per day, and $Y_k(t)$ denotes the number of outages in service unit k at time t, extracted from the state $\mathbf{S}$ according to (5). This framework primarily evaluates system functionality based on the number of customers experiencing outages (Y). It is worth emphasizing that it is flexible to incorporate additional metrics, including the number of recovered customers (R) and unaffected customers (U), if needed.

The transportation and inventory constraints of the mobile genereator deployment are as follows:

$$q_{0k} = 0, \quad \forall k \in \mathcal{K} \tag{9}$$

$$q_{0w} = Q_w, \quad \forall w \in \mathcal{W} \tag{10}$$

$$q_{tk} = \sum_{k'\in\mathcal{V}} \sum_{\substack{t'\in\mathcal{T} \\ t'+\delta_{k'k}\le t}} x_{t'k'k} - \sum_{k'\in\mathcal{V}} x_{tkk'}, \quad \forall k \in \mathcal{V},\ t \in \mathcal{T} \tag{11}$$

$$x_{tkk'} \le C, \quad \forall t \in \mathcal{T}, (k, k') \in \mathcal{E} \tag{12}$$

$$\sum_{k\in\mathcal{V}/\{k'\}, t\in\mathcal{T}} x_{tkk'} = \sum_{k\in\mathcal{V}/\{k'\}, t\in\mathcal{T}} x_{tk'k}, \quad \forall k' \in \mathcal{V}, \tag{13}$$

$$q_{kk'}, x_{tkk'} \in \mathbb{Z}_+, \quad \forall (k, k') \in \mathcal{E}, \forall t \in \mathcal{T}. \tag{14}$$

Constraints (9) and (10) establish the initial generator stock at each location. Equation (11) tracks the generator stock at each location, accounting for the travel time δ taken for incoming shipments. Inequality constraint (12) limits the flow between service units or warehouses in $\mathcal{E}$ to a maximum capacity C. Finally, equation (13) ensures flow conservation, requiring that the total inflow of generators equals the total outflow over the planning horizon for each node in the network. As a result, all generators return to the staging areas after the events.

To implement the deployment strategy based on our GDF framework, we first predict outage levels $\hat{Y}_k(t)$ for all service units using

the neural ODE model specified in (5). The model is learned by minimizing the GDF loss defined in (6). These global-deision-focused predictions are then integrated into the objective function (8) of the mobile generator deployment problem, and the decisions are derived by solving a mixed-integer linear programming (MILP). Figure 3 illustrates the optimal transportation schedules derived from the GDF predictions using an MILP solver, compared to the schedules produced by the Two-stage and reactive online approaches in a stylized example of the mobile generator deployment problem.

5 EXPERIMENTS

In this section, we evaluate the proposed GDF on two grid resilience management problems—mobile generator deployment and power line undergrounding. The numerical results demonstrate its superior performance compared to conventional Two-stage methods, enabling better decision-making in the face of natural hazards.

5.1 Dataset and Experimental Setup

We evaluate the proposed GDF framework using both real and synthetic datasets. The real dataset captures county-level customer outages during the 2018 Nor'easter in Massachusetts [46], augmented with meteorological features from NOAA's HRRR model [32] and socioeconomic variables from the U.S. Census Bureau's American Community Survey [42]. These include wind speed, temperature, pressure, income, age, unemployment rate, poverty rate, college enrollment, and household size—factors that affect outage propagation and recovery. The event includes three storms over a 15-day period, allowing us to treat them as arising from the same data distribution. We use the first storm for training and the second for testing, omitting the third due to its minor impact. Fig. 1 illustrates the spatiotemporal patterns.

To support controlled experimentation under diverse scenarios, we generate synthetic datasets using a simplified SIR model, treating each county as an independent population subject to outages and recovery, governed by the dynamics in (5). Transmission rates are modulated by simulated weather to mimic varying hazard severity. This setup enables systematic stress-testing of resilience strategies across configurations.

Experiments focus on two applications: mobile generator deployment and power line undergrounding [3, 15, 33, 41]. Predictions are made for the entire horizon before the hazard surpasses a predefined threshold (1%), reflecting real-world challenges in data availability [49]. For the generator problem, we assume a centralized warehouse as the only source of units, prohibiting inter-county transfers. Generators must return for refueling, and a uniform travel delay δ_t is assumed across all routes. A myopic online baseline with an observation lag is used to mimic reactive deployment.

Model training consists of two stages: first, a neural ODE is trained with standard MSE loss to establish a predictive baseline (Two-stage). Then, it is fine-tuned using the decision-focused objective in (6) via Algorithm 1. We compare GDF against the Two-stage and ground truth optimal solutions, evaluating performance using (i) MSE for prediction accuracy, (ii) deployment cost or SAIDI for decision loss, and (iii) regret relative to the true optimal. For synthetic settings, results are averaged over three random seeds with different outage and weather profiles.

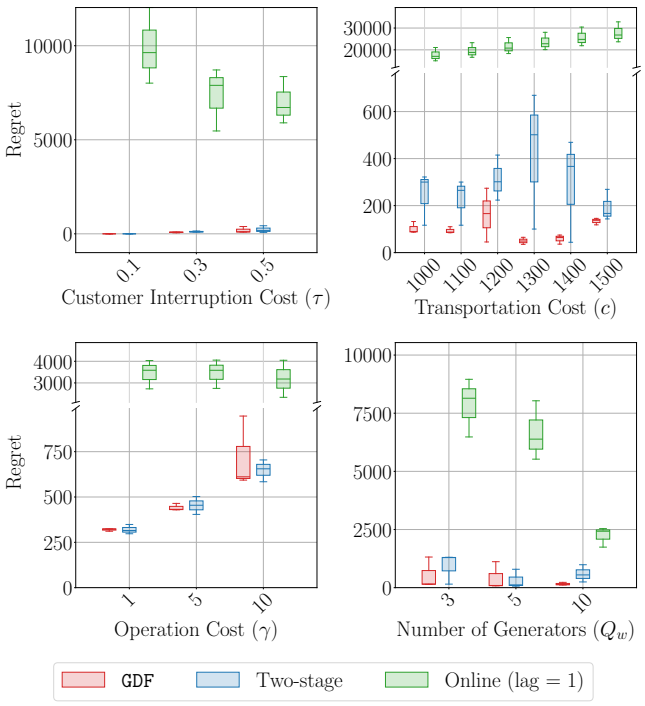

Figure 4: Performance Comparison for Synthetic Mobile Generator Deployment: A detailed comparison of regret outcomes for GDF, Two-Stage, and Online methods under varying customer interruption costs (τ), transportation cost factors (c), operational costs (γ), and numbers of generators (Q_w).

5.2 Results

This section presents the results of GDF on mobile generator deployment and power line undergrounding, evaluated in terms of predictive accuracy and decision quality. Results on both synthetic and real datasets demonstrate that GDF improves decision-making compared to conventional Two-stage methods, enabling a more effective response to natural hazards.

5.2.1 Mobile Generator Deployment Problem. Table 1 summarizes the out-of-sample performance for the generator deployment problem on synthetic data under different travel time settings (δ_t = 1, 5, 10). The proposed method consistently delivers the best decision quality, as reflected in the lowest regret. Notably, as δ_t increases—representing longer transportation delays—the performance improvement of GDF over the two-stage approach becomes increasingly significant. A similar trend is observed under higher transportation cost factors, as shown in Fig. 4, until transportation costs become so high that the best strategy is effectively to remain idle, at which point the improvement from proactive actions with GDF diminishes.

These findings highlight the importance of proactive scheduling and early interventions when facing longer deployment delays or higher transportation costs. It also demonstrates the advantage of GDF in optimizing resource allocation under more challenging, time-sensitive conditions in the face of extreme hazards that threaten

Table 1: Out-of-Sample Performance for Generator Distribution Problem with Synthetic Data. Results are averaged over 3 repeated experiments with standard error (SE) in brackets.

Model	$\delta_t = 1$			$\delta_t = 5$			$\delta_t = 10$		
	MSE	Cost	Regret	MSE	Cost	Regret	MSE	Cost	Regret
Ground Truth	/	10316.9 (73.9)	0	/	11031.7 (69.0)	0	/	12473.7 (63.8)	0
Online (lag = 1)	/	63663.2 (3762.5)	53346.3 (3769.9)	/	18862.8 (965.7)	7831.1 (916.1)	/	19681.8 (2069.3)	7208.1 (2121.0)
Online (lag = 3)	/	84769.4 (122.8)	74452.6 (7246.8)	/	19816.9 (73.6)	8785.2 (1216.9)	/	38378.6 (3020.2)	4962.5 (792.7)
Two-stage	9336.0 (975.6)	10459.8 (90.6)	142.9 (90.3)	9336.0 (975.6)	11506.6 (98.1)	474.9 (95.7)	9336.0 (975.6)	12746.2 (196.7)	272.4 (207.2)
GDF	**7204.1 (1804.8)**	10409.9 (79.7)	**93.0 (11.8)**	**7531.5 (1805.8)**	11452.1 (80.3)	**420.3 (24.3)**	**8907.7 (2645.2)**	12601.9 (58.1)	**128.2 (49.3)**

Table 2: Out-of-Sample Performance of Power Line Undergrounding Problem with Real and Synthetic Data

Models	Nor'easter, MA, 2018			Synthetic		
	MSE ($\times 10^3$)	SAIDI	Regret	MSE ($\times 10^3$)	SAIDI	Regret
True Optimal	/	218.2	/	/	15.6	/
Two-stage	**117.4**	328.3	112.1	13.5	16.1	0.5
GDF	165.6	**312.7**	**94.5**	**13.1**	**15.6**	**0**

grid resilience. As visualized in Fig. 3, the GDF model slightly overestimate outages in critical regions first hit by outage, enabling the reallocation of additional resources to mitigate potential disruptions. While this adjustment in the forecast is subtle, it leads to a significant reduction in regret. Remarkably, the GDF models also achieve MSE comparable to that of the MSE-trained model in Table 1, albeit with higher variance. This indicates that improved decision quality was not achieved at the expense of predictive accuracy but rather through targeted and meaningful adjustments in the forecasts.

Additionally, we conducted extensive experiments on synthetic data for the mobile generator deployment problem under various settings, as shown in Fig. 4. We observed that the performance advantage of GDF over the Two-stage and greedy methods narrows as customer interruption costs, operational costs, or the number of available generators increase. This behavior is expected, as reduced flexibility or higher system resources diminish the need for globally optimized interventions—resulting in less improvement with GDF.

5.2.2 Power Line Undergrounding. Table 2 presents the out-of-sample performance on both the Nor'easter, MA, 2018 event and the synthetic dataset. For the real dataset, the proposed GDF model—despite a slightly higher MSE—delivers improved decision quality, achieving a lower SAIDI and reduced regret compared to the Two-stage baseline. On the synthetic dataset, GDF similarly outperforms the Two-stage method in decision quality while maintaining a low MSE, possibly due to the relatively simple prediction tasks. These results confirm that GDF leads to better overall decision performance in real and synthetic settings.

Overall, the experimental results on both real and synthetic data demonstrate that GDF achieves substantial improvements in decision quality compared to the traditional two-stage approach. Although both models exhibit comparable prediction accuracy, GDF consistently yields lower regret values, highlighting the benefits of proactive scheduling and decision-focused training.

6 CONCLUSION

This paper presented a novel framework for grid resilience management that jointly optimized prediction accuracy and global decision quality. Experimental results on both real and synthetic datasets demonstrated that, while achieving comparable MSE results to the conventional two-stage approach, GDF consistently yielded lower regret and better decisions. These findings suggest that integrating decision-focused learning enhances proactive scheduling and resource allocation, enabling system operators to more effectively mitigate outages. Future work will explore integrating real-time data streams to improve responsiveness, incorporating renewable energy forecasting, and adapting the framework to larger network systems.

Our research provides actionable insights for grid resilience practitioners

Our findings offer practical implications for supply chain and infrastructure resilience, especially in scenarios with costly or delayed logistics. By embedding decision-aware objectives into forecasting, the model enables earlier and more strategic allocation of limited resources across the network. This proactive approach is critical for mitigating disruptions, minimizing economic losses, and supporting robust, data-driven planning in high-stakes operational settings. Future work includes extending to real-time data integration, renewable forecasting, and scaling to larger networks.

REFERENCES

[1] 2020. *Massachusetts Power Outages.* Technical Report. Massachusetts Emergency Management Agency, Massachusetts, USA.

[2] 2022. IEEE Guide for Electric Power Distribution Reliability Indices. *IEEE Std 1366-2022 (Revision of IEEE Std 1366-2012)* (2022), 1–44. https://doi.org/10.1109/IEEESTD.2022.9955492

[3] Nicholas Abi-Samra, Lee Willis, and Marvin Moon. 2013. *Hardening the System.* https://www.tdworld.com/vegetation-management/article/20962556/hardening-the-system

[4] A. Agrawal, B. Amos, S. Barratt, S. Boyd, S. Diamond, and Z. Kolter. 2019. Differentiable Convex Optimization Layers. In *Advances in Neural Information Processing Systems.*

[5] Nabil A. Ahmed and Mohamad F. AlHajri. 2019. Distributed Generators Optimal Placement and Sizing in Power Systems. In *2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS).* 1–5. https://doi.org/10.1109/ICETAS48360.2019.9117454

[6] Brandon Amos and J. Zico Kolter. 2017. OptNet: Differentiable Optimization as a Layer in Neural Networks. In *Proceedings of the 34th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 70).* PMLR, 136–145.

[7] H. Aydin, R. Melhem, D. Mosse, and P. Mejia-Alvarez. 2001. Dynamic and aggressive scheduling techniques for power-aware real-time systems. In *Proceedings 22nd IEEE Real-Time Systems Symposium (RTSS 2001) (Cat. No.01PR1420).* 95–105. https://doi.org/10.1109/REAL.2001.990600

[8] Quentin Berthet, Mathieu Blondel, Olivier Teboul, Marco Cuturi, Jean-Philippe Vert, and Francis Bach. 2020. Learning with differentiable pertubed optimizers. *Advances in neural information processing systems* 33 (2020), 9508–9519.

[9] Shengrong Bu, F. Richard Yu, Peter X. Liu, and Peng Zhang. 2011. Distributed Scheduling in Smart Grid Communications with Dynamic Power Demands and Intermittent Renewable Energy Resources. In *2011 IEEE International Conference on Communications Workshops (ICC)*. 1–5. https://doi.org/10.1109/iccw.2011.5963580

[10] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. 2018. Neural ordinary differential equations. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (Montréal, Canada) *(NIPS'18)*. Curran Associates Inc., Red Hook, NY, USA, 6572–6583.

[11] Shuyi Chen, Kaize Ding, and Shixiang Zhu. 2025. Uncertainty-Aware Robust Learning on Noisy Graphs. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1–5. https://doi.org/10.1109/ICASSP49660.2025.10888672

[12] Rashmi Deshmukh and Amol Kalage. 2018. Optimal Placement and Sizing of Distributed Generator in Distribution System Using Artificial Bee Colony Algorithm. In *2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN)*. 178–181. https://doi.org/10.1109/GCWCN.2018.8668633

[13] Adam N. Elmachtoub and Paul Grigas. 2022. Smart "Predict, then Optimize". *Management Science* 68, 1 (Jan. 2022), 9–26. https://doi.org/10.1287/mnsc.2020.3922 Publisher: INFORMS.

[14] Rozhin Eskandarpour and Amin Khodaei. 2017. Machine Learning Based Power Grid Outage Prediction in Response to Extreme Events. *IEEE Transactions on Power Systems* 32, 4 (2017), 3315–3316. https://doi.org/10.1109/TPWRS.2016.2631895

[15] Peter Fairley. 2018. Utilities bury transmission lines. *IEEE Spectrum* 55, 2 (2018), 9–10. https://doi.org/10.1109/MSPEC.2018.8278121

[16] Carlos Fernández-Loría and Foster Provost. 2022. Causal decision making and causal effect estimation are not the same… and why it matters. *INFORMS Journal on Data Science* 1, 1 (2022), 4–16.

[17] Jay Giri. 2015. Proactive Management of the Future Grid. *IEEE Power and Energy Technology Systems Journal* 2, 2 (2015), 43–52. https://doi.org/10.1109/JPETS.2015.2408212

[18] Stephen Gould, Richard Hartley, and Dylan Campbell. 2021. Deep declarative networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 8 (2021), 3988–4004.

[19] Wangyi Guo, Zhanbo Xu, Zhequn Zhou, Jinhui Liu, Jiang Wu, Haoming Zhao, and Xiaohong Guan. 2024. Integrating end-to-end prediction-with-optimization for distributed hydrogen energy system scheduling*. In *2024 IEEE 20th International Conference on Automation Science and Engineering (CASE)*. 2774–2779. https://doi.org/10.1109/CASE59546.2024.10711393

[20] John Handmer, Yasushi Honda, Zbigniew W Kundzewicz, Nigel Arnell, Gerardo Benito, Jerry Hatfield, Ismail Fadl Mohamed, Pascal Peduzzi, Shaohong Wu, Boris Sherstyukov, et al. 2012. Changes in impacts of climate extremes: human systems and ecosystems. In *Managing the risks of extreme events and disasters to advance climate change adaptation special report of the intergovernmental panel on climate change*. Intergovernmental Panel on Climate Change, 231–290.

[21] Alexandre Jacquillat, Michael Lingzhi Li, Martin Ramé, and Kai Wang. 2024. Branch-and-Price for Prescriptive Contagion Analytics. *Operations Research* Ahead of Print (2024). https://doi.org/10.1287/opre.2023.0308

[22] Alyson Kenward and Urooj Raja. 2014. Blackout: Extreme weather climate change and power outages. *Climate central* 10 (2014), 1–23.

[23] Chrysoula Kosma, Giannis Nikolentzos, George Panagopoulos, Jean-Marc Steyaert, and Michalis Vazirgiannis. 2023. Neural Ordinary Differential Equations for Modeling Epidemic Spreading. *Transactions on Machine Learning Research* (2023).

[24] James Kotary, My H. Dinh, and Ferdinando Fioretto. 2023. Backpropagation of Unrolled Solvers with Folded Optimization. *arXiv preprint arXiv:2301.12047* (2023).

[25] James Kotary, Ferdinando Fioretto, Pascal Van Hentenryck, and Bryan Wilder. 2021. End-to-End Constrained Optimization Learning: A Survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*. 4475–4482. https://doi.org/10.24963/ijcai.2021/610

[26] S.R. Kramer, T.J. Rodenbaugh, and M.W. Conroy. 1994. The use of trenchless technologies for transmission and distribution projects. In *Proceedings of IEEE/PES Transmission and Distribution Conference*. 302–308. https://doi.org/10.1109/TDC.1994.328395

[27] Martin Macaš, Sergio Orlando, Stefan Costea, Petr Novák, Oleksiy Chumak, Petr Kadera, and Petr Kopejtko. 2020. Impact of forecasting errors on microgrid optimal power management. In *2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)*. 1–6. https://doi.org/10.1109/EEEIC/ICPSEurope49358.2020.9160513

[28] Jayanta Mandi, James Kotary, Senne Berden, Maxime Mulamba, Victor Bucarey, Tias Guns, and Ferdinando Fioretto. 2024. Decision-Focused Learning: Foundations, State of the Art, Benchmark and Future Opportunities. *Journal of Artificial Intelligence Research* 80 (Aug. 2024), 1623–1701. https://doi.org/10.1613/jair.1.15320

[29] Jayanta Mandi, Peter J Stuckey, Tias Guns, et al. 2020. Smart predict-and-optimize for hard combinatorial optimization problems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 1603–1610.

[30] Mohamed Massaoudi, Maymouna Ez Eddin, Ali Ghrayeb, Haitham Abu-Rub, and Shady S. Refaat. 2025. Advancing Coherent Power Grid Partitioning: A Review Embracing Machine and Deep Learning. *IEEE Open Access Journal of Power and Energy* 12 (2025), 59–75. https://doi.org/10.1109/OAJPE.2025.3535709

[31] John W. Muhs, Masood Parvania, and Mohammad Shahidehpour. 2020. Wildfire Risk Mitigation: A Paradigm Shift in Power Systems Planning and Operation. *IEEE Open Access Journal of Power and Energy* 7 (2020), 366–375. https://doi.org/10.1109/OAJPE.2020.3030023

[32] National Oceanic and Atmospheric Administration. 2024. High-Resolution Rapid Refresh (HRRR) Model. https://rapidrefresh.noaa.gov/hrrr/. Accessed: 2024-11-11.

[33] Mathaios Panteli, Dimitris N. Trakas, Pierluigi Mancarella, and Nikos D. Hatziargyriou. 2017. Power Systems Resilience Assessment: Hardening and Smart Operational Enhancement Strategies. *Proc. IEEE* 105, 7 (2017), 1202–1213. https://doi.org/10.1109/JPROC.2017.2691357

[34] Cheng Qian and Aiyuan Wang. 2021. Power Grid Disturbance Prediction and Analysis Method Based on SIR Model. In *2021 IEEE 4th Student Conference on Electric Machines and Systems (SCEMS)*. 1–5. https://doi.org/10.1109/SCEMS52239.2021.9646114

[35] Hu Qin, Anton Moriakin, Gangyan Xu, and Jiliu Li. 2024. The generator distribution problem for base stations during emergency power outage: A branch-and-price-and-cut approach. *European Journal of Operational Research* 318, 3 (2024), 752–767. https://doi.org/10.1016/j.ejor.2024.06.007

[36] Jing Qiu, Luke J. Reedman, Zhao Yang Dong, Ke Meng, Huiqiao Tian, and Junhua Zhao. 2017. Network reinforcement for grid resiliency under extreme events. In *2017 IEEE Power & Energy Society General Meeting*. 1–5. https://doi.org/10.1109/PESGM.2017.8273985

[37] Vetrivel S. Rajkumar, Alexandru Ştefanov, José Luis Rueda Torres, and Peter Palensky. 2024. Dynamical Analysis of Power System Cascading Failures Caused by Cyber Attacks. *IEEE Transactions on Industrial Informatics* 20, 6 (2024), 8807–8817. https://doi.org/10.1109/TII.2024.3372024

[38] Anubhav Ratha, Emil Iggland, and Göran Andersson. 2013. Value of Lost Load: How much is supply security worth?. In *2013 IEEE Power & Energy Society General Meeting*. 1–5. https://doi.org/10.1109/PESMG.2013.6672826

[39] Reuters. 2024. Over 1.3 million Florida customers without power due to Hurricane Milton. *Reuters* (10 October 2024). https://www.reuters.com/business/energy/over-13-million-florida-customers-without-power-due-hurricane-milton-2024-10-10/ Accessed: 2025-02-15.

[40] Reuters. 2025. Californian utility SoCal Edison shuts power to over 114,000 customers due to wildfire. *Reuters* (8 January 2025). https://www.reuters.com/business/energy/californian-utility-socal-edison-shuts-power-over-114000-customers-due-wildfire-2025-01-08/ Accessed: 2025-02-15.

[41] Daniel Shea. 2018. *Hardening the Grid: How States Are Working to Establish a Resilient and Reliable Electric System*. Available online: urlhttps://www.ncsl.org/research/energy/hardening-the-grid-how-states-are-working-to-establish-a-resilient-and-reliable-electric-system.aspx [Accessed: February 14, 2025].

[42] U.S. Census Bureau. 2017. American Community Survey 1-Year Estimates, Data Profiles: Massachusetts. https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2017/ Accessed: 2025-02-23.

[43] Vincenzo Di Vito, Mostafa Mohammadian, Kyri Baker, and Ferdinando Fioretto. 2025. Learning To Solve Differential Equation Constrained Optimization Problems. In *International Conference on Learning Representations*, Vol. 13. https://doi.org/10.48550/arXiv.2410.01786

[44] Marin Vlastelica, Anselm Paulus, Vít Musil, Georg Martius, and Michal Rolínek. 2019. Differentiation of blackbox combinatorial solvers. *arXiv preprint arXiv:1912.02175* (2019).

[45] Prince Zizhuang Wang, Jinhao Liang, Shuyi Chen, Ferdinando Fioretto, and Shixiang Zhu. 2025. Gen-DFL: Decision-Focused Generative Learning for Robust Decision Making. arXiv:2502.05468 [cs.LG] https://arxiv.org/abs/2502.05468

[46] Wikipedia contributors. 2023. March 11–15, 2018 nor'easter. https://en.wikipedia.org/wiki/March_11%E2%80%9315,_2018_nor%27easter [Online; accessed 23-February-2025].

[47] Bryan Wilder, Bistra Dilkina, and Milind Tambe. 2019. Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 1658–1665.

[48] Bryan Wilder, Eric Ewing, Bistra Dilkina, and Milind Tambe. 2019. End to end learning and optimization on graphs. *Advances in Neural Information Processing Systems* 32 (2019).

[49] Le Xie, Xiangtian Zheng, Yannan Sun, Tong Huang, and Tony Bruton. 2023. Massively Digitized Power Grid: Opportunities and Challenges of Use-Inspired AI. *Proc. IEEE* 111, 7 (2023), 762–787. https://doi.org/10.1109/JPROC.2022.3175070

[50] Aoran Zhang, Wenbin Zhou, Liyan Xie, and Shixiang Zhu. 2024. Recurrent Neural Goodness-of-Fit Test for Time Series. *arXiv preprint arXiv:2410.13986*

(2024).
[51] Shixiang Zhu, Liyan Xie, Minghe Zhang, Rui Gao, and Yao Xie. 2022. Distributionally robust weighted k-nearest neighbors. *Advances in Neural Information Processing Systems* 35 (2022), 29088–29100.
[52] Shixiang Zhu, Rui Yao, Yao Xie, Feng Qiu, Yueming Qiu, and Xuan Wu. 2021. Quantifying Grid Resilience Against Extreme Weather Using Large-Scale Customer Power Outage Data. arXiv:2109.09711 [stat.AP]

A ABLATION STUDY

To evaluate the impact of the prediction error weight λ in (6) on balancing prediction accuracy and decision quality, we conducted an ablation study with various λ values on synthetic dataset. Results in Table 3 indicate that a lower λ shifts the model's focus toward decision quality, reducing decision regret by aligning predictions with resilience goals, albeit with a slight trade-off in MSE, compared to Two-stage method. We note that such a design offers better flexibility and interpretability for the GDF-trained decision-making models.

Table 3: Effect of λ on Mobile Generator Deployment Problem.

Metric	λ			Baselines	
	0	1	10	Two-stage	Online
MSE ($\times 10^3$)	9.7	9.5	9.6	9.3	/
Cost	6668.7	6609.5	6611.8	6631.7	8163.7
Regret	153.0	135.7	138.0	155.8	1648.1

B DETAILED DESCRIPTION ON ONLINE ALGORITHMS FOR MOBILE GENERATOR DEPLOYMENT PROBLEM

We include pesudocode for the proposed online baseline allocation methods in Alg. 2.

Algorithm 2 Observe-then-optimize algorithm for mobile generator deployment problem

Input: Observations or forecasts $Y_{t,i}$, parameters $(\tau,\ N_g,\ \gamma)$, warehouse stock s_w, etc.
Output: Shipping decisions $\{x^{\text{to}}_{t,i}, x^{\text{back}}_{t,i}\}$ for each time t and city i.

for $t = 1$ to T **do**
  — *update warehouse and city stocks from previous shipments* —
  for each city i **do**
    Compute demand shortfall $d_{t,i} \leftarrow \max(0,\ \lceil Y_{t,i}/N_H \rceil - q_{t,i})$.
    $x^{\text{to}}_{t,i} \leftarrow \min(s_w(t),\ d_{t,i})$ *// send enough to cover shortfall*
    $x^{\text{back}}_{t,i} \leftarrow 0$ *// no return shipments*
  end for
end for

C POWER LINE UNDERGROUNDING

Power line undergrounding is a grid-hardening measure against extreme meteorological events (such as hurricanes and heavy snowfalls) [3, 15, 33, 41]. Although effective, it involves substantial costs for the authority and causes significant disruptions to local communities [26, 31]. The objective of the power line undergrounding problem is to select an optimal subset of locations for underground interventions under budget constraints [3, 33, 41], in anticipation of an incoming hazard.

To formalize the decision-making problem, let x_k be a binary variable indicating whether city k is selected for undergrounding. With K cities in total, the decision vector is $\boldsymbol{x} = [x_1, \ldots, x_K]^\top$, and $\hat{\mathrm{S}}$ denotes the predicted outage states. We then define:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & g(\boldsymbol{x}, \hat{\mathrm{S}}), \\ \text{s.t.} \quad & \sum_{k=1}^{K} x_k \ \le\ C, \\ & x_k \ \in\ \{0,1\}, \quad k = 1, \ldots, K, \end{aligned} \tag{15}$$

where $g(\boldsymbol{x}, \hat{\mathrm{S}})$ is the decision loss that quantifies the impact of outages given the chosen undergrounding plan $\boldsymbol{x}$.

We adopt the System Average Interruption Duration Index (SAIDI) [2] to measure how outages affect the population. Let $Y_k(t)$ be the (true) number of outages at city k and time t, and N_k be the total number of customers in city k. Since undergrounding is assumed fully effective, a city k with $x_k = 1$ incurs no further outages from the event. Hence, the decision loss is:

$$g(\boldsymbol{x}, \mathrm{S}) \ =\ \frac{1}{K} \sum_{k=1}^{K} \frac{1}{N_k} \int_0^\infty \Big[\big(1 - x_k\big)\, Y_k(t) \Big]\, dt. \tag{16}$$

The optimal solution $\boldsymbol{x}^*$ to (15) is then the subset of cities to be undergrounded in order to minimize the total outage impact. Its performance is evaluated via $g(\boldsymbol{x}^*, \mathrm{S})$ using true outage data S.

D IMPLEMENTATION DETAILS OF GDF

To get the gradient of (7), let matrix H encode coefficients for all the linear constraints $H\boldsymbol{x} \le \boldsymbol{a}$, and ξ^i represents a reformulation of the ground truth cost factors derived from S^i, such that $g(\boldsymbol{x}, \mathrm{S}^i) = {\xi^i}^T \boldsymbol{x} + \boldsymbol{b}$. Using the KKT conditions of the Lagrangian of the problem, the gradient of the QP in (7) is:

$$\nabla_\theta L_{\text{QP}} = {\xi^i}^T K^{-1} \nabla_\theta \hat{\mathrm{S}}^i, \quad K = \begin{bmatrix} 2\rho_x I & H^T \\ H & 0 \end{bmatrix}, \tag{17}$$

And $\nabla_{\beta,\gamma} \hat{\mathrm{S}}^i$ can obtained via backpropagation through the neural ODE model parameters $\{\theta_U, \theta_R\}$ [10]. This gradient aims to improve decision quality across all cities and all events $i \in \{1, \ldots, I\}$. Therefore, we refer to it as *decision-focused gradient*.

To be more specific about derivation of (17), the optimal solution $\boldsymbol{x}^*$ must satisfy the KKT conditions. We define the Lagrangian:

$$\mathcal{L}(\boldsymbol{x}, \lambda) = g(\boldsymbol{x}, \mathrm{S}) + \rho \|\boldsymbol{x}\|_2^2 + \lambda^T (H\boldsymbol{x} - \boldsymbol{a}). \tag{18}$$

The stationarity of x^* for optimality gives:

$$\nabla_x g(\boldsymbol{x}, \mathrm{S}) + 2\rho\boldsymbol{x} + H^T \lambda = 0. \tag{19}$$

Since the optimal decision $\boldsymbol{x}^*$ satisfies the above KKT system, we can apply implicit differentiation. Taking the total derivative with respect to the predicted system state $\hat{\mathbf{S}}$:

$$\begin{bmatrix} \nabla^2_{xx}\mathcal{L} & H^T \\ H & 0 \end{bmatrix} \begin{bmatrix} \frac{d\boldsymbol{x}^*}{d\hat{\mathbf{S}}} \\ \frac{d\lambda}{d\hat{\mathbf{S}}} \end{bmatrix} = \begin{bmatrix} \frac{d(-\nabla_x g(\boldsymbol{x},\mathbf{S}))}{d\hat{\mathbf{S}}} \\ 0 \end{bmatrix}. \tag{20}$$

where

$$\frac{d\boldsymbol{x}^*}{d\hat{\mathbf{S}}} = -K^{-1}\frac{d\nabla_x g}{d\hat{\mathbf{S}}}, \tag{21}$$

and the KKT matrix K is defined as:

$$K = \begin{bmatrix} 2\rho I & H^T \\ H & 0 \end{bmatrix}. \tag{22}$$

since $\nabla^2_{xx}\mathcal{L} = 2\rho I$ etc.

This allows us to compute the gradient of the loss function with respect to model parameters:

$$\nabla_\theta L_{\text{QP}} = \xi^{i\,T} K^{-1} \nabla_\theta \hat{\mathbf{S}}^i. \tag{23}$$

In practice, we regularize the GDF loss with prediction-focused gradient, as the prediction error is localized and requires fine-grained information for each individual sample unit. Specifically, we construct mini-batches $\mathcal{B} \subset \mathcal{D}$ from the training dataset $\mathcal{D} = \{z_k^i, y_k^i(t)\}$.

The neural ODE model generates forecasts $\hat{\mathbf{S}}_\mathcal{B} = f_{\beta,\gamma}(z_\mathcal{B})$ for each sample in the batch. From these forecasts, we extract the predicted values $\hat{Y}_\mathcal{B}$, which are then used to compute the MSE loss:

$$L_{\text{MSE}} = \sum_{\mathcal{B}\subset\mathcal{D}} \frac{1}{|\mathcal{B}|} \sum_{(i,k,t)\in\mathcal{B}} \left(y_k^i(t) - \hat{Y}_k^i(t)\right)^2. \tag{24}$$

Finally, for each epoch, the model parameters θ is updated using a combination of both loss gradients, balanced by a hyperparameter λ:

$$\nabla_\theta L_{\text{total}} = \nabla_\theta L_{\text{GDF}} + \lambda \nabla_\theta L_{\text{MSE}}. \tag{25}$$

E ADDITIONAL RESULTS FOR THE MOBILE GENERATOR DEPLOYMENT PROBLEM

This section presents additional results and visualizations for the mobile generator deployment problem.

As demonstrated in the ablation study, when λ is large, the MSE dominates model training, reducing the advantage of GDF over MSE-trained models in decision quality. For more detailed ablation results in the mobile generator deployment problem, see Fig. 5, which shows that as λ increases, the GDF results become similar to those of the Two-stage method, resulting in larger regret and higher variance.

Further more, Table 4 summarizes the out-of-sample performance for the generator deployment problem on synthetic data across three transportation cost factors (100, 500, and 1000). As the transportation cost increases, the improvement in decision quality for GDF compared to the Two-stage methods becomes more apparent. This highlights the importance of scheduling and proactive actions when transportation costs are high, demonstrating the clear advantage of GDF.

We also provide additional visualizations of the deployment schemes under varying conditions. Comparing Fig. 6 and Fig. 8, we observe that when travel costs are low, the online strategy closely

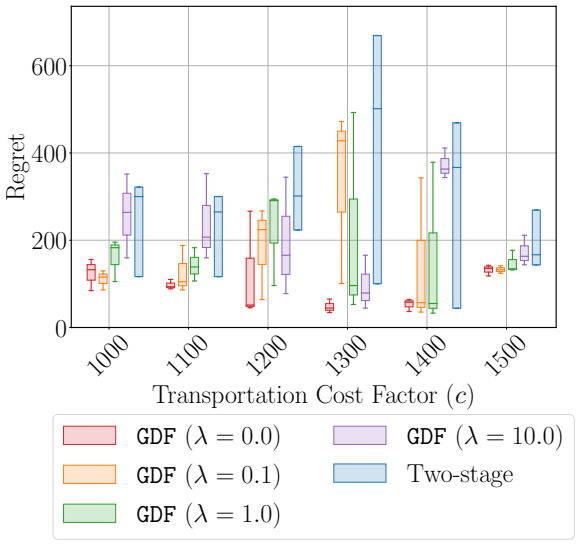

Figure 5: Regret performance of the GDF method with varying λ values across different transportation cost factors in synthetic data for mobile generator deployment problem, benchmarked against the Two-stage method.

approximates the optimal strategy, resulting in small regret. This is because low travel costs allow the online method to frequently move generators based on previous day data at minimal expense, leading to near-optimal regret, whereas the Two-stage and GDF methods rely more on predictions, and the associated noise can diminish the benefits of prediction or proactive allocation under these conditions.

In contrast, comparing Fig.7 and Fig.8, we find that limited resources degrade the online strategy's performance. Without proactive planning, fewer generators must be relocated more frequently with an online method, incurring higher transportation costs and overall regret compared to GDF and Two-stage approaches.

F ADDITIONAL RESULTS FOR POWER LINE UNDERGROUNDING

Fig. 9 shows the predicted outage trajectories for all Massachusetts counties using the GDF model compared with groundtruth and Two-stage. Notably, the model exaggerates outages for selected counties—a deliberate strategy to prioritize resource allocation. This controlled overestimation, while slightly increasing MSE. Overall, it effectively reduces SAIDI and regret compared to the Two-stage baseline, demonstrating that decision-focused training can enhance overall scheduling performance.

G EXTENDED LITERATURE REVIEW FOR DFL AND DIFFERENTIABLE OPTIMIZATION

Decision-Focused Learning. Decision-focused learning (DFL) has emerged as a powerful framework for integrating predictive models with downstream optimization tasks. Unlike traditional two-stage approaches, which first train standalone prediction models and then use their predictions as input parameters to optimal decision

Table 4: Out-of-Sample Performance for Generator Distribution problem with Synthetic Data. Results are averaged over 3 repeated experiments with standard error (SE) in the brackets.

Model	Transportation Cost = 100			Transportation Cost = 500			Transportation Cost = 1000		
	MSE	Cost	Regret	MSE	Cost	Regret	MSE	Cost	Regret
Ground Truth	/	6515.6 (74.4)	0	/	11271	0	/	16271	0
Online (lag = 1)	/	8163.7 (293.4)	1648.1 (358.6)	/	19630.4 (1514.8)	8358.7 (1576.8)	/	33963.7 (3042.2)	17692.1 (3104.4)
Online (lag = 3)	/	8978.6 (277.6)	2462.9 (313.8)	/	22045.2 (1493.3)	10773.6 (1522.9)	/	38378.6 (3020.2)	22106.9 (3050.4)
Online (lag = 5)	/	9618.9 (57.3)	4939.9 (111.1)	/	22152.2 (498.5)	16497.9 (557.8)	/	37818.9 (1074.7)	32164.5 (1135.0)
Two-stage	**9336.0 (975.6)**	6671.5 (4328.7)	155.8 (33.3)	9336.0 (975.6)	11465.8 (94.6)	194.2 (107.3)	9336.0 (975.6)	16517.8 (133.9)	246.1 (112.7)
GDF	9709.6 (3055.6)	6668.7 (4365.9)	**153.0 (19.8)**	**8671.2 (1553.7)**	11428.1 (87.1)	**156.4 (44.2)**	**6981.1 (1874.0)**	16395.9 (89.0)	**124.2 (36.3)**

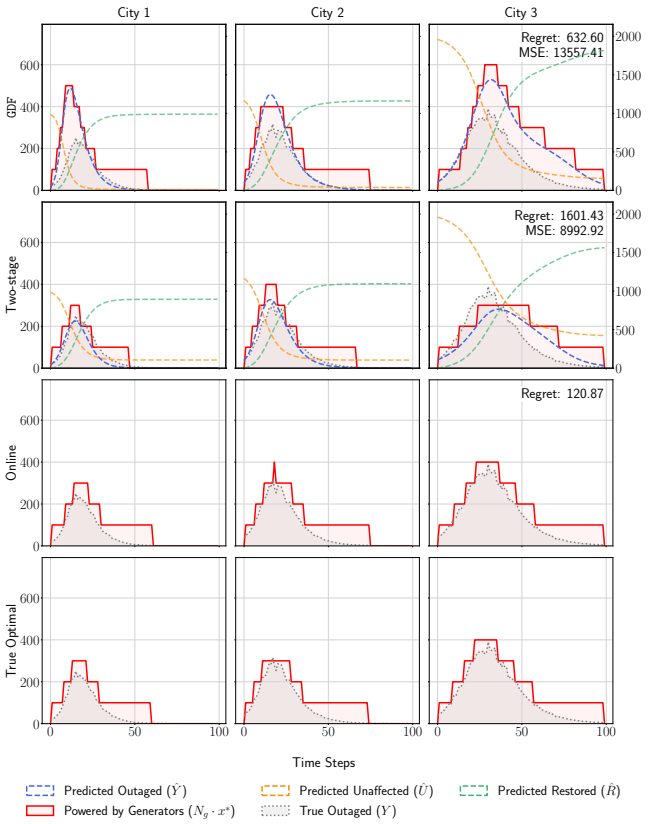

Figure 6: A synthetic instance of the mobile generator deployment problem for a system with three cities and five generators ($Q_w = 10$). The y-axis shows the number of households experiencing outages over time. In this example, the transportation cost is set to $c = 10$, the customer interruption cost to $\tau = 1$, and the operational cost to $\gamma = 2$. The online method operates with a one-day observation lag. Travel time $\delta_t = 0$ is neglected in this case.

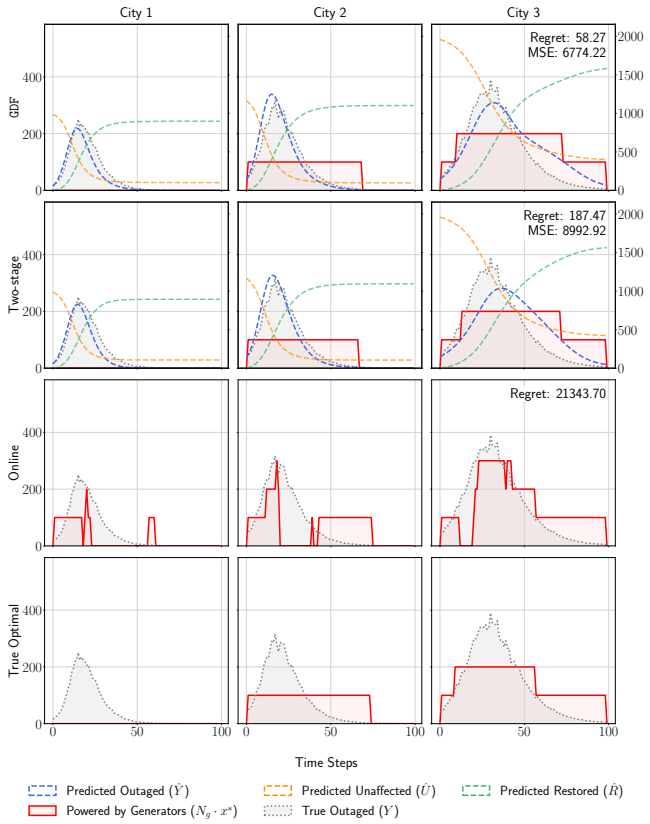

Figure 7: A synthetic instance of the mobile generator deployment problem for a system with three cities and five generators ($Q_w = 3$). The y-axis shows the number of households experiencing outages over time. In this example, the transportation cost is set to $c = 1000$, the customer interruption cost to $\tau = 1$, and the operational cost to $\gamma = 2$. The online method operates with a one-day observation lag. Travel time $\delta_t = 0$ is neglected in this case.

models, DFL aligns the prediction model's training loss with the objective function of the downstream optimization. This concept is enabled in gradient descent training by backpropagating gradients through the solution to an optimization problem. When the optimization is a differentiable function of its parameters, this can be implemented via implicit differentiation of optimality conditions such as KKT conditions [6, 18] or fixed-point conditions [24, 48]. When the optimization is nondifferentiable, it can instead be implemented by means of various approximation techniques [25, 28]. Unlike traditional two-stage approaches, which first train standalone prediction models and then use their predictions as inputs for decision-making, DFL directly embeds the optimization problem within the learning process. This allows the learning model to focus on the variables that matter most for the final decision [28].

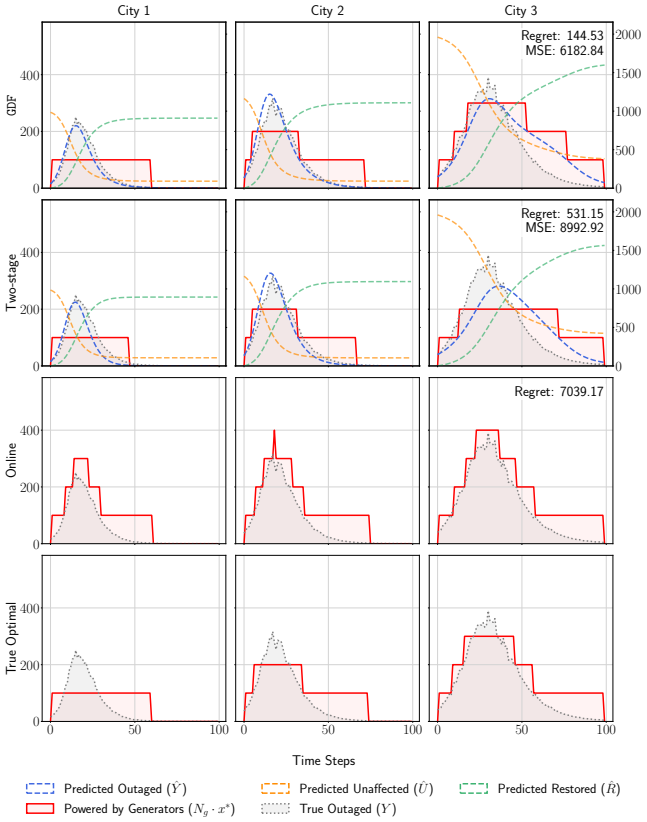

Figure 8: A synthetic instance of the mobile generator deployment problem for a system with three cities and five generators ($Q_w = 10$). The y-axis shows the number of households experiencing outages over time. In this example, the transportation cost is set to $c = 1000$, the customer interruption cost to $\tau = 1$, and the operational cost to $\gamma = 2$. The online method operates with a one-day observation lag. Travel time $\delta_t = 0$ is neglected in this case.

Elmachtoub and Grigas [13] first proposed the *Smart Predict-and-Optimize* (SPO) framework, which introduced a novel method for formulating optimization problems in the prediction process. SPO essentially bridges the gap between predictive modeling and optimization by constructing a decision-driven loss function that reflects the downstream task. However, the SPO framework only addresses linear optimization problems and does not extend well to more complex combinatorial tasks.

The most-studied class of nondifferentiable optimization problems in decision-focused learning (DFL) involves linear programs (LPs). Notably, the Smart Predict-and-Optimize (SPO) framework by Elmachtoub and Grigas [13] introduced a convex surrogate upper bound to approximate subgradients for minimizing the suboptimality of LP solutions based on predicted cost coefficients. The most-studied class of nondifferentiable optimizations are linear programs (LPs). Elmachtoub and Grigas [13] proposed the Smart Predict-and-Optimize (SPO) framework for minimizing the suboptimality of solutions to a linear program as a function of its predicted cost coefficients. Despite this function being inherently

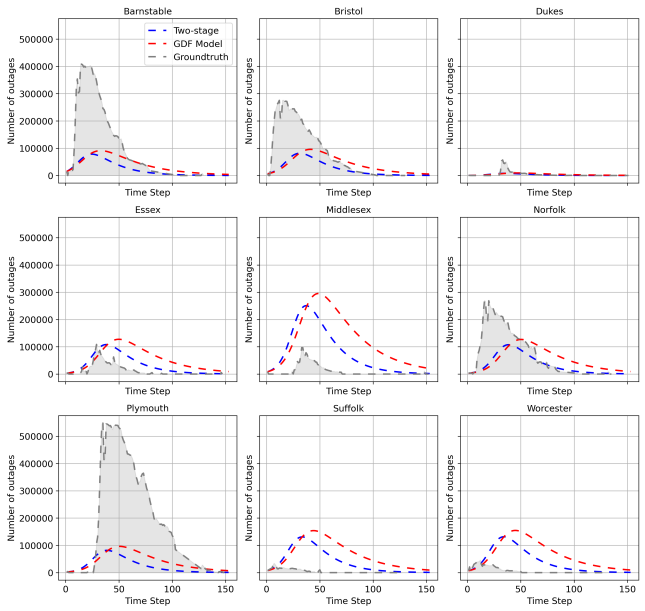

Figure 9: GDF is overestimating outage in certain counties to prioritize resource allocation at the cost of MSE.

non-differentiable, a convex surrogate upper-bound is used to derive informative subgradients. Wilder et al. [47] propose to smooth linear programs by augmenting their objectives with small quadratic terms [6] and differentiating the resulting KKT conditions. A method of smoothing LP's by noise perturbations was proposed in [8]. Differentiation through combinatorial problems, such as mixed-integer programs (MIPs), is generally performed by adapting the approaches proposed for LP's, either directly or on their LP relaxations. For example, Mandi et al. [29] demonstrated the effectiveness of the SPO method in predicting cost coefficients to MIPs. Vlastelica et al. [44] demonstrated their method directly on MIPs, and Wilder et al. [47] evaluated their approach on LP relaxations of MIPs.

Building on these works, we extend DFL to spatio-temporal decision-making for power grid resilience management. Our approach employs quadratic relaxations to enable gradient backpropagation through MIPs [47], thereby integrating a spatio-temporal ODE model for power outage forecasting directly into the optimization process. Additionally, we introduce a Global Decision-Focused Framework that combine prediciton error with decision losses across geophysical units, improving grid resilience against extreme natural events and bridging the gap between localized predictions and system-wide decisions.

Differentiable Optimization. Differentiable optimization (DO) techniques have demonstrated significant potential in integrating predictive models with optimization problems. By enabling the computation of gradients through optimization processes, DO facilitates the seamless incorporation of complex system objectives into machine learning models, thereby enhancing decision-making capabilities [4]. Recent extensions of DO methods have tackled challenges beyond standard optimization tasks. For example, distributionally

robust optimization (DRO) problems have been addressed using differentiable frameworks to handle prediction tasks under worst-case scenarios. For instance, [11, 51] employed DO-based techniques to improve uncertainty quantification and robust learning, effectively addressing data scarcity and enhancing resilience modeling.

Beyond predictive modeling, DO has advanced solutions in combinatorial and nonlinear optimization. Techniques such as implicit differentiation of KKT conditions [6] and fixed-point conditions [24] address differentiable constraints, while approximation methods, including noise perturbation [8] and smoothing techniques [44], enable gradient computation for nondifferentiable tasks.

These advancements underscore DO's pivotal role in bridging predictive modeling and optimization, especially where decision quality critically affects system resilience. In this work, DO is employed to align spatio-temporal outage predictions with grid optimization objectives, enabling robust strategies for generator deployment and power line undergrounding.