

# Data-Driven Stochastic Distribution System Hardening Based on Bayesian Online Learning

Wenlong Shi<sup>1</sup>, Member, IEEE, Hongyi Li<sup>1</sup>, Member, IEEE, and Zhaoyu Wang<sup>1</sup>, Senior Member, IEEE

**Abstract**—Extreme weather frequently cause widespread outages in distribution systems (DSs), demonstrating the importance of hardening strategies for resilience enhancement. However, the well-utilization of real-world outage data with associated weather conditions to make informed hardening decisions in DSs is still an open issue. To bridge this research gap, this paper proposes a data-driven stochastic distribution line (DL) hardening strategy. First, a deep neural network (DNN) regression model is developed to predict the probabilistic evolution of outage scenarios under various hardening decisions. Based on the DNN predictions, the problem is formulated as a decision-dependent distributionally robust optimization (DRO) model, accounting for uncertainties in outage scenario distributions using a data-driven ambiguity set. To address decision-dependent uncertainty, a Bayesian online learning algorithm is proposed. This algorithm decomposes the original problem into inner and outer problems. Then, it iteratively refines hardening decisions by sequentially incorporating outage data and dynamically updating decision-specific ambiguity sets by using Bayes' theorem and Bayesian Inference. Also, the convergence of the algorithm is proven through dynamic regret analysis. Finally, case studies are implemented on a real-world DS in Redfield, Iowa, USA. A dataset spanning 24 years (2001–2024) is constructed based on the utility outage records. The simulation results validates the effectiveness of the proposed strategy.

**Index Terms**—Distribution line, distribution system, extreme weather, hardening, outage data, online learning, resilience.

## NOMENCLATURE

### A. Parameters

|                      |   |
|----------------------|---|
| $\mathbf{1}$         | Column vector with three all 1 entries.   |
| $B^{\max}$           | Maximum available hardening budget.   |
| $C_z^{pl}, C_z^{ud}$ | Hardening costs of line segment $z$ in terms of pole upgrading and line undergrounding. |
| $C_z^{pd}$           | Hardening cost of pad-mounted transformer.  |
| $D_i^p, D_i^q$       | Maximum Active and reactive load demand.  |

|                                      |  |    |
|--------------------------------------|--|----|
| $\tilde{I}_{ij}^{\max}$              | Maximum current capacity of line $(i, j)$ .                            | 55 |
| $M$                                  | A sufficiently large number fo big- $M$ .                              | 56 |
| $P_k^{\max}, Q_k^{\max}$             | Rated active and reactive capacity of source $k$ .                     | 57 |
| $r_{ij}, x_{ij}$                     | Resistance and reactance matrices of line $(i, j)$ .                   | 58 |
| $\tilde{V}^{\min}, \tilde{V}^{\max}$ | Maximum and minimum nodal voltage limits.                              | 59 |
| $\omega_i$                           | Load weight at node $i$ for priority consideration.                    | 60 |
| $\alpha_z$                           | Binary indicator, 0 if faults occur on line segment $z$ , otherwise 1. | 61 |
| $\Delta t_s$                         | Outage duration under scenario $s$ .                                   | 62 |

### B. Variables

|                                      |   |    |
|--------------------------------------|---|----|
| $F_{ij}$                             | Fictitious commodity flow on line $(i, j)$ .  | 63 |
| $h_z^{ud}, h_z^{pl}$                 | Binary variables, 1 if undergrounding lines and upgrading poles are chosen for line segment $z$ . | 64 |
| $\tilde{I}_{ij}^k$                   | 3-phase squared current magnitude on line $(i, j)$ with respect to source $k$ .                   | 65 |
| $p_i, q_i$                           | Active and reactive load demand at node $i$ .   | 66 |
| $\tilde{p}_{ij}^k, \tilde{q}_{ij}^k$ | Slack variables for active and reactive power balance when switch $(i, j)$ is open.               | 67 |
| $P_{ij}^k, Q_{ij}^k$                 | 3-phase active and reactive power flow on line $(i, j)$ with respect to source $k$ .              | 68 |
| $u_i^k$                              | Binary variable, equals 1 if node $i$ is activated by source node $k$ , otherwise 0.              | 69 |
| $v_j^k$                              | Slack variable for voltage at node $j$ when switch $(i, j)$ is disconnected.                      | 70 |
| $\tilde{V}_i^k$                      | 3-phase squares of voltage magnitude of node $i$ with respect to node $k$ .                       | 71 |
| $\beta_i$                            | Binary vairable, equals 0 if the load at node $i$ is not served, otherwise 1.                     | 72 |
| $\gamma_{ij}$                        | Binary variable, 0 if switch $(i, j)$ is open.  | 73 |
| $\Gamma_{ij}$                        | Binary, 0 if line $(i, j)$ in the fictitious graph is disconnected to maintain radiality.         | 74 |
| $\sigma_{i,s}$                       | Load demand not being served at node $i$ .  | 75 |

### C. Indices and Sets

|          |   |    |
|----------|---|----|
| $i, j$   | Indices for nodes.  | 76 |
| $(i, j)$ | Indices for lines, switches, fuses, reclosers between nodes $i$ and $j$ . | 77 |
| $m$      | Index of training dataset.  | 78 |

AQ:1 Received 19 March 2025; revised 15 August 2025; accepted 27 October 2025. This work was supported in part by the Power System Engineering and Research Center under Grant PSERC S-110, in part by U.S. Department of Energy's Office of Electricity under Award DE-OE000986, in part by U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy under Grant DE-EE0011234, and in part by the National Science Foundation under Grant ECCS 2042314. Paper no. TSG-00489-2025. (Corresponding author: Zhaoyu Wang.)

The authors are with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: wshi5@iastate.edu; hongyili@iastate.edu; wzy@iastate.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TSG.2025.3628228>.

Digital Object Identifier 10.1109/TSG.2025.3628228

|     |                                |   |
|-----|--------------------------------|---|
| 100 | $k \in \mathcal{K}$            | Index of source nodes in the network.   |
| 101 | $s \in \mathcal{S}$            | Index of outage scenarios.  |
| 102 | $t$                            | Iteration index of the online algorithm.  |
| 103 | $z \in \mathcal{Z}$            | Index of components for hardening.  |
| 104 | $\mathcal{B}, \mathcal{B}_z$   | Sets of switches and boundaries of segment $z$ .  |
| 105 |                                |   |
| 106 | $\mathcal{C}_i$                | Set of child nodes of node $i$ .  |
| 107 | $\mathcal{U}, \mathcal{P}$     | Sets of geographic restrictions on undergrounding and pole upgrading.                     |
| 108 |                                |   |
| 109 | $\mathcal{Z}_s, \mathcal{Z}_t$ | Sets of components considering line segments and distribution transformers, respectively. |
| 110 |                                |   |
| 111 |                                |   |
| 112 | $\phi$                         | Index of phase in three-phase systems.  |

## I. INTRODUCTION

114 **E**XTREME weather events, such as hurricanes and severe  
 115 storms, impose significant challenges to the resilience of  
 116 power distribution systems (DSs). These natural disasters often  
 117 result in widespread outages, disrupting electricity supply for  
 118 extended durations [1]. For example, Hurricane Ian in 2022  
 119 caused 5-hours electricity interruption, which affected over 2.7  
 120 million customers in Florida, USA [2]. And, the 2021 Texas  
 121 winter storm resulted in 52,000 MW of power capacity being  
 122 offline [3]. To mitigate these negative impacts, DS hardening  
 123 measures such as undergrounding lines and upgrading poles  
 124 are necessary. These measures are very commonly adopted by  
 125 utilities, which are proven effective in resilience enhancement.

126 However, obtaining which poles to upgrade and which line  
 127 segments to underground is challenging. Hardening strategies  
 128 relying on empirical analysis obtain hardening decisions based  
 129 on iterative simulation and assessment [4], [5]. Despite the  
 130 approaches are intuitive, their optimality cannot be guar-  
 131 anteed. In other words, theoretical analysis is necessary to  
 132 deal with outage complexities. In this respect, optimization  
 133 provides an effective way. In particular, robust optimization  
 134 (RO), stochastic programming (SP), and distributionally robust  
 135 optimization (DRO) are useful in tackling uncertainties. For  
 136 example, RO employs uncertainty sets to represent uncertain  
 137 parameters. It is developed in a min-max form to obtain  
 138 hardening decisions against the worst-case scenario [6], [7],  
 139 [8], [9], [10], [11], [12]. As RO contains only deterministic  
 140 variables, their performance can degrade when addressing the  
 141 stochastic nature of extreme events. In addition, SP considers  
 142 stochasticity via scenario sets with probability distributions.  
 143 It optimizes the expected cost value throughout all outage  
 144 scenarios [13], [14], [15], [16]. DRO constructs an ambiguity  
 145 set of probability distributions to address the randomness. It  
 146 considers the worst-case expected hardening performance [17].  
 147 However, both SP and DRO requires probability distributions,  
 148 which should be accurate to avoid under- or over-estimation.

149 Modern power systems have transformed into more intelli-  
 150 gent networks, driven by advanced sensing and communication  
 151 technologies [18]. These advancements facilitate the collection  
 152 of network conditions, making data-driven hardening strategies  
 153 possible. For example, in [19], [20], and [21], the statistics are  
 154 extracted from outage records to establish data-driven ambi-  
 155 guity sets. Accordingly, a modified DRO is developed such  
 156 that hardening decisions can be obtained without relying on

explicit models. Nonetheless, these research are predominantly  
 investigated for transmission systems. They do not adapt to the  
 unique features of outages and restoration at the distribution  
 level.

In practice, DS operators monitor the network condition rou-  
 tinely via Supervisory Control and Data Acquisition (SCADA)  
 systems. The outage related information becomes available at  
 the same time. For example, for isolating downstream faults,  
 the clearing devices, such as fuses and reclosers installed along  
 main feeders or at distribution transformers, are recorded. The  
 event time stamp, restoration times, and affected customers are  
 also documented [22], [23]. This dataset have the potential to  
 support hardening decision-making. However, how to achieve  
 this remains underexplored.

According to the aforementioned issue, this paper proposes  
 a data-driven stochastic DL hardening strategy. Specifically,  
 we employ DRO formulation to address the stochasticity, and  
 we develop an online learning process to solve the data-driven  
 challenge. The main contributions are as follows:

- 1) A unique feature of real-world outage data is that it  
 is collected under fixed network conditions. To address  
 this challenge, a deep neural network (DNN) regres-  
 sion model is proposed in this paper. By learning the  
 nonlinear relationships between hardening measures and  
 their impacts on outage outcomes, the model effectively  
 captures the complex evolution of outage scenarios.
- 2) For utilizing outage data from SCADA systems and  
 weather information from stations, and enable stochastic  
 analysis, a data-driven stochastic DL hardening problem  
 is proposed. The problem is formulated as a decision-  
 dependent DRO model with a data-driven ambiguity set.  
 Note that the problem can not be solved using traditional  
 algorithms due to the decision-dependent uncertainty.
- 3) To address the decision-dependent issue and solve the  
 problem efficiently, a Bayesian online learning algo-  
 rithm is developed. This algorithm first decomposes  
 the original problem into inner and outer problems.  
 Then, it refines hardening decisions in an iterative man-  
 ner by sequentially learning from historical data and  
 dynamically updating the ambiguity set based on Bayes'  
 Theorem and Bayesian Inference. The convergence is  
 also proven based on dynamic regret analysis.

The remainder of this paper is structured as follows. Sec-  
 tion II presents the system model, including the network  
 model and the DNN regression model for outage prediction.  
 Section III formulates the data-driven stochastic DL hardening  
 problem as a decision-dependent DRO model. In Section IV,  
 a Bayesian online learning algorithm is developed. Section V  
 evaluates the proposed strategy through numerical experiments  
 on a real-world distribution network in Redfield, Iowa, USA.  
 Finally, Section VI concludes the paper and outlines future  
 directions.

## II. SYSTEM MODEL

In this section, the distribution network model considering  
 real-world outage data is presented. A DNN regression model  
 is proposed to predict the probability distribution of outage  
 scenarios under varying hardening decisions.

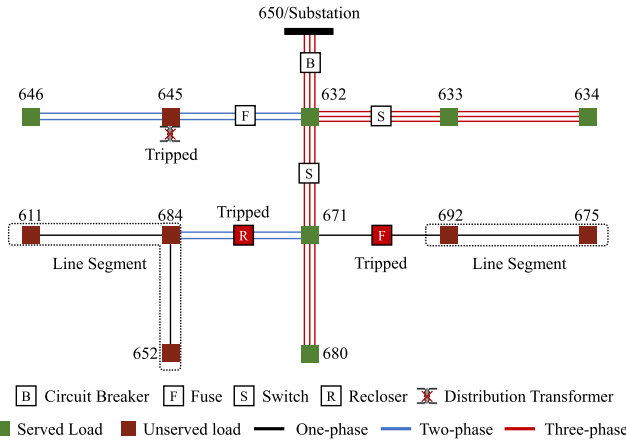


Fig. 1. An illustration of the three-phase unbalanced DS model considering real-world outage data and protection devices.

### A. Distribution Network Model

In this paper, the distribution network model is developed to facilitate the utilization of real-world outage data. Specifically, we consider three-phase unbalanced systems. The distribution network is modeled as a graph  $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ , where  $\mathcal{N}$  represents the set of nodes and  $\mathcal{L}$  represents the set of lines. A line between two nodes can represent a DL, a switch or a protection device. In addition, the protection devices are assumed to install along main feeders, including circuit breakers, reclosers, and fuses to isolate faults on main feeders. We also have fuses installed at distribution transformers to clear faults on laterals. Then, the portion of lines between two or more protection devices, is defined as a line segment. To enhance system resilience against extreme weather events, hardening measures including upgrading poles and undergrounding lines are applied to line segments. The deployment of pad-mounted transformers is employed to protect distribution transformers. Beyond protection devices, switches are also installed along main feeders for topology reconfiguration [24].

An example based on IEEE 13-node test feeder is shown in Fig. 1. Specifically, a recloser and a fuse are tripped to isolate the faults downstream of node 684 and 692, respectively. Also, the distribution transformer at node 645 is tripped to isolate the fault on laterals. Furthermore, the faults are represented as a set of outage scenarios. The outage scenarios are classified into two types. First, when a fault happens on the main feeder, it can be easily located to one specific line segment between protection devices. However, the exact fault location within this line segment does not influence the automated operation of the protection devices [25]. Consequently, each line segment is corresponding to an outage scenario. Second, when a fault occurs on the lateral, the fuse at the distribution transformer will trip to isolate the fault. This defines another type of outage scenarios that isolate the downstream of the transformer.

### B. DNN Regression Model

For DL hardening problem, a significant challenge lies in the fact that outage data is predominantly collected under existing and fixed network conditions. In practice, it is nearly impossible to obtain outage data corresponding to the multitude

of potential hardening solutions, as implementing and testing these solutions across diverse weather conditions require impractical levels of time, resources, and infrastructure. To overcome this limitation, a DNN regression model is proposed. The objective is to predict how a real-world outage scenario, that is observed under fixed network conditions, will evolve when subjected to hypothetical hardening measures.

We first construct the training dataset. Specifically, considering an outage scenario contains a set of  $\mathcal{Z}'$  fault components including line segments and distribution transformers. We use  $\mathcal{I}_z \in [0, 1]$  to represent the improvement in the failure probability of component  $z$  if hardening measures are taken [26]. The improvement is integrated into the dataset by assigning the scenario with a label indicating an occurrence of  $\prod_{z \in \mathcal{Z}'} (1 - \mathcal{I}_z)$ . The remaining probability  $1 - \prod_{z \in \mathcal{Z}'} (1 - \mathcal{I}_z)$  is evenly distributed across the other scenarios with fault components  $z' \in \mathcal{Z}' \setminus z$  for normalization. Such an adjustment assumes that the effects of hardening different components are independent. It is practical, as physically reinforcing one component does not influence the failure probability of others in most distribution systems. The training dataset can be represented as  $\mathcal{T} = \{(\mathbf{h}_m, \mathbf{c}_m, \mathbf{o}_m, \mathbf{o}'_m)\}_{m=1}^M$ , where  $m$  is the instance index. Specifically,  $\mathbf{h}_m = [h_{m,1}, h_{m,2}, \dots, h_{m,|\mathcal{Z}'|}] \in \{0, 1\}^{|\mathcal{Z}'|}$  represents a binary vector with  $|\mathcal{Z}'|$  components, each corresponding to a hardening decision  $h_{m,z}$ , where  $h_{k,z} = 1$  when component  $z$  is hardened.  $\mathbf{c}_m = [c_{m,1}, c_{m,2}, \dots, c_{m,|\mathcal{M}|}] \in \mathbb{R}^{|\mathcal{M}|}$  contains  $|\mathcal{M}|$  weather parameters, such as wind speed, humidity and temperature. Also,  $\mathbf{o}_m = [o_{m,1}, o_{m,2}, \dots, o_{m,|\mathcal{S}|}] \in \{0, 1\}^{|\mathcal{S}|}$  is a binary vector for outage scenarios, where  $o_{m,s} = 1$  if scenario  $s$  occurs. And,  $\mathbf{o}'_m = [o'_{m,1}, o'_{m,2}, \dots, o'_{m,|\mathcal{S}|}] \in [0, 1]^{|\mathcal{S}|}$  denotes the outage occurrence, serving as labels for learning and making predictions. Furthermore, the DNN model consists of an input layer, hidden layers, and an output layer. The input layer vector is  $\{\mathbf{h}, \mathbf{c}, \mathbf{o}\}$  including  $|\mathcal{Z}'| + |\mathcal{M}| + |\mathcal{S}|$  neurons. The output layer contains  $|\mathcal{S}|$  neurons each corresponding to the probability of an outage scenario. In addition, to ensure the output of the DNN sums up to 1, the softmax function is applied as the activation function of the output layer. Note that the trained DNN serves as a regression model for scenario translation. It can be described as a function  $f_w : (\mathbf{h}, \mathbf{c}, \mathbf{o}) \rightarrow \mathbf{o}'$ , where  $w$  represents the learnable parameters. It transforms data collected under fixed network conditions into predictions to reflect the effects of hypothetical hardening decisions. Note that, in this paper, the DNN is trained on outage records associated with wind-related weather covariates. The network architecture is inherently flexible and not restricted to specific input types. Additional ambient variables, such as temperature, precipitation, ice accretion, or lightning frequency, can be included into the input feature vector. As a result, the DNN can be adapted to other geographic regions where different climatic factors are more prominent, provided that corresponding historical data are available for training.

## III. DATA-DRIVEN STOCHASTIC DL HARDENING PROBLEM FORMULATION

In this paper, we aim to bridge the research gap by leveraging long-term historical outage data to inform and optimize the stochastic DL hardening decision against extreme weather.

312 To this end, the problem is formulated as a decision-dependent  
313 DRO model with a data-driven ambiguity set. In the rest of  
314 this section, the problem constraints are presented, and the  
315 problem objective is discussed in details.

### 316 A. Problem Constraints

317 In this section, the problem constraints are presented in  
318 detail. Note that except for the investment constraints, all con-  
319 straints are scenario-specific. However, for notation simplicity,  
320 outage scenario  $s$  is not explicitly indicated in each equation.

321 1) *Operational Constraints*: To ensure an efficient opera-  
322 tion for restoration, the following constraints are applied:

$$323 \gamma_{ij} \leq \alpha_z, \forall (i, j) \in \mathcal{B}_z, z \in \mathcal{Z}_s, \quad (1)$$

$$324 \beta_i \leq \alpha_z, i = z, \forall z \in \mathcal{Z}_t, \quad (2)$$

$$325 0 \leq \mathbf{p}_i \leq \beta_i \mathbf{D}_i^p, 0 \leq \mathbf{q}_i \leq \beta_i \mathbf{D}_i^q, \forall i \in \mathcal{N}, \quad (3)$$

$$326 \boldsymbol{\sigma}_i = \mathbf{D}_i^p - \mathbf{p}_i, \forall i \in \mathcal{N}, \quad (4)$$

$$327 u_i^k = 1, \forall k \in \mathcal{K}, \quad (5)$$

$$328 \sum_{k \in \mathcal{K}} u_i^k \leq 1, \forall i \in \mathcal{N}, \quad (6)$$

$$329 u_j^k \leq u_i^k, \forall j \in \mathcal{C}_i, \quad (7)$$

$$330 u_i^k + u_j^k \leq 1 + \gamma_{ij}, \forall (i, j) \in \mathcal{B}, k \in \mathcal{K}, \quad (8)$$

$$331 |u_i^k - u_j^k| \leq 1 - \gamma_{ij}, \forall (i, j) \in \mathcal{B}, k \in \mathcal{K}. \quad (9)$$

332 Constraint (1) models fault isolation, implying that if faults  
333 occur on line segment  $z$ , the switches on the boundaries must  
334 be open. Constraint (2) means that the load can not be served if  
335 the distribution transformer is tripped due to faults on laterals.  
336 Constraints (3)-(4) are for unserved loads [27]. Constraints (5)-  
337 (6) imply one node can only be activated by one source node  
338 for voltage and frequency regulation. Constraint (7) means that  
339 a child node can be activated only when its parent node is  
340 activated. Constraints (8)-(9) ensure two ends of a switch  $(i, j)$   
341 are disconnected if the switch is open. In addition, if line  $(i, j)$   
342 is hard connected, meaning it is non-switchable, the activation  
343 status of nodes  $i$  and  $j$  must remain consistent.

344 2) *Power Flow Constraints*: In this paper, the branch flow  
345 model is adopted as constraints to provide a set of closed-form  
346 power flow equations [28]. The three-phase active and reactive  
347 power balance equations can be given by

$$348 \sum_{j \in \mathcal{C}_i} \mathbf{P}_{jj}^k = \mathbf{P}_{ij}^k - \mathbf{p}_j - \mathbf{r}_{ij} \tilde{\mathbf{I}}_{ij}^k + \hat{\mathbf{p}}_{ij}^k, \forall (i, j) \in \mathcal{L}, \quad (10)$$

$$349 \sum_{j \in \mathcal{C}_i} \mathbf{Q}_{jj}^k = \mathbf{Q}_{ij}^k - \mathbf{q}_j - \mathbf{x}_{ij} \tilde{\mathbf{I}}_{ij}^k + \hat{\mathbf{q}}_{ij}^k, \forall (i, j) \in \mathcal{L}, \quad (11)$$

350 where  $\tilde{\mathbf{I}}_{ij}^k$  is the squared magnitude of the current from node  $i$   
351 to  $j$  with respect to the source  $k$ . In addition, the power flow  
352 constraints after angle and conic relaxations can be given by

$$353 \tilde{\mathbf{I}}_{ij}^k \odot \tilde{\mathbf{V}}_i^k \geq (\mathbf{P}_{ij}^k)^{\odot 2} + (\mathbf{Q}_{ij}^k)^{\odot 2}, \forall (i, j) \in \mathcal{L}, \quad (12)$$

$$354 \tilde{\mathbf{V}}_i^k - \tilde{\mathbf{V}}_j^k = 2(\hat{\mathbf{r}}_{ij} \mathbf{P}_{ij}^k + \hat{\mathbf{x}}_{ij} \mathbf{Q}_{ij}^k) + |\hat{\mathbf{z}}_{ij}|^2 \tilde{\mathbf{I}}_{ij}^k + \mathbf{v}_j^k, \quad (13)$$

355 where  $\hat{\mathbf{z}}_{ij} = \hat{\mathbf{x}}_{ij} + j\hat{\mathbf{r}}_{ij}$  is the equivalent impedance matrix [29].  
356 Notice that constraint (12) is a second-order cone constraint  
357 which is convex. The terms  $(\cdot)^{\odot 2}$  denote slack variables, such that  
358 constraint (10)-(13) can be feasible if switch  $(i, j)$  is opened.  
359 To achieve this, the following constraints are applied:

$$360 \mathbf{0} \leq \hat{\mathbf{p}}_{ij}^k, \hat{\mathbf{q}}_{ij}^k \leq (1 - u_j^k) \mathbf{M}, \forall (i, j) \in \mathcal{B}, \quad (14)$$

$$361 \mathbf{0} \leq \mathbf{v}_j^k \leq (1 - u_j^k) \mathbf{M}, \forall j \in \mathcal{C}_i. \quad (15)$$

362 Moreover, we have the following constraints for the nodes  
363 if they are not restored by the  $k_{th}$  source node [30]:

$$364 \sum_{k \in \mathcal{K}} u_i^k \tilde{\mathbf{V}}_i^{\min} \leq \sum_{k \in \mathcal{K}} \tilde{\mathbf{V}}_i^k \leq \sum_{k \in \mathcal{K}} u_i^k \tilde{\mathbf{V}}_i^{\max}, \quad (16)$$

$$365 \mathbf{0} \leq \mathbf{P}_{ij}^k \leq u_j^k \mathbf{P}_k^{\max}, \forall j \in \mathcal{C}_i, k \in \mathcal{K}, \quad (17)$$

$$366 \mathbf{0} \leq \mathbf{Q}_{ij}^k \leq u_j^k \mathbf{Q}_k^{\max}, \forall j \in \mathcal{C}_i, k \in \mathcal{K}, \quad (18)$$

$$367 0 \leq \tilde{\mathbf{I}}_{ij}^k \leq u_j^k \tilde{\mathbf{I}}_{ij}^{\max}, \forall j \in \mathcal{C}_i, k \in \mathcal{K}. \quad (19)$$

368 Constraint (16) means that if node  $i$  is not activated by any  
369 source  $k$ , its voltage should be 0. This constraint also ensures  
370 that all the nodal voltages remain within the specified limits  
371 if the nodes are activated. Constraints (17)-(19) enforce power  
372 flow and current on line  $(i, j)$  to 0, if node  $j$  is not powered by  
373 source  $k$ . Note that Constraints (17)-(18) also limit the output  
374 of source  $k$  within rated capacity.

375 3) *Radial Network Constraints*: The operation of tie-lines  
376 can introduce potential loops. Hence, the following constraints  
377 based on single commodity flow model are essential to main-  
378 tain a spanning-tree after topology reconfiguration [31].

$$379 \sum_{i \in \pi_j} F_{ji} - \sum_{i \in \pi_j} F_{ij} = -1, \forall j \notin \mathcal{K}, \quad (20)$$

$$380 \sum_{i \in \pi_k} F_{ki} \geq 1, \forall k \in \mathcal{K}, \quad (21)$$

$$381 \sum_{(i,j) \in \mathcal{B}} \beta_{ij} = |\mathcal{N}| - |\mathcal{K}|, \quad (22)$$

$$382 -\Gamma_{ij} \mathbf{M} \leq F_{ij} \leq \Gamma_{ij} \mathbf{M}, \forall (i, j) \in \mathcal{B}, \quad (23)$$

$$383 \gamma_{ij} \leq \Gamma_{ij}, \forall (i, j) \in \mathcal{B}. \quad (24)$$

384 Constraint (20)-(22) ensures radial connectivity by requiring  
385 that exactly one unit of a fictitious commodity flows from the  
386 source node to every other node in the network. Constraint  
387 (23) enforces the fictitious flow on line  $(i, j)$  to 0 if it is dis-  
388 connected. These constraints achieves radiality by duplicating  
389 the distribution network into a fictitious network with the same  
390 topology. Ensuring radiality in the fictitious network inherently  
391 guarantees radiality in the original network. Constraint (24)  
392 ensures that a switch can only be operated when the radiality  
393 condition is satisfied.

394 4) *Investment Constraints*: To ensure that the hardening  
395 decisions in the DS are financially viable and in accor-  
396 dance with the available budget, the following constraints are  
397 applied:

$$398 \sum_{z \in \mathcal{Z}_s} (c_z^{\text{pl}} h_z^{\text{pl}} + c_z^{\text{ud}} h_z^{\text{ud}}) + \sum_{z \in \mathcal{Z}_t} c_z^{\text{pd}} h_z^{\text{pd}} \leq B^{\max}, \quad (25)$$

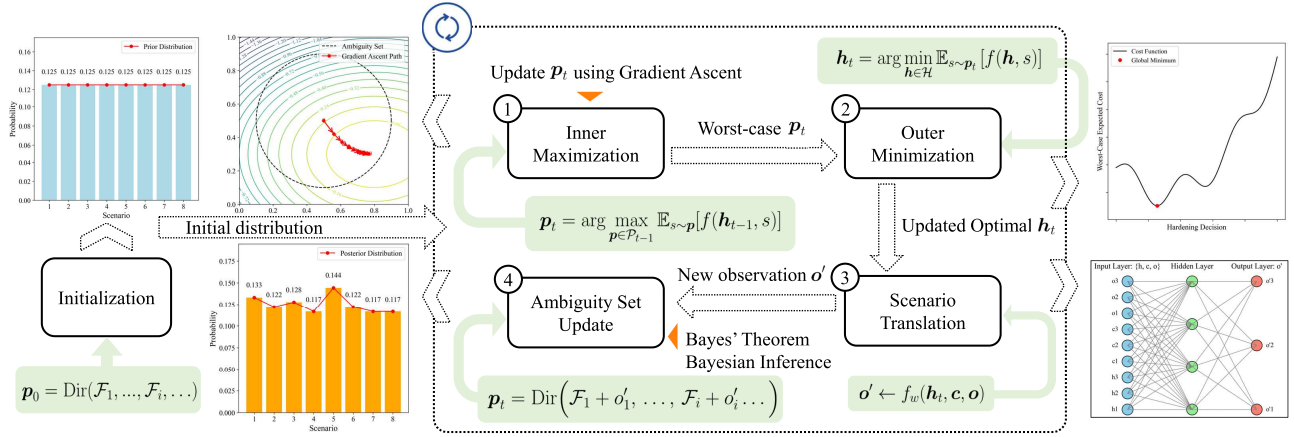
$$399 h_z^{\text{pl}} + h_z^{\text{ud}} \leq h_z, z \in \mathcal{Z}_s, \quad (26)$$

$$400 h_z^{\text{pl}} \leq 0, z \in \mathcal{P} \text{ or } h_z^{\text{ud}} \leq 0, z \in \mathcal{U}. \quad (27)$$

401 Constraint (25) ensures the total investment must be within  
402 the budget. Constraint (26) stipulates that each line segment  
403 can be hardened by only one measure. Constraint (27) includes  
404 geographic requirements based on local conditions, allowing  
405 only pole upgrading or line undergrounding in certain areas.

### 411 B. Problem Objective

412 The data-driven stochastic DL hardening problem is formu-  
413 lated as a decision-dependent DRO model. The objective is



414 Fig. 2. An illustration of the proposed Bayesian online learning algorithm.

415 to minimize the worst-case expected unserved load over an  
416 ambiguity set of probability distributions:

$$417 \quad (P1) \quad \min_{h \in \mathcal{H}} \max_{p \in \mathcal{P}} \mathbb{E}_{s \sim p} [f(\mathbf{h}, s)], \quad (28)$$

418 where  $\mathbb{E}_{s \sim p} [f(\mathbf{h}, s)]$  denotes the expected cost over all possible  
419 scenarios. For a finite outage scenario set, the expectation  
420 simplifies to  $\mathbb{E}_{s \sim p} [f(\mathbf{h}, s)] = \sum_i p_i f(\mathbf{h}, s_i)$  based on sample  
421 average approximation (SAA), where  $f(\mathbf{h}, s)$  is given by

$$422 \quad f(\mathbf{h}, s) = f_1(\mathbf{h}, s) + f_2(\mathbf{h}, s), \quad (29)$$

$$423 \quad f_1(\mathbf{h}, s) = \sum_{i \in \mathcal{N}, \phi \in \{a, b, c\}} \omega_i \sigma_{i,s} \Delta t_s, \quad (30)$$

$$424 \quad f_2(\mathbf{h}, s) = \sum_{(i,j) \in \mathcal{L}, \phi \in \{a, b, c\}} \left( r_{ij} \sum_{k \in \mathcal{K}} \tilde{I}_{ij}^k \right), \quad (31)$$

426 where  $f_1(\mathbf{h}, s)$  is the total weighted unserved load function,  
427 with  $\sigma_{i,s} \Delta t_s$  denoting the unserved load at node  $i$  under  
428 scenario  $s$ . Also,  $f_2(\mathbf{h}, s)$  is the total power losses function.  
429 This term is added such that the exactness of convex relaxation  
430 of the second-order cone constraint (12) is guaranteed [29].

431 To model the uncertainty in the probability distribution of  
432 outage scenarios, an ambiguity set  $\mathcal{P}$  is constructed around the  
433 empirical distribution  $\hat{\mathbf{p}}$ . This set uses the  $\ell_2$ -norm to define  
434 the proximity to the empirical distribution:

$$435 \quad \mathcal{P} = \{ \mathbf{p} \in \mathbb{R}^n : \|\mathbf{p} - \hat{\mathbf{p}}\|_2 \leq d \}, \quad (32)$$

436 where  $\mathbf{p}$  denotes all potential distributions of damage scenarios  
437 within a distance  $d$  of the empirical distribution. The parameter  
438  $d$  captures the uncertainty, reflecting the unmodeled variability  
439 in the DNN regression model. Note that the empirical distribu-  
440 tion  $\hat{\mathbf{p}}$  can be extracted from historical data, while it varies  
441 with respect to hardening decision  $\mathbf{h}$ , making the DRO model  
442 decision dependent. Hence, it cannot be directly solved using  
443 traditional approaches such as Benders' decomposition.

444 The  $\ell_2$ -norm is selected for both statistical and computa-  
445 tional reasons. Statistically, it can offer finite sample coverage  
446 guarantees for empirical distributions. Computationally, it  
447 enables closed-form gradient evaluation and projection via a  
448 simple bisection method with complexity  $\mathcal{O}(|\mathcal{S}|)$ . Its smooth-  
449 ness and convexity also facilitate dynamic regret analysis.  
450 While the Wasserstein distance provides advantages in con-  
451 tinuous or spatially structured spaces, its application requires

452 solving large-scale linear programs with  $\mathcal{O}(|\mathcal{S}|^2)$  constraints.  
453 However, our strategy is not restricted to the Euclidean metric.  
454 If future applications require more complicated representations  
455 of uncertainty, such as Wasserstein-based ambiguity sets, the  
456 framework can be extended by modifying the projection step,  
457 without requiring fundamental changes to the overall structure.

#### 458 IV. DATA-DRIVEN STOCHASTIC DL HARDENING 459 PROBLEM SOLUTION

460 To address the decision-dependent uncertainty and enhance  
461 computational efficiency, we propose a Bayesian online learn-  
462 ing algorithm. As illustrated in Fig. 2, the algorithm iteratively  
463 refines hardening decisions by sequentially incorporating out-  
464 age data. Specifically, at each iteration, the algorithm employs  
465 the proposed DNN regression model to translate an observed  
466 outage instance obtained under fixed network conditions into a  
467 probability distribution that quantifies how a specific hardening  
468 decision alters outage likelihoods. The DNN prediction is then  
469 used to update the posterior distribution via Bayes' Theorem,  
470 yielding a refined empirical distribution that acts as the center  
471 of a decision-specific ambiguity set. After that, the original  
472 decision-dependent DRO problem is decomposed into an inner  
473 problem and an outer problem. The inner problem utilizes  
474 projected gradient ascent to identify the worst-case probability  
475 distribution. The outer problem updates the hardening decision  
476 by minimizing the worst-case expected unserved load. Then,  
477 the posterior distribution is employed as the global prior for  
478 the next iteration. Through the online learning, the underlying  
479 outage uncertainties are iteratively refined, and the hardening  
480 decision ultimately converges to robust solutions that capture  
481 real-world outage complexities.

482 Even though integrating multiple components, the proposed  
483 algorithm ensures robustness and stability. First, the DNN  
484 predicts the probabilistic distribution over outage scenarios,  
485 with prediction errors accounted for through the ambiguity set.  
486 Second, the DRO model, convex in the scenario probabilities,  
487 provides both the worst-case expected performance and its  
488 gradient to the online learning. It ensures that even a single  
489 projected gradient ascent step suffices for useful updates, with-  
490 out solving the inner loop to full optimality. Third, ambiguity  
491

**Algorithm 1** Bayesian Online Learning Algorithm

**Require:** Cost function  $f(\mathbf{h}, s)$ , budget  $B^{\max}$ , and real-world data.

**Ensure:** Hardening decision  $\mathbf{h}_T$ .

1: **Initialization:**

2: Set  $t = 0$  and initialize hardening decision  $\mathbf{h}_0$  arbitrarily.

3: Initialize Dirichlet prior:

$$\mathbf{p}_0 = \text{Dir}(\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_{|S|}),$$

where  $\mathcal{F}_i = 1$  for a uniform prior.

4: Construct ambiguity set  $\mathcal{P}_0$  centered at  $\mathbf{p}_0$  with radius  $d_0$ .

5: **for**  $t = 1, 2, \dots, T$  **do**

6: **Inner Maximization:** Identify the worst-case distribution

7: Compute the gradient ascent update:

$$\tilde{\mathbf{p}}_t = \mathbf{p}_{t-1} + \eta \nabla_{\mathbf{p}} \mathbb{E}_{s \sim \mathbf{p}_{t-1}} [f(\mathbf{h}_{t-1}, s)].$$

8: Project  $\tilde{\mathbf{p}}_t$  onto  $\mathcal{P}_{t-1}$  to ensure feasibility:

$$\mathbf{p}_t = \arg \min_{\mathbf{p} \in \mathcal{P}_{t-1}} \frac{1}{2} \|\mathbf{p} - \tilde{\mathbf{p}}_t\|_2^2.$$

9: **Outer Minimization:** Update hardening decision

$$\mathbf{h}_t = \arg \min_{\mathbf{h} \in \mathcal{H}} \mathbb{E}_{s \sim \mathbf{p}_t} [f(\mathbf{h}, s)].$$

10: **Scenario Translation:** Transform outage scenarios

11:

$$\mathbf{o}' \leftarrow f_w(\mathbf{h}_t, \mathbf{c}, \mathbf{o}).$$

12: **Ambiguity Set Update:** Update the probability distribution.

13: Update the Dirichlet parameters with fractional counts:

$$\mathbf{p}_t(D_t, \mathbf{h}_t) = \text{Dir}(\mathcal{F}_1 + o'_1, \dots, \mathcal{F}_{|S|} + o'_{|S|}).$$

14: Compute the posterior mean for each scenario:

$$\bar{\mathbf{p}}_t = \left[ \frac{\mathcal{F}_1 + o'_1}{\sum_{j=1}^{|S|} (\mathcal{F}_j + o'_j)}, \dots, \frac{\mathcal{F}_{|S|} + o'_{|S|}}{\sum_{j=1}^{|S|} (\mathcal{F}_j + o'_j)} \right].$$

15: Update the ambiguity set for the next iteration as:

$$\mathcal{P}_{t+1}(\mathbf{h}_t) = \left\{ \mathbf{p} \in \mathbb{R}^{|S|} : \|\mathbf{p} - \bar{\mathbf{p}}_t\|_2 \leq d_{t+1} \right\},$$

with the confidence radius  $d_t = \sqrt{2|S| \log(2/\delta_t)/t}$

16: **end for**

17: **Output:** Return the final hardening decision  $\mathbf{h}_T$ .

sets are updated online via Dirichlet counts based on observed scenario occurrences, without requiring DNN retraining.

### A. Bayesian Online Learning Algorithm

The pseudo code for the algorithm is presented in Algorithm 1. Main steps of the algorithm are explained as follows.

1) *Initial Setup and Ambiguity Set Definition:* At the beginning of the learning process, initial hardening decisions are arbitrarily chosen to provide a starting point for the algorithm. An initial prior distribution is defined as:

$$\mathbf{p}_0 = \text{Dir}(\mathcal{F}_1, \dots, \mathcal{F}_i, \dots, \mathcal{F}_{|S|}), \quad (33)$$

where  $\text{Dir}(\cdot)$  denotes the Dirichlet distribution, and  $\mathcal{F}_i$  represents the initial cumulative fractional count for scenario  $s_i$ . Before any data is learned,  $\mathbf{p}_0$  can be initialized as uniform distribution by having all  $\mathcal{F}_i = 1$ . It can also be initialized based on expert knowledge. Note that  $\mathbf{p}_0$  acts as the empirical distribution of the initial ambiguity set. And, as the algorithm progresses, the  $\mathcal{F}$  will be updated by adding fractional counts derived from the DNN's output.

2) *Inner Maximization:* Based on hardening decision at last step  $\mathbf{h}_{t-1}$ , the inner problem is solved to identify the worst-case probability distribution within the ambiguity set  $\mathcal{P}_{t-1}$  that maximizes the expected cost, given by

$$(P2) \quad \mathbf{p}_t = \arg \max_{\mathbf{p} \in \mathcal{P}_{t-1}} \mathbb{E}_{s \sim \mathbf{p}} [f(\mathbf{h}_{t-1}, s)], \quad (34)$$

where  $\mathbf{p}_t$  represents the most adverse probability corresponding to the given decision  $\mathbf{h}_{t-1}$ . Since  $\mathbf{p} \in \mathcal{P}_{t-1}$  is a probability distribution over a finite scenario set  $\mathcal{S}$ , the inner problem is linear in  $\mathbf{p}$ . Consequently, the inner problem is differentiable w.r.t  $\mathbf{p}$ , making it suitable to the gradient ascent method:

$$\tilde{\mathbf{p}}_t = \mathbf{p}_{t-1} + \eta \nabla_{\mathbf{p}} \mathbb{E}_{s \sim \mathbf{p}_{t-1}} [f(\mathbf{h}_{t-1}, s)], \quad (35)$$

where  $\eta$  is a small step size, and  $\tilde{\mathbf{p}}_t$  is an approximation of  $\mathbf{p}_t$ . Equation (35) allows for adjustment of  $\mathbf{p}$  in the direction of the gradient, which increases the expected cost for maximization. Note that  $\tilde{\mathbf{p}}_t$  is an intermediate distribution, which may not yet satisfy the constraints of the ambiguity set, such as  $\tilde{\mathbf{p}}_t(s_i) \geq 0$ ,  $\sum_i \tilde{\mathbf{p}}_t(s_i) = 1$ , and  $\|\tilde{\mathbf{p}}_t - \mathbf{p}_{t-1}\|_2 \leq d_{t-1}$ . After gradient ascent, if  $\tilde{\mathbf{p}}_t$  violates the constraints even after normalization, it is therefore projected back onto the ambiguity set  $\mathcal{P}_{t-1}$  to ensure feasibility. This projection can be achieved by solving a convex quadratic programming problem:

$$\mathbf{p}_t = \arg \min_{\mathbf{p} \in \mathcal{P}_{t-1}} \frac{1}{2} \|\mathbf{p} - \tilde{\mathbf{p}}_t\|_2^2, \quad (36)$$

where  $\|\mathbf{p} - \tilde{\mathbf{p}}_t\|_2^2$  is the regularization term utilized to prevent large jumps from  $\mathbf{p}_{t-1}$ . By combining equations (35)–(36), the probability  $\mathbf{p}_t$  for the next iteration can be optimized by a compact-form optimization problem which simultaneously considers the gradient ascent and regularization, given by

$$\min_{\mathbf{p} \in \mathcal{P}_{t-1}} \langle \eta \nabla_{\mathbf{p}} \mathbb{E}_{s \sim \mathbf{p}_{t-1}} [f(\mathbf{h}_{t-1}, s)], \mathbf{p} \rangle + \frac{1}{2} \|\mathbf{p} - \mathbf{p}_{t-1}\|_2^2 \quad (37)$$

$$s.t. \quad \mathbf{p} \geq \mathbf{0}, \langle \mathbf{p}, \mathbf{1} \rangle = 1, \|\mathbf{p} - \mathbf{p}_{t-1}\|_2 \leq d_{t-1}$$

where the first inner product term  $\langle \cdot, \cdot \rangle$  represents how much the updated  $\mathbf{p}_t$  aligns with the gradient direction. Notice that convergence of the inner problem is guaranteed as long as  $\mathbf{p}$  is updated in a valid ascent direction at each iteration. Solving the inner maximization problem to full optimality is not required. Even though performing multiple projected gradient steps per iteration may yield a tighter approximation of the worst-case distribution, it is unnecessary in practice considering the trade-off between convergence and computational efficiency.

These three equations collectively describe the same projected gradient update procedure, but from different and complementary perspectives. Equation (35) describes an unconstrained gradient ascent step, which updates the scenario probability vector in the direction of the gradient without

enforcing any feasibility constraints. As a result, the intermediate update  $\tilde{\mathbf{p}}_t$  may fall outside the feasible region. Equation (36) then introduces a projection step, which maps  $\tilde{\mathbf{p}}_t$  back onto the admissible set defined by the  $\ell_2$ -norm trust region around the previous iteration. This projection not only ensures feasibility but also serves as a regularization mechanism that stabilizes the update and limits the step size. Finally, equation (37) presents the whole projected gradient update into a single constrained optimization problem, explicitly incorporating non-negativity, normalization, and proximity constraints. While equation (37) alone is sufficient from an algorithmic standpoint, we remain equations (35)-(36) and their respective explanations to make the underlying “gradient-then-project” logic transparent. This helps readers better understand how the direction of update is first determined and then regularized through projection.

3) *Outer Minimization:* Using the worst-case distribution  $\mathbf{p}_t$ , the outer problem is further solved to update the hardening decision  $\mathbf{h}_t$ . The goal is to minimize the expected cost under the identified worst-case scenario, given by

$$(P3) \quad \mathbf{h}_t = \arg \min_{\mathbf{h} \in \mathcal{H}} \mathbb{E}_{s \sim \mathbf{p}_t} [f(\mathbf{h}, s)]. \quad (38)$$

4) *Scenario Translation:* Once the DL hardening  $\mathbf{h}_t$  decision at step  $t$  is obtained, it is input into the DNN regression model along with randomly selected outage data instance:

$$\mathbf{o}' \leftarrow f_w(\mathbf{h}_t, \mathbf{c}, \mathbf{o}). \quad (39)$$

where  $\mathbf{o}'$  denotes the DNN’s predicted probability distribution over outage scenarios at the current iteration. This process directly addresses decision-dependent uncertainty, simulating the transformation of outage scenarios from data collected under fixed network conditions to those expected under hypothetical hardening decision  $\mathbf{h}_t$ .

5) *Ambiguity Set Update:* At each iteration, the probability distribution over all scenarios  $s$  output by the DNN model is used to update the likelihood of outage scenarios, given by

$$p(D_t | \mathbf{p}_{t-1}, \mathbf{h}_t) = \left[ o'_1, o'_2, \dots, o'_{|S|} \right], \quad (40)$$

where  $o'_s$  represents the fractional counts of scenarios, serving as the newest observation of the underlying outage uncertainty. Accordingly, by using Bayes’ Theorem, the posterior distribution is updated as

$$\mathbf{p}_t(D_t, \mathbf{h}_t) \propto p(D_t | \mathbf{p}_{t-1}, \mathbf{h}_t) \mathbf{p}_{t-1}. \quad (41)$$

Because the Dirichlet distribution is the conjugate prior for the multinomial distribution, the posterior distribution remains the Dirichlet form, with its parameters updated by adding the new fractional counts  $o'_i$  to the existing counts, given by

$$\mathbf{p}_t(D_t, \mathbf{h}_t) = \text{Dir}(\mathcal{F}_1 + o'_1, \mathcal{F}_2 + o'_2, \dots, \mathcal{F}_{|S|} + o'_{|S|}), \quad (42)$$

where  $\mathcal{F}_i$  denotes the fractional counts for  $s_i$  from previous iterations. The Dirichlet prior is adopted, as it offers a natural and analytically tractable approach for modeling uncertainty over discrete scenario probabilities. Its conjugacy with the categorical likelihood facilitates efficient Bayesian updating as new scenario observations are acquired. Equation 42 imply that the new observation  $\mathbf{o}'$  is governed by the prior belief  $\mathcal{F}$ ,

which reflects the sequential learning process of the proposed algorithm. Then, by applying the expectation for the Dirichlet distribution, we have

$$\mathbb{E}[p(s_i)] = \frac{\mathcal{F}_i + o'_i}{\sum_{j=1}^{|S|} (\mathcal{F}_j + o'_j)}. \quad (43)$$

Note that  $\mathbb{E}[p(s_i)]$  provides the current best probability estimation for each scenario  $s_i$ . Therefore, based on the posterior mean, the decision-specific ambiguity set is constructed:

$$\mathcal{P}_{t+1}(\mathbf{h}_t) = \{ \mathbf{p} \in \mathbb{R}^n : \|\mathbf{p} - \bar{\mathbf{p}}_t\|_2 \leq d_{t+1} \}. \quad (44)$$

where  $\bar{\mathbf{p}}_t = [\mathbb{E}[p(s_1)], \mathbb{E}[p(s_2)], \dots, \mathbb{E}[p(s_{|S|})]]$ . Also,  $d_{t+1}$  at the current step reflects our confidence level in the updated posterior in representing the empirical distribution. Note that the term “decision-specific ambiguity set” refers to the fact that the posterior update is conditioned on the hardening decision implemented. Specifically, at iteration  $t$ , a hardening vector  $\mathbf{h}_t$  is selected, and the DNN is utilized to infer the scenario probability distribution corresponding to the observed outage under this decision. This updates the Dirichlet prior via Bayes’ rule, yielding a posterior mean  $\bar{\mathbf{p}}_t$ , which is then used to define the ambiguity set for the next iteration.

By employing Bayesian Inference, the posterior distribution from one decision is used as the global prior in the subsequent iteration to represent an updated belief about the uncertainty. A critical aspect of the online learning algorithm is its ability to address decision-dependent uncertainty by constructing data-driven and decision-specific ambiguity sets  $\mathcal{P}_{t+1}(\mathbf{h}_t)$ , while iteratively learning decisions based on incoming data. The data learned under one decision not only updates the posterior and ambiguity set for this decision, but also influences other decisions through the global prior, facilitating the knowledge transfer between decisions. Furthermore, as more data is incorporated, the empirical distribution becomes increasingly precise, and the distance of the ambiguity set shrinks. This reduced distance signifies that the ambiguity set aligns more closely with the true underlying distribution, mitigating over-conservatism and enhancing decision-making accuracy. This

shrinking distance can be modeled as  $d_t = \sqrt{2|S| \log\left(\frac{2}{\delta_t}\right)}/t$ ,

where  $\delta_t = \frac{6\delta}{\pi^2 t^2}$  [32]. Since the algorithm processes outage records sequentially, one data instance per iteration, the iteration index  $t$  also represents the cumulative number of observed samples. Note that  $\delta_t$  ensures the confidence level, initially  $1 - \delta$ , is maintained over time. Also, the distance selection is based on the  $\ell_2$ -norm, which guarantees that the true scenario distribution lies within  $\mathcal{P}_t$  with probability at least  $1 - \delta_t$ . An important insight of the online learning algorithm is that a hardening decision impacts only the uncertainty of damage scenarios, rather than the outcomes themselves. As a result, the ambiguity sets for all decisions are equivalent at each iteration. This simplifies the algorithm by requiring only a single ambiguity set to be maintained during the iterative process, significantly reducing computational complexity.

## B. Convergence Analysis Using Dynamic Regret

To demonstrate the performance of the proposed Bayesian online learning algorithm, its convergence is analyzed.

Specifically, the dynamic regret is applied to evaluate the cumulative performance gap between the worst-case performance and the optimal worst-case performance of solving the DRO at each time step  $t$ . The dynamic regret can be stated as

$$\mathcal{D}_T = \frac{1}{T} \sum_{t=1}^T \left( \mathbb{E}_{s \sim p_t} [f(\mathbf{h}_t, s)] - \mathbb{E}_{s \sim p_t^*} [f(\mathbf{h}_t^*, s)] + \epsilon_t \right), \quad (45)$$

where  $p_t$  is the worst-case distribution. And,  $\mathbf{h}_t$  represent the hardening decision at iteration  $t$  produced by the algorithm,  $\mathbf{h}_t^*$  represent the optimal decision for the same ambiguity set  $\mathcal{P}_t$ . As the online algorithm include the DNN regression model in the loop, the prediction error can propagate through the empirical distribution  $\hat{p}_t$ . To evaluate its impact on the convergence, a residual  $\epsilon_t$  which quantifies the prediction error is added in equation (45). Note that  $\epsilon_t = \|\hat{p}_t - p_t^*\|_2$ , where  $p_t^*$  is the true distribution. Based on the fact that the residual  $\epsilon_t$  is bounded by the shrinking distance  $d_t$  of the ambiguity set, i.e.,  $d_t = O\left(\sqrt{\log(t)/t}\right)$ , and the empirical distribution stabilizes as more data is learned over time, hence the cumulative effect of residuals across iterations asymptotically decays.

*Theorem.* The dynamic regret is bounded:

$$\mathcal{D}_T \leq B \left( \sqrt{L/T} + O\left(\sqrt{\log(T)/T}\right) \right), \quad (46)$$

where  $B$  is the bound for function  $f(\mathbf{h}, s)$ , i.e.,  $f(\mathbf{h}, s) \leq B$ .  $L$  is the cumulative path length of the ambiguity sets  $\mathcal{P}_t$  over all iterations, defined as  $L = \sum_{t=1}^T \|p_t - p_{t-1}\|_2^2$ . Equation (46) implies that the regret decreases sublinearly with  $T$ , ensuring convergence as the algorithm progresses. In other words, the hardening decision derived by the proposed online algorithm become increasingly closer to the exact DRO solutions as more data is learned sequentially.

*Proof:* Using the triangle inequality and the boundedness of function  $f(\mathbf{h}, s)$ , the deviation at each step is stated as:

$$\mathbb{E}_{s \sim p_t} [f(\mathbf{h}_t, s)] - \mathbb{E}_{s \sim p_t^*} [f(\mathbf{h}_t^*, s)] + \epsilon_t \leq B \|p_t - p_t^*\|_2, \quad (47)$$

where  $p_t$  and  $p_t^*$  denote the worst-case distribution by using the online algorithm and solving the DRO exactly at time step  $t$ , respectively. Then, summing over all iterations, the dynamic regret can be derived as

$$\mathcal{D}_T \leq \frac{1}{T} \sum_{t=1}^T B \|p_t - p_t^*\|_2 + \frac{1}{T} \sum_{t=1}^T \epsilon_t. \quad (48)$$

Since the term  $\|p_t - p_t^*\|_2$  can be bounded as  $\|p_t - p_t^*\|_2 \leq \|p_t - p_{t-1}\|_2 + \|p_{t-1} - p_t^*\|_2$ , the right hand side becomes:

$$\frac{1}{T} B \left( \sum_{t=1}^T \|p_t - p_{t-1}\|_2 + \sum_{t=1}^T \|p_{t-1} - p_t^*\|_2 \right) + \frac{1}{T} \sum_{t=1}^T \epsilon_t. \quad (49)$$

For the first term in the bracket, using the Cauchy-Schwarz inequality, the following inequality can be derived,

$$\sum_{t=1}^T \|p_t - p_{t-1}\|_2 \leq \sqrt{T \sum_{t=1}^T \|p_t - p_{t-1}\|_2^2}, \quad (50)$$

which can simply rewritten as  $\sum_{t=1}^T \|p_t - p_{t-1}\|_2 \leq \sqrt{TL}$ .

For the second term in the bracket, since  $\|p_{t-1} - p_t^*\|_2 \leq d_t$ , where  $d_t = \sqrt{\frac{2|\mathcal{S}| \log(\frac{2}{\delta_t})}{t}}$ ,  $\delta_t = \frac{6\delta}{\pi^2 T}$ , we have

$$\sum_{t=1}^T d_t \leq 4 \sqrt{2|\mathcal{S}|T} \left( \sqrt{\log(T)} + \sqrt{\log\left(\frac{\pi^2}{3\delta}\right)} \right). \quad (51)$$

The term  $\frac{1}{T} \sum_{t=1}^T \epsilon_t$  is the average residual. As  $\epsilon_t = \|p_t - p_t^*\|_2$  is bounded by the distance  $d_t$ , by substituting  $d_t$  and  $\delta_t$ , the bound for the average residual over time can be derived:

$$\frac{1}{T} \sum_{t=1}^T \epsilon_t \leq \frac{\sqrt{2|\mathcal{S}|} \cdot (\log(T))^{3/2}}{1.5 T}, \quad (52)$$

which means the asymptotic bound on the average residual is  $O((\log(T))^{3/2}/T)$ , which is faster than  $O(1/\sqrt{T})$ .

Combining these results, the final dynamic regret is

$$\mathcal{D}_T \leq B \left( \sqrt{L/T} + O\left(\sqrt{\log(T)/T}\right) \right), \quad (53)$$

where we omit the Big- $O$  term for the average residual, since its contribution is asymptotically dominated by others.

### C. Scalability Analysis

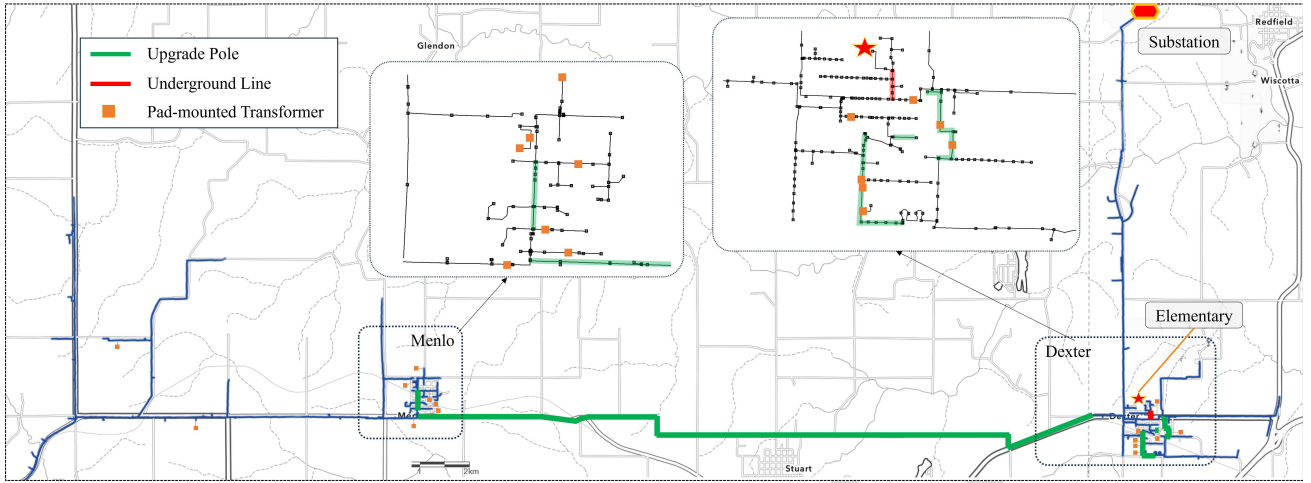
The proposed strategy is mathematically scalable and practically applicable to much larger systems in both network size and scenario complexity. The computational cost each iteration is dominated by four components: DNN inference, ambiguity set update, projected gradient ascent for the inner problem, and the outer mixed-integer second-order cone programming. The first three are lightweight and scale linearly with the number of scenarios  $\mathcal{O}(|\mathcal{S}|)$ . Specifically, the DNN inference produces a scenario probability vector; the Bayesian update involves simple Dirichlet count increments; and projection onto the  $\ell_2$ -norm ambiguity set is a convex optimization problem. Thus, the projected gradient update similarly incurs low cost due to convexity. The outer problem is the most computationally intensive step, yet it is solved by using commercial solvers such as Gurobi, which employs convex relaxations and presolving to enhance efficiency. We want to note that the entire learning and optimization process is performed offline as part of long-term infrastructure planning. Unlike real-time restoration tasks, hardening plans are determined in advance. Therefore, even for systems comprising thousands of nodes and hundreds of candidate hardening options, the strategy remains tractable.

## V. CASE STUDY

In this section, the test system is set up. The DNN regression model is trained using real-world outage data. The simulation results of Bayesian online learning algorithm and data-driven stochastic DL hardening strategy are presented. The comparisons between different hardening strategies are discussed.

### A. Test System and Data Preparation

1) *Test System Setup:* In this paper, a real-world 24.9kV power distribution network located in Redfield, Iowa, USA is selected for simulation and case study. The coordination



800 Fig. 3. A geographic illustration of the Redfield distribution network.

801 TABLE I  
802 REAL-WORLD OUTAGE DATA STRUCTURE

| $s$ | Time             | Clear Devices | S (mph) | H (%) | T ( $^{\circ}$ C) |
|-----|------------------|---------------|---------|-------|-------------------|
| 1   | 2015-05-12 14:32 | Transformer   | 30      | 65    | 20.9              |
| 2   | 2017-08-25 17:15 | Recloser      | 40      | 74    | 27.7              |
| 3   | 2019-11-03 09:45 | Fuse          | 25      | 62    | 18.2              |
| 4   | 2021-03-15 12:05 | Transformer   | 35      | 68    | 12.3              |
| 5   | 2023-07-07 19:20 | Fuse          | 45      | 93    | 30.1              |

803 is  $< 41^{\circ}35'36.1''N, 94^{\circ}13'37.3''W >$ . The network features  
 804 a single feeder with a U-shaped configuration. It serves  
 805 two major load groups located in Dexter, Iowa, and Melon,  
 806 Iowa, respectively. An illustration of the distribution net-  
 807 work is depicted in Fig. 3. Specifically, the total length of  
 808 the main feeder is 47.54 miles, including 597 nodes, 226  
 809 distribution transformers, 19 switches, 1 circuit breaker, 2  
 810 recloser, and 60 fuses. The total load demand are 1,813kW  
 811 and 1,683kVar, which are obtained from the utility based  
 812 on historical electricity consumption. The outage data, span-  
 813 ning from 2001 to 2024, is collected by the utility via the  
 814 SCADA systems. This dataset contains time stamps, outage  
 815 duration, and clearing devices. Examples of outage scenarios  
 816 are listed in Table I. Weather conditions including wind  
 817 speed and direction, humidity, and temperature, are obtained  
 818 from the Iowa Environmental Mesonet [33]. The hardening  
 819 cost for main feeders is 0.3M/mile for upgrading poles [34],  
 820 and 3.0M/mile for undergrounding lines [35], respectively.  
 821 The hardening cost for laterals is 0.05M for pad-mounted  
 822 transformers [36]. In addition, an Intel Xeon W-1370 work-  
 823 station, with 2.90 GHz and 32 GB RAM, and Python 3.13,  
 824 TensorFlow 2.16 and Gurobi solver 12.0 are used as a test  
 825 platform.

826 2) *DNN Training and Validation*: Before training the  
 827 DNN, the raw outage dataset are preprocessed through data  
 828 augmentation. Specifically, for weather conditions, subtle per-  
 829 turbations are introduced by adding Gaussian noise with a  
 830 small standard deviation. Similarly, for outage events, minor  
 831 variations are introduced by adjusting the parameters of  
 832 clearing devices to mimic those of adjacent devices, thereby

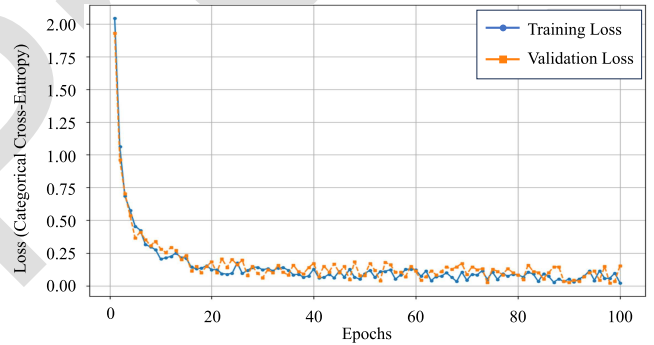
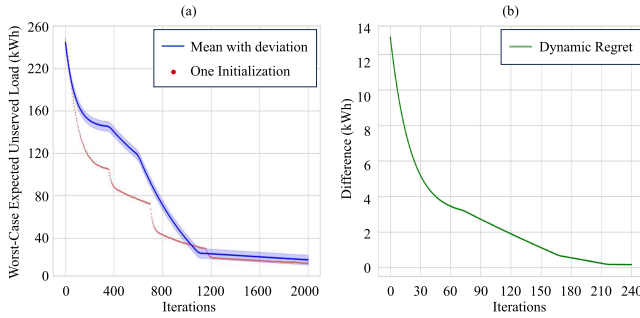


Fig. 4. An illustration of the learning process of the DNN model.

834 accounting for spatial uncertainties. To train the DNN model,  
 835 the outage dataset  $\mathcal{T}$  is randomly divided into three sub-  
 836 sets, including a training set for optimizing the parameters  
 837  $w$ , a validation set to fine-tune hyperparameters, and a test  
 838 set to assess the DNN's performance. The DNN model  
 839 consists of three hidden layers, with each layer containing  
 840 64 neurons. The Rectified Linear Unit (ReLU) is selected  
 841 as the activation function. The model is trained with a  
 842 mini-batch size of 32 samples for up to 100 epochs using  
 843 the Adam optimizer with an initial learning rate of 0.001.  
 844 Early stopping was applied by monitoring the validation  
 845 loss. Also, the categorical cross-entropy is adopted as the  
 846 loss function to measure the difference between the DNN  
 847 regression model's predicted and labeled distributions of fault  
 848 scenarios.

849 Fig. 4 depicts the training loss and validation loss curves.  
 850 We can see both the loss curves decrease obviously, which  
 851 effectively demonstrate the learning process. In particular,  
 852 the validation loss stabilizes around epoch 45. It means the DNN  
 853 model converges with minimal overfitting, and the validation  
 854 performance does not deteriorate despite additional training.  
 855 Furthermore, we evaluate the DNN model based on the test  
 856 set. The results show the model achieves an accuracy of 92%,  
 857 with a precision of 90% and a recall of 88%, to  
 858 predict the outage scenario with the highest probability. In  
 859 addition, the predicted probability distributions over all outage



AQ:2 Fig. 5. An illustration of Bayesian online learning process.

861 TABLE II

862 COMPUTATIONAL PERFORMANCE OF ONLINE LEARNING

| $ S $ | Online Learning | DRO Each Step | Efficiency Improvement |
|-------|-----------------|---------------|------------------------|
| 10    | 1.22s           | 2.34s         | 47.86%                 |
| 25    | 3.61s           | 7.64s         | 52.75%                 |
| 50    | 12.3s           | 32.0s         | 61.56%                 |
| 100   | 21.5s           | 85.6s         | 74.88%                 |

863 scenarios were assessed, resulting in a mean absolute error of  
 864 0.05 and a root mean squared error of 0.07. These results  
 865 confirm that the proposed DNN model accurately captures  
 866 the evolution of outage scenarios under different hardening  
 867 measures.

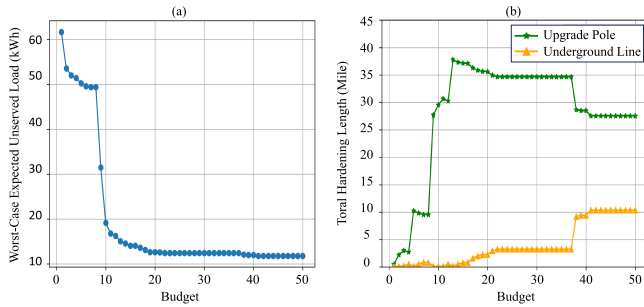
### 868 B. Bayesian Online Learning Implementation

869 In this section, we evaluate the performance of the Bayesian  
 870 online learning algorithm. The results are averaged over 50  
 871 random initializations, with both the mean and standard deviation  
 872 plotted to illustrate the stability of the algorithm. In  
 873 Fig. 5(a), the blue solid line represents the evolution of the  
 874 average worst-case expected load shed over 2,000 iterations.  
 875 Initially, the mean value is very high because the ambiguity  
 876 set is large due to limited data, meaning the empirical distri-  
 877 bution remains largely random. Accordingly, the hardening  
 878 decisions at the beginning can be overly conservative. Never-  
 879 theless, as iterations progress, more data are incorporated  
 880 and learned, the empirical distribution becomes more reliable  
 881 and the ambiguity set naturally shrinks. Consequently, the  
 882 worst-case expected unserved load exhibits a steady decline.  
 883 And, after iteration 1100, the objective value stabilizes around  
 884 around 20.00kWh. It implies the hardening decisions become  
 885 less conservative and more consistent with the actual outage  
 886 uncertainty. The red dot line in Fig. 5(a) shows the online  
 887 learning process of one initialization. The results reveal three  
 888 abrupt improvements at iteration 367, 691, and 1314. These  
 889 reflect the mixed-integer nature of the proposed hardening  
 890 problem. Also, the algorithm demonstrates the ability to escape  
 891 local optima and explore superior solutions. In addition, Fig.  
 892 5(b) presents the convergence of the dynamic regret, which  
 893 quantifies the performance gap between the online learning  
 894 algorithm and the exact DRO solution at each iteration. The  
 895 fact that the dynamic regret tends to zero after approximately  
 896 240 iterations indicates that the algorithm has accumulated  
 897 sufficient information to generate solutions that are effectively

equivalent to those of the exact DRO problem with the current  
 data. In other words, from that point onward, the online  
 learning algorithm is producing decisions that are reliably  
 near-optimal for the evolving problem. However, achieving  
 near-zero dynamic regret does not imply convergence of the  
 objective value. Since the ambiguity set continues to be  
 refined as new data are incorporated, the worst-case expecta-  
 tion may continue to change. This process continues until  
 around iteration 1100, when both the ambiguity set and the  
 decision vector  $h_t$  cease to change meaningfully. To further  
 demonstrate the computational performance, we conduct com-  
 parisons between the Bayesian online learning algorithm and  
 the exact DRO in terms of the average computational times  
 per iteration. The comparisons are based on 10, 25, 50, and  
 100 randomly selected outage scenarios. The results listed in  
 Table II indicate that the proposed online learning algorithm  
 requires substantially less computational time per iteration,  
 especially as the number of scenarios increases. This effi-  
 ciency is primarily due to our decomposition strategy, which  
 bypasses the need to solve the complete DRO problem in each  
 round.

### 919 C. Numerical Results and Discussion

920 In this section, we discuss about the numerical results  
 921 for the Redfield distribution network under both fixed and  
 922 variable budget scenarios. For the fixed case, a total hard-  
 923 ening budget of 10.0M is considered. Then, we determine  
 924 the hardening decision through the Bayesian online learning  
 925 algorithm. The results show that a combined solution of 29.5  
 926 miles of pole upgrading, 0.114 miles of undergrounding, and  
 927 the reinforcement of 16 distribution transformers can minimize  
 928 the worst-case expected unserved load to 19.17kWh. Fig. 3  
 929 illustrates the hardening solution. First, the components whose  
 930 failures can result in significant losses are hardened. For  
 931 instance, the long line segment connecting the two concen-  
 932 trated load groups with a length of 9.18 miles is prioritized  
 933 for hardening. This result is reasonable, as protecting critical  
 934 links between major load areas is essential. If this line seg-  
 935 ment is not hardened as suggested, the worst-case expected  
 936 unserved load will increase to 92.09kWh. Also, under the 10.0M  
 937 fixed budget, this long line segment can only be hardened  
 938 through upgrading poles due to the budget limits. Second, the  
 939 solution prioritizes the uninterrupted service of critical loads.  
 940 For instance, a key line segment supplying power to the West  
 941 Elementary School is specifically undergrounded, demonstrat-  
 942 ing the proposed strategy can effectively incorporate real-world  
 943 outage data and allocate resources efficiently. If this line seg-  
 944 ment is not hardened as suggested, the worst-case expected  
 945 unserved load would increase to 48.20kWh. Third, distribution  
 946 transformers with substantial unserved load scenarios are re-  
 947 placed with pad-mounted transformers. Furthermore, to evaluate  
 948 the impact of DNN prediction accuracy, we perform a sensitiv-  
 949 ity analysis by perturbing the DNN output probabilities with  
 950 zero-mean Gaussian noise with a standard deviation of 10%,  
 951 followed by renormalization. The hardening strategy is then  
 952 re-executed across 50 such perturbed scenarios. Simulation  
 953 results show that the average worst-case load shed increases  
 954 by only 3.59%, while the selected hardening lines maintains a



955 Fig. 6. Results of varying budget in terms of hardening outcomes.

956 93.8% overlap with the original plan. These findings indicate  
 957 that the DRO-based ambiguity set offers robustness against  
 958 moderate prediction errors, preserving decision quality under  
 959 uncertainty.

960 For the variable budget case, the comparative results are  
 961 shown in Fig. 6. Specifically, as the budget increases, the  
 962 worst-case expected unserved load is decreased. It demon-  
 963 strates that higher investments enhance system resilience. In  
 964 Fig. 6(b), at higher budget levels, it is obvious that the  
 965 solution increasingly prioritizes undergrounding over pole  
 966 upgrading. This shift occurs because undergrounding substan-  
 967 tially reduces outage probabilities. Conversely, when financial  
 968 resources are limited, pole upgrading remains a more cost-  
 969 effective alternative. For example, when the budget exceeds  
 970 15M, sufficient financial resources become available for under-  
 971 grounding. However, the reduction in the expected unserved  
 972 load beyond this threshold is less pronounced. This find-  
 973 ing highlights the importance of optimal balancing pole  
 974 upgrading and undergrounding rather than assuming that  
 975 increased undergrounding always yields the most resilient  
 976 measures.

#### 977 D. Comparative Study

978 To further demonstrate the advantages of the proposed data-  
 979 driven stochastic DL hardening strategy, we compare its  
 980 performance against the following strategies:

- 981 1) Robust optimization [9]: Minimizes the worst-case  
 982 unserved load considering the most severe outage sce-  
 983 nario.
- 984 2) Stochastic programming [13]: Minimizes unserved load  
 985 based on the expected performance over all scenarios.
- 986 3) DRO [17]: Minimizes the worst-case expected unserved  
 987 load by considering distributions using an ambiguity set.

988 Among these strategies, robust optimization is a determin-  
 989 stic approach without consideration of probabilities. In contrast,  
 990 both stochastic programming and DRO derive the probability  
 991 distributions of outage scenarios from mathematical models,  
 992 while our proposed data-driven strategy leverages real-world  
 993 outages. For the comparison, each hardening strategy is evalu-  
 994 ated based on multiple trials. Specifically, for each hardening  
 995 decision of each strategy, we implement 50 independent trials,  
 996 each trial includes 50 randomly generated scenarios. For each  
 997 scenario, the unserved load is calculated in terms of the fixed  
 998 hardening decision, and the average unserved load for the trial

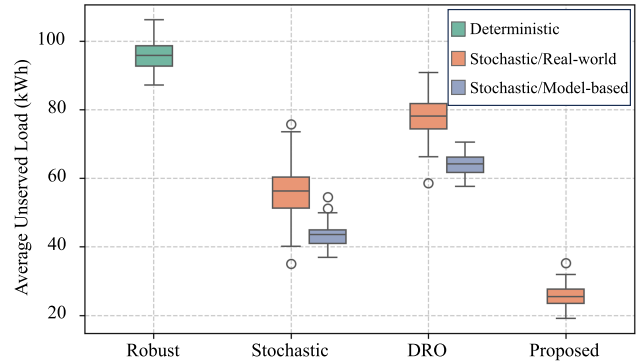


Fig. 7. Comparative results between hardening strategies.

is computed. This procedure is repeated to construct a distri-  
 bution of average unserved load values. From this distribution,  
 key statistical measures, including the mean, upper bound, and  
 lower bound, are derived, as presented in Fig. 7.

From these results, we can observe that hardening strategy  
 using robust optimization derives the highest average  
 unserved load, which is 96.31kWh. The reason is that robust  
 optimization explicitly protects against the worst-case sce-  
 nario. In contrast, stochastic programming maximizes expected  
 performance across all scenarios, which results in a lower  
 average unserved load of 57.11kWh. However, we can see  
 the discrepancy between the mean and upper bound is  
 larger in terms of real-world scenarios and model-based  
 scenarios. It means that stochastic programming can under-  
 estimate risk if the model-based distribution deviates from  
 actual outage behavior. In addition, the results of DRO lies  
 between RO and SP, offering a balance between conservatism  
 and adaptability. This is achieved by considering an ambi-  
 guity set of distributions. However, if the ambiguity set is  
 derived from an inaccurate model, discrepancies can still  
 arise when evaluated against real-world data. Furthermore,  
 the proposed data-driven strategy achieves the best over-  
 all performance with the lowest average unserved load of  
 24.05kWh. The reason is its ability to continuously update  
 the outage distributions based on online learning of real-world  
 outages, capturing the stochastic nature of system failures  
 more accurately.

## VI. CONCLUSION

In this paper, a data-driven stochastic DL hardening strategy  
 is proposed to leverage real-world data for making informed  
 hardening decisions. By integrating a DNN regression model  
 for outage distribution prediction with a decision-dependent  
 DRO model, the proposed strategy effectively incorporates  
 historical outage information into a data-driven ambiguity set.  
 A Bayesian online learning algorithm is proposed to address  
 the decision-dependent uncertainty and enhance computational  
 efficiency. This algorithm decompose the problem into inner  
 and outer problems based on Bayes' Theorem and Bayesian  
 Inference, then iteratively updates the empirical distribution  
 and the hardening decisions by sequential learning. Simu-  
 lation results on a real-world Redfield distribution network  
 demonstrate the strategy's capability to minimize worst-case

1042 expected unserved load while outperforming model-based  
1043 strategies.

1044 Future research can be extended in several promising  
1045 directions. First, the proposed strategy may be enhanced  
1046 by incorporating high-frequency phasor measurement unit  
1047 (PMU) data to capture transient phenomena such as cold load  
1048 pickup following extended outages. This integration would  
1049 enable more accurate modeling of load recovery dynamics  
1050 and further strengthen distribution system resilience. Second,  
1051 to address the computational challenges posed by the mixed-  
1052 integer optimization in the outer loop, learning-based surrogate  
1053 models or neural warm-start generators can be explored.  
1054 These tools have the potential to produce high-quality initial  
1055 solutions, reducing solution times and improving overall  
1056 scalability.

#### 1057 ACKNOWLEDGMENT

AQ:3 1058 The views expressed herein do not necessarily represent  
1059 the views of U.S. Department of Energy or United States  
1060 Government.

#### 1061 REFERENCES

- 1062 [1] Q. Shi, W. Liu, B. Zeng, H. Hui, and F. Li, "Enhancing distribution  
1063 system resilience against extreme weather events: Concept review,  
1064 algorithm summary, and future vision," *Int. J. Electr. Power Energy*  
1065 *Syst.*, vol. 138, Jun. 2022, Art. no. 107860.
- 1066 [2] U.S. Department of Energy. (2022). *Hurricane Ian*. [Online].  
1067 Available: [https://www.energy.gov/sites/default/files/2022-10/  
1068 Ian\\_DOE%20Situation%20Update\\_18.pdf?utm\\_source=chatgpt.com](https://www.energy.gov/sites/default/files/2022-10/Ian_DOE%20Situation%20Update_18.pdf?utm_source=chatgpt.com)
- 1069 [3] P. R. White, *A Method to Model the System-Wide Impacts of Residential*  
1070 *Heating Electrification Under Various Future Load Scenarios in Texas*.  
1071 Austin, Texas: Univ. Texas at Austin, 2021.
- 1072 [4] A. Ahmad and I. Dobson, "Towards using utility data to quantify how  
1073 investments would have increased the wind resilience of distribution  
1074 systems," *IEEE Trans. Power Syst.*, vol. 39, no. 4, pp. 5956–5968, Jul.  
1075 2024.
- 1076 [5] A. Ahmad and I. Dobson, "Quantifying distribution system resilience  
1077 from utility data: Large event risk and benefits of investments," *IET*  
1078 *Conf. Proc.*, vol. 2024, no. 27, pp. 118–122, 2025.
- 1079 [6] X. Wang, Z. Li, M. Shahidepour, and C. Jiang, "Robust line hardening  
1080 strategies for improving the resilience of distribution systems with  
1081 variable renewable resources," *IEEE Trans. Sustain. Energy*, vol. 10,  
1082 no. 1, pp. 386–395, Jan. 2019.
- 1083 [7] T. Li, X. Han, W. Wu, and H. Sun, "Robust expansion planning and  
1084 hardening strategy of meshed multi-energy distribution networks for  
1085 resilience enhancement," *Appl. Energy*, vol. 341, Jul. 2023, Art. no.  
1086 121066.
- 1087 [8] C. He, C. Dai, L. Wu, and T. Liu, "Robust network hardening strategy for  
1088 enhancing resilience of integrated electricity and natural gas distribution  
1089 systems against natural disasters," *IEEE Trans. Power Syst.*, vol. 33,  
1090 no. 5, pp. 5787–5798, Sep. 2018.
- 1091 [9] Y. Lin and Z. Bie, "Tri-level optimal hardening plan for a resilient  
1092 distribution system considering reconfiguration and DG islanding," *Appl.*  
1093 *Energy*, vol. 210, pp. 1266–1279, Jan. 2018.
- 1094 [10] H. Zhang, S. Ma, T. Ding, Y. Lin, and M. Shahidepour, "Multi-  
1095 stage multi-zone defender-attacker-defender model for optimal resilience  
1096 strategy with distribution line hardening and energy storage system  
1097 deployment," *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 1194–1205,  
1098 Mar. 2021.
- 1099 [11] C. Wang, K. Pang, M. Shahidepour, F. Wen, and S. Duan, "Two-  
1100 stage robust design of resilient active distribution networks considering  
1101 random tie line outages and outage propagation," *IEEE Trans. Smart*  
1102 *Grid*, vol. 14, no. 4, pp. 2630–2644, Jul. 2023.
- 1103 [12] W. Gan et al., "A tri-level planning approach to resilient expansion and  
1104 hardening of coupled power distribution and transportation systems,"  
1105 *IEEE Trans. Power Syst.*, vol. 37, no. 2, pp. 1495–1507, Mar. 2022.
- 1106 [13] S. Ma, B. Chen, and Z. Wang, "Resilience enhancement strategy for  
1107 distribution systems under extreme weather events," *IEEE Trans. Smart*  
1108 *Grid*, vol. 9, no. 2, pp. 1442–1451, Mar. 2018.
- 1109 [14] H. Hou et al., "Resilience enhancement of distribution network under  
1110 typhoon disaster based on two-stage stochastic programming," *Appl.*  
1111 *Energy*, vol. 338, May 2023, Art. no. 120892.
- 1112 [15] H. Chen, J. Wang, J. Zhu, X. Xiong, W. Wang, and H. Yang, "A  
1113 two-stage stochastic mixed-integer programming model for resilience  
1114 enhancement of active distribution networks," *J. Modern Power Syst.*  
1115 *Clean Energy*, vol. 11, no. 1, pp. 94–106, Jan. 2023.
- 1116 [16] Y. Tan, A. K. Das, P. Arabshahi, and D. S. Kirschen, "Distribution  
1117 systems hardening against natural disasters," *IEEE Trans. Power Syst.*,  
1118 vol. 33, no. 6, pp. 6849–6860, Nov. 2018.
- 1119 [17] Y. Li, S. Lei, W. Sun, C. Hu, and Y. Hou, "A distributionally robust  
1120 resilience enhancement strategy for distribution networks considering  
1121 decision-dependent contingencies," *IEEE Trans. Smart Grid*, vol. 15,  
1122 no. 2, pp. 1450–1465, Mar. 2024.
- 1123 [18] Z. Wang, B. Chen, J. Wang, M. M. Begovic, and C. Chen, "Coordinated  
1124 energy management of networked microgrids in distribution systems,"  
1125 *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 45–53, Jan. 2015.
- 1126 [19] A. Bagheri, C. Zhao, F. Qiu, and J. Wang, "Resilient transmission  
1127 hardening planning in a high renewable penetration era," *IEEE Trans.*  
1128 *Power Syst.*, vol. 34, no. 2, pp. 873–882, Mar. 2019.
- 1129 [20] A. Bagheri and C. Zhao, "Distributionally robust reliability assessment  
1130 for transmission system hardening plan under  $N-k$  security criterion,"  
1131 *IEEE Trans. Rel.*, vol. 68, no. 2, pp. 653–662, Jun. 2019.
- 1132 [21] S. Rahim, Z. Wang, K. Sun, and H. Chen, "A Wasserstein distribu-  
1133 tionally robust model for transmission expansion planning with  
1134 renewable-based microgrid penetration," *IET Gener. Transm. Distrib.*,  
1135 vol. 18, no. 17, pp. 2793–2808, 2024.
- 1136 [22] Y. Zhang, T. Huang, and E. F. Bompard, "Big data analytics in  
1137 smart grids: A review," *Energy Informat.*, vol. 1, no. 1, pp. 1–24,  
1138 Dec. 2018.
- 1139 [23] N. K. Carrington, I. Dobson, and Z. Wang, "Extracting resilience  
1140 metrics from distribution utility data using outage and restore process  
1141 statistics," *IEEE Trans. Power Syst.*, vol. 36, no. 6, pp. 5814–5823,  
1142 Nov. 2021.
- 1143 [24] Z. Wang and J. Wang, "Self-healing resilient distribution systems based  
1144 on sectionalization into microgrids," *IEEE Trans. Power Syst.*, vol. 30,  
1145 no. 6, pp. 3139–3149, Nov. 2015.
- 1146 [25] W. A. dos Santos Fonseca, U. H. Bezerra, M. V. A. Nunes, F. G. N. Bar-  
1147 ros, and J. A. P. Moutinho, "Simultaneous fault section estimation  
1148 and protective device failure detection using percentage values of the  
1149 protective devices alarms," *IEEE Trans. Power Syst.*, vol. 28, no. 1,  
1150 pp. 170–180, Feb. 2013.
- 1151 [26] Y. M. Darestani and A. Shafieezadeh, "Multi-dimensional wind fragility  
1152 functions for wood utility poles," *Eng. Struct.*, vol. 183, pp. 937–948,  
1153 Mar. 2019.
- 1154 [27] Z. Wang and J. Wang, "Time-varying stochastic assessment of conserva-  
1155 tion voltage reduction based on load modeling," *IEEE Trans. Power*  
1156 *Syst.*, vol. 29, no. 5, pp. 2321–2328, Sep. 2014.
- 1157 [28] M. Farivar and S. H. Low, "Branch flow model: Relaxations and  
1158 convexification—Part I," *IEEE Trans. Power Syst.*, vol. 28, no. 3,  
1159 pp. 2554–2564, Aug. 2013.
- 1160 [29] R. R. Nejad and W. Sun, "Distributed load restoration in unbalanced  
1161 active distribution systems," *IEEE Trans. Smart Grid*, vol. 10, no. 5,  
1162 pp. 5759–5769, Sep. 2019.
- 1163 [30] Z. Ma, Y. Xiang, and Z. Wang, "Robust conservation voltage  
1164 reduction evaluation using soft constrained gradient analysis,"  
1165 *IEEE Trans. Power Syst.*, vol. 37, no. 6, pp. 4485–4496,  
1166 Nov. 2022.
- 1167 [31] T. Ding, Y. Lin, G. Li, and Z. Bie, "A new model for resilient distribution  
1168 systems by microgrids formation," *IEEE Trans. Power Syst.*, vol. 32,  
1169 no. 5, pp. 4145–4147, Sep. 2017.
- 1170 [32] H. Rahimian and S. Mehrotra, "Distributionally robust optimization: A  
1171 review," 2019, *arXiv:1908.05659*.
- 1172 [33] I. E. Mesonet. (2025). *Asos-Awos-Metar Data*. [Online]. Available:  
1173 <https://mesonet.agron.iastate.edu/>
- 1174 [34] U.S. Department of Energy. (2024). *Low-Cost*  
1175 *Grid Resilience Projects*. [Online]. Available:  
1176 [https://www.energy.gov/sites/default/files/2024-01/2024-01-18%20Low-  
1177 Cost%20Grid%20Resilience%20Projects%20Fact%20Sheet.pdf](https://www.energy.gov/sites/default/files/2024-01/2024-01-18%20Low-Cost%20Grid%20Resilience%20Projects%20Fact%20Sheet.pdf)
- 1178 [35] California Public Utilities Commission. (2019). *CPUC*  
1179 *Undergrounding Programs Description*. [Online]. Available:  
1180 [https://docs.cpuc.ca.gov/PublishedDocs/Efile/G000/M258/K116/  
1181 258116720.PDF](https://docs.cpuc.ca.gov/PublishedDocs/Efile/G000/M258/K116/258116720.PDF)
- 1182 [36] Pacific Gas and Electric Company. (2024). *PG&E's*  
1183 *Unit Cost Guide*. [Online]. Available: [https://  
1184 www.pge.com/assets/pge/docs/about/doing-business-with-pge/  
1185 unit-cost-guide.pdf](https://www.pge.com/assets/pge/docs/about/doing-business-with-pge/unit-cost-guide.pdf)

1186  
1187  
1188  
1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1198  
1199



**Wenlong Shi** (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Jilin University, China, in 2011 and 2014, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Alberta, Edmonton, Canada, in 2024.

He is currently a Post-Doctoral Research Associate with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, USA. His research interests include power system resilience, data-driven and stochastic decision-making, outage analytics, and distribution system restoration under climate extremes. He also works on the integration of electrification, distributed energy resources, and policy feasibility in future resilient energy systems.

1200  
1201  
1202  
1203  
1204  
1205  
1206  
1207  
1208



**Hongyi Li** (Member, IEEE) received the B.Eng. degree from the Huazhong University of Science and Technology in 2020, the M.Sc. degree from Imperial College London in 2021, and the Ph.D. degree from the University of Macau in 2024. He is currently a Post-Doctoral Researcher with Iowa State University. His research interests include AI-aided optimization, distributed energy resource operation, and other smart energy technologies.



**Zhaoyu Wang** (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Shanghai Jiao Tong University and the M.S. and Ph.D. degrees in electrical and computer engineering from Georgia Institute of Technology. Since 2015, he has been an Assistant, an Associate, and a Full Professor with Iowa State University. He is the lead Principal Investigator for over 24M projects funded by the National Science Foundation, the Department of Energy, National Laboratories, PSERC, and Iowa Economic Development Authority.

His research interests include optimization and data analytics in power distribution systems and microgrids. He was a recipient of the National Science Foundation CAREER Award, the IEEE Power and Energy Society (PES) Outstanding Young Engineer Award, the ISU Award for Mid-Career Achievement in Research, the Northrop Grumman Endowment, the ISU College of Engineering's Early Achievement in Research Award, and the Harpole-Pentair Young Faculty Award Endowment. He is the Secretary and the Technical Committee Program Chair (TCPC) of the IEEE Power System Operation, Planning and Economics (PSOPE) Committee, the Vice Chair of the IEEE Distribution System Operation and Planning Subcommittee, the Secretary of the IEEE Task Force on IEEE P3102 Standard for Conservation Voltage Reduction (CVR) Data Collection and Management Procedures, and the Chair of IEEE Task Force on Advances in Natural Disaster Mitigation Methods. He is an Associate Editor of IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, IEEE OPEN ACCESS JOURNAL OF POWER AND ENERGY, IEEE POWER ENGINEERING LETTERS, and *IET Smart Grid*. He was an Associate Editor of IEEE TRANSACTIONS ON POWER SYSTEMS and IEEE TRANSACTIONS ON SMART GRID. He is an IEEE PES Distinguished Lecturer.

1209  
AQ:4  
1211  
1212  
1213  
1214  
1215  
1216  
1217  
1218  
1219  
1220  
1221  
1222  
1223  
1224  
1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238